EJONS International Journal on Mathematic, Engineering and Natural Sciences | ISSN 2602 - 4136

Article Arrival Date 18.05.2020 | Article Type Research Article | Article Published Date 12.06.2020

Doi Number: http://dx.doi.org/10.38063/ejons.255

EVALUATION OF GAIT BEHAVIOR WITH STATE SPACE VECTORS FOR CLASSIFICATION OF NEURODEGENERATIVE DISEASES

Derya YILMAZ

Assistant Prof., Başkent University, Department of Electrical and Electronics Engineering, Ankara/Turkey

derya@baskent.edu.tr, ORCID: 0000-0002-1903-7132

ABSTRACT

Walking is a set of movements that take place with the control of the motor nervous system. Since neurodegenerative diseases caused by motor control disorders change the gait dynamics of individuals, analysis of gait signals is used in the detection of such diseases. In this study, it is considered that the gait behavior can be evaluated from a systemic perspective to reveal its characteristics. With this motivation, it is proposed to analyze the gait signals with the state space approach. In accordance with this approach, the gait behavior was evaluated with the state vectors constructed in the state spaces of different dimensions. The state spaces were formed for the only left foot, only right foot, and the whole system, and it was also studied with the state spaces that enable the left and right feet to be evaluated together. The new series were obtained by calculating the sizes of the state vectors obtained from each state space and some statistical features were extracted from these series. To evaluate the success of the proposed approach in discriminating neurodegenerative diseases from healthy ones, classification was done with the K Nearest Neighbor (KNN) algorithm by using ten-fold cross-correlation validation. The results reveal that state vectors are effective in detecting neurodegenerative diseases, with 90.5% accuracy. It is considered that the proposed approach may yield useful results in the analysis of other physiological signals.

Keywords: Gait dynamics, State space, State vector, KNN.

1. INTRODUCTION

Walking, which is a physical form of movement, is an action conducted by the nervous system. Therefore, gait characteristics and gait analysis are used for the detection of neurodegenerative diseases with motor disorders due to damages to the nervous system. These neurodegenerative diseases, known as Parkinson's (PARK), Amyotrophic Lateral Sclerosis (ALS) and, Huntington (HUNT), caused by the death of neurons, loss of structure or function, cannot be cured completely, but it is important to begin early treatment, for slowing their progression. Therefore, in recent years, the development of automated and non-invasive methods based on gait dynamics has become an important issue of work for the diagnosis of these diseases (Carletti et al., 2006, Jin et al., 2018).

Gait-related features have a fluctuation that varies from step to step, even in healthy individuals. In the studies conducted for the detection of neurodegenerative diseases, the focus is on determining the characteristic features that can usually distinguish the diseased gait from healthy. For people with neurodegenerative disease, the dynamics of these fluctuations show different properties from the healthy ones, and these changes can be seen in the structural characteristics of the gait signal, primarily in statistical features (Pailhous, 1992; Hausdorff et al., 1995; Hausdorff et al., 1998; Hausdorff et al., 2000). Hausdorff et al., (2000), evaluated the magnitudes of step-to-step fluctuations and gait rhythm for healthy individuals and subjects with ALS, PARK, and HUNT.

They showed that the gait variability increased in all three diseased groups. In recent years, gait signals and some machine learning methods have been used together to detect neurodegenerative diseases. Wu and Krishnan, (2009) were able to distinguish ALS patients from healthy individuals with an accuracy of 89.66% by determining the fluctuations in the oscillation interval data using the signal turn numbers. In another similar study by the same authors, the gait variability was evaluated by taking into account signal turns for stride interval signals, and a significant difference was found between healthy and PARK patients (Wu and Krishnan, 2010). The Parzen window method was used to calculate the probability density functions of the stride signals of PARK patients, and an accuracy of 90.32% was achieved. Dutta et al., (2009) achieved the best average accuracy of 87.1% in discriminating of the healthy and diseased groups consisting of PARK, ALS, and HUNT patients, using Elman's repetitive neural network and various statistical parameters. In the study of Daliri, (2012); when ALS, PARK, and HUNT patients and healthy individuals were classified separately with the Support Vector Machine (SVM), 96.79%, 89.33%, and 90.28% accuracies were obtained, respectively. When it evaluated all diseased groups together, it reached 90.63% accuracy. In another study evaluating the features of gait by using the wavelet transform for the detection of Parkinson's disease, classification accuracy of 77.33% was obtained (Lee and Lim, 2012). Baratin et al., (2015) also achieved an average performance of 85% by using a wavelet-based approach and SVM classifier for patients with PARK, HUNT, and ALS, and the healthy individuals. Xia et al., (2015) evaluated nine statistical features obtained from gait signals in their study with four different classifiers to differentiate neurodegenerative diseases (ALS, PARK, HUNT) from healthy individuals, and achieved the highest performance as 96.83% with SVM. In the study of Kamath (2017)'s, the left and right stride intervals were examined by using two Poincaré plot symbolic complexity quantities. The Receiver Operating Characteristic (ROC) analysis was applied for comparisons between the healthy and three neurodegenerative disorder groups, for two features and for two feet, separately. The best accuracies obtained in their study were 100%, 94.4%, and 87.5% for ALS, HUNT, and PARK groups, with right foot stride signal, respectively. Aydın and Aslan, (2017) used wavelet transform and several classifiers to detect neurodegenerative diseases of ALS, HUNT, and PARK. They achieved 92.18% accuracy in the comparison between the healthy and diseased groups with an Additive Logistic Regression algorithm. Ye et al., (2018) classified the gait signals obtained from PARK, HUNT, and ALS patients, with an Adaptive Neural Fuzzy Inference System (ANFIS) and they were able to differentiate healthy individuals with 90.63% accuracy.

Considering the results of the studies in the literature; in the presence of neurological disorders, gait dynamics such as gait rhythm, speed, step-to-step gait variability, and gait symmetry are different from healthy individuals. In the literature, time domain features, statistical and nonlinear features were calculated to find these differences and applied to classification algorithms for disease detection. In these studies, generally, signals belonging to a single foot or two feet were used. In some others which are using all signals, the signals were individually evaluated. In light of the realized works in the literature, it was studied on the evaluation of gait signals together in the state spaces with the approach that the neurodegenerative diseases cause some changes in the system behavior, and therefore change the state vector sizes.

It is anticipated that recording more than one signal from a system at the same time and evaluating these signals together in investigating the system behavior can provide useful results. In this study, unlike previous studies, the behavior of the system over time is evaluated in the state space. For this purpose and with this approach, multidimensional state space vectors were obtained by using gait signals. For the behavior of each foot, the signals recorded from the left and right feet were evaluated in separate state spaces, and for the total behavior of the system, all signals were used together in the higher dimensional state space. Besides, the relationship between the left and right feet was evaluated, and for this purpose, signal type-based state spaces were constructed. Statistics based features were calculated from the sizes of the state space vectors. To evaluate the success of the proposed approach, the classification task was carried out with the obtained features.

2. MATERIALS AND METHODS

2.1. Gait Data

In this study, the gait dataset recorded by Hausdorff et al., (1995; 1998 and 2000) and used by many other studies (Wu and Krishnan, 2010; Dutta et al., 2009; Daliri, 2012; Lee and Lim, 2012; Baratin et al., 2015; Xia et al., 2015; Ye et al., 2018) was used and downloaded from PhysioNet database (Goldberger et al., 2003). The dataset contains a total of 64 gait recordings of four different groups: 20 HUNT, 15 PARK, 13 ALS, and 16 healthy controls (CO). Since the record for the HUNT subject number of 20 was disordered it was excluded from the study. According to the experiment protocol, the recordings were obtained from the force-sensitive sensors, placed on their shoes while the subjects were walking with a normal speed at a 77 m long straight corridor for 300 seconds. Gait signals were recorded by sampling at 300 Hz with a 12-bit analog-to-digital converter. The first 20 seconds of the recordings were removed to eliminate the starting effect, a median filter was applied to filtering the noises caused by the return at the end of the corridor (Hausdorff et al., 2000), and the signals were normalized. The dataset has the stride time interval (the time between consecutive contacts of the foot to the ground), the swing time interval (the time elapsed with the foot in the air), stance time interval (the time the person stands on their foot) and the double support time interval (the time that both feet touch the ground) recordings, from left and right feet and listed in Table 1. Figure 1 shows the left foot stride intervals before (at the left column) and after the filtering and normalization process (at the right column) for the subjects number of 1 for all groups.

Table 1.The gait dataset recordings from PhysioNet.

No	Data
1	Elapsed Time (sec)
2	Left Stride Interval (sec)
3	Right Stride Interval (sec)
4	Left Swing Interval (sec)
5	Right Swing Interval (sec)
6	Left Swing Interval (% of stride)
7	Right Swing Interval (% of stride)
8	Left Stance Interval (sec)
9	Right Stance Interval (sec)
10	Left Stance Interval (% of stride)
11	Right Stance Interval (% of stride)
12	Double Support Interval (sec)
13	Double Support Interval (% of stride)

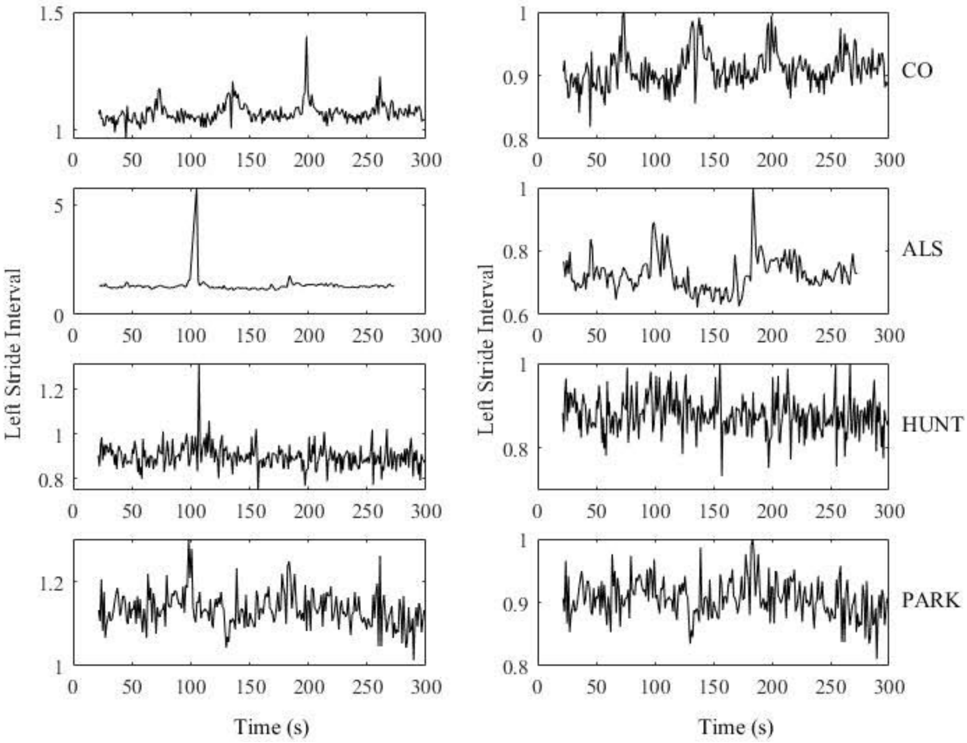

Figure 1. Left stride intervals before (left column) and after the preprocessing (right column) for all groups.

2.2. State Space Approach and State Space Vectors

State space is a way of understanding the behavior of a dynamic system (Williams, 1997). It allows us to evaluate all the states that the system takes over time. The axes of the state space represent the variables of the system. In other words, every variable obtained from the system corresponds to one dimension of the state space. Therefore, as the number of variables increases, the dimensions of the state space increases. While it is possible to visualize two and three dimensional state spaces, this is not possible in higher dimensions. Each point in the state space, whose coordinates are formed by variables, is a state vector that represents the current state of the system. If $x_n=\{x_1, x_2,...,x_N\}$ and $y_n=\{y_1, y_2,...,y_N\}$ are the signals of two different variables obtained from a system, where N is the total number of samples. The two-dimensional state vectors (X_n, n=1,2,...,N) in the two-dimensional state space are formed as in Equation 1 and are shown as in Figure 2.

$$X_n = [x_n, y_n] \tag{1}$$

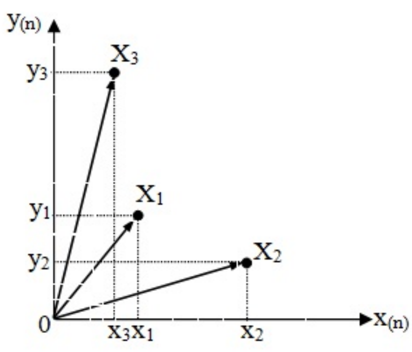

Figure 2. State vectors in two-dimensional state space.

The sizes of the state vectors in the state space are found by calculating the Euclidean distance (Williams, 1997; Spiegel et al. 2009). The size of one-dimensional vectors is the absolute value of the numerical difference of point coordinates. The Euclidean distance for two dimensions is calculated by Equation 2. This Euclidean distance equation, taking into account that the starting point of each vector is in the origin of the state space (point 0, initial point), is used as in Equation 3 in this study.

$$D_1 = \sqrt{(x_1 - x_0)^2 + (y_1 - y_0)^2} \quad (2)$$

$$D_n = \sqrt{x_n^2 + y_n^2} \quad (3)$$

If a new variable such as $z_n=\{z_1, z_2, ..., z_N\}$ is added to the system exemplified above, three-dimensional vectors, $X_n=[x_n, y_n, z_n]$, will be obtained, and in which case the three-dimensional Euclidean distance equation given in Equation 4 is used. In this study, Equation 4 has become in Equation 5 for the calculation of vector sizes. In Equation 3 and 5, n=1,2,...,N and N is the total number of samples and vectors.

$$D_1 = \sqrt{(x_1 - x_0)^2 + (y_1 - y_0)^2 + (z_1 - z_0)^2} \quad (4)$$

$$D_n = \sqrt{x_n^2 + y_n^2 + z_n^2} \quad (5)$$

The sizes of the vectors in the higher dimensions are also calculated as n-dimensional Euclidean distance. If there are n variables (signals), all variables are added to the equation as in Equation 5, and the Euclidean distance in n-dimensional state space is calculated in a similar way.

2.3. The New Series and Extracted Features

In this study, many state spaces were obtained. The sizes (distances) of the state vectors in each of these state spaces were calculated using Euclidean distance equations, and these distances were obtained as the new series. In this case, the number of new series containing Euclidean distances is equal to the number of state spaces. The state spaces formed by using gait intervals and the names of the new series are given in Table 2. In the study, the behaviors of the left and right feet were evaluated in three-dimensional state spaces, separately. These state spaces are indicated in Table 2 as 6th (for left foot) and 7th (for right foot) state spaces and the distances series obtained from these three-dimensional vectors were named as the DL and DR, respectively. The behavior of the entire system was evaluated in six-dimensional state space using all of the left and right feet gait intervals which in seconds (shown in Table 2 as 8th state space). The DA is another new distance series obtained from six-dimensional state vectors in this case. In addition, in order to examine the behaviors of both feet at the same time, the signals belonging to the same variable obtained from the left and right feet were evaluated in two-dimensional state spaces. In this case, one of the axes of these two-dimensional state spaces was the signals received from the left and the other axis from the right foot. In this way, five state spaces were constructed (from 1 to 5 in Table 2) and five new series (D1-D5) were obtained from the two-dimensional vector distances in these spaces.

Table 2. The properties of state spaces and the names of the distance series.

Number of state space	Dimension of state space	The gait intervals used in state spaces	Name of new series
1	2	Left Stride Interval (sec) Right Stride Interval (sec)	D1
2	2	Left Swing Interval (sec) Right Swing Interval (sec)	D2
3	2	Left Swing Interval (% of stride) Right Swing Interval (% of stride)	D3
4	2	Left Stance Interval (sec) Right Stance Interval (sec)	D4
5	2	Left Stance Interval (% of stride) Right Stance Interval (% of stride)	D5
6	3	Left Stride Interval (sec) Left Swing Interval (sec) Left Stance Interval (sec)	DL
7	3	Right Stride Interval (sec) Right Swing Interval (sec) Right Stance Interval (sec)	DR
8	6	Left Stride Interval (sec) Left Swing Interval (sec) Left Stance Interval (sec) Right Stride Interval (sec) Right Swing Interval (sec) Right Stance Interval (sec)	DA

In Figure 3, two-dimensional state spaces constructed by using left and right stride intervals in seconds for the subjects number of 1 from all groups are shown (1st state space in Table 2). The sizes of vectors in these state spaces were calculated and D1 distance signals were obtained for all subjects. The other two dimensional vector sizes (D2, D3, D4, and D5) were also calculated from the other two dimensional state spaces indicated in Table 2. The D1-D5 series calculated for all groups (from subject numbers of 1) are presented in Figure 4.

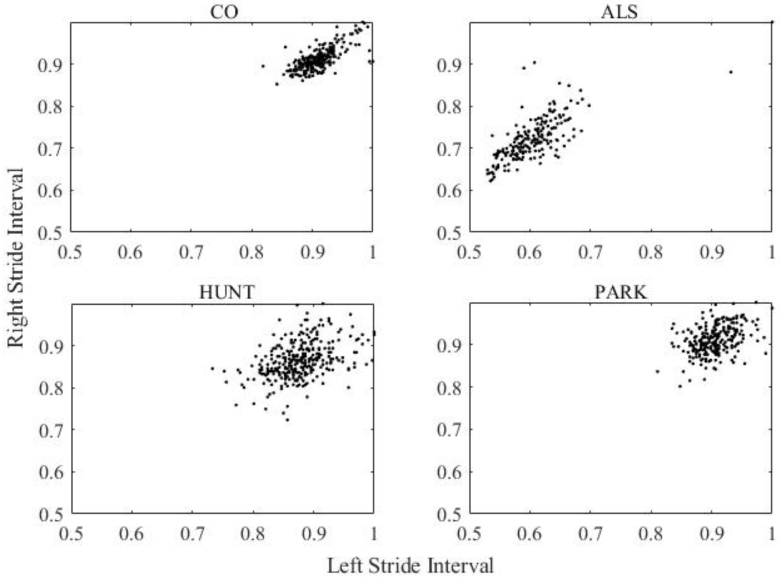

Figure 3. Two-dimensional state spaces for left and right stride intervals.

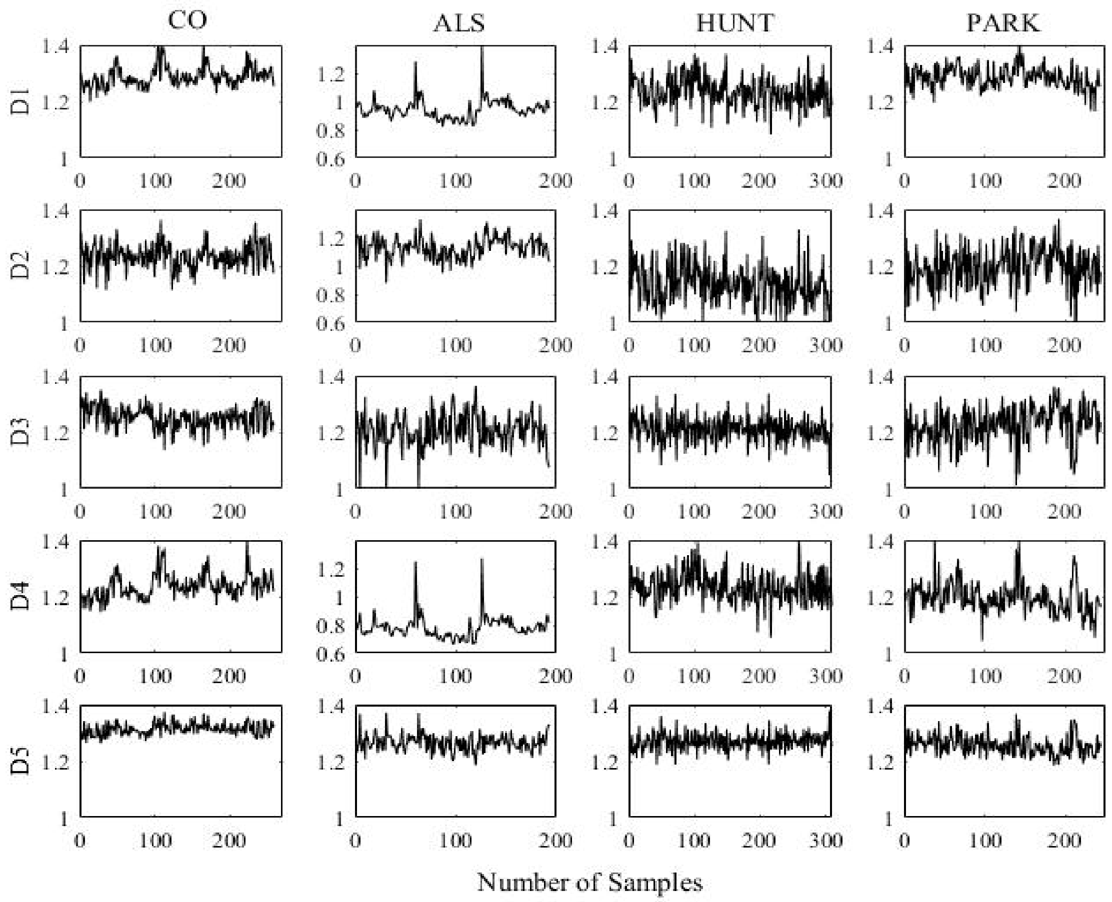

Figure 4. D1-D5 series from two-dimensional state vectors for all groups (subjects no:1).

The three-dimensional state spaces obtained from left foot signals are presented in Figure 5 for all groups. In three-dimensional state spaces formed for each foot, the stride (x-axis), swing (y-axis) and stance (z-axis) intervals (in seconds) were used (6th and 7th state spaces in Table 2). The DL (left) and DR (right) vector distance series were calculated by using three-dimensional Euclidean distance. Finally, the DA (all) distance series was obtained from six-dimensional state space formed by using six gait intervals (in seconds) from both feet (8th state space in Table 2). The DL, DR, and DA distance series are given in Figure 6 for four groups with the subjects number of 1.

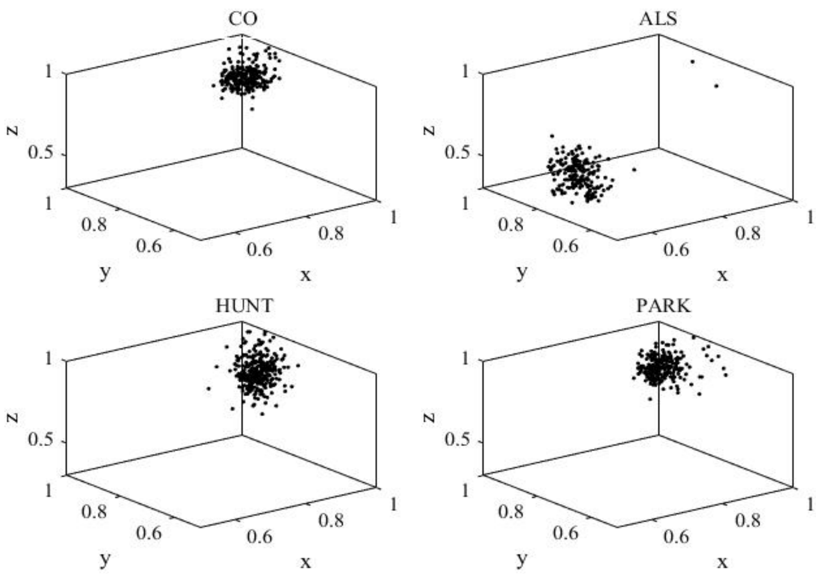

Figure 5. Three-dimensional state spaces for stride, swing and stance intervals of left foot for all groups (subjects no:1).

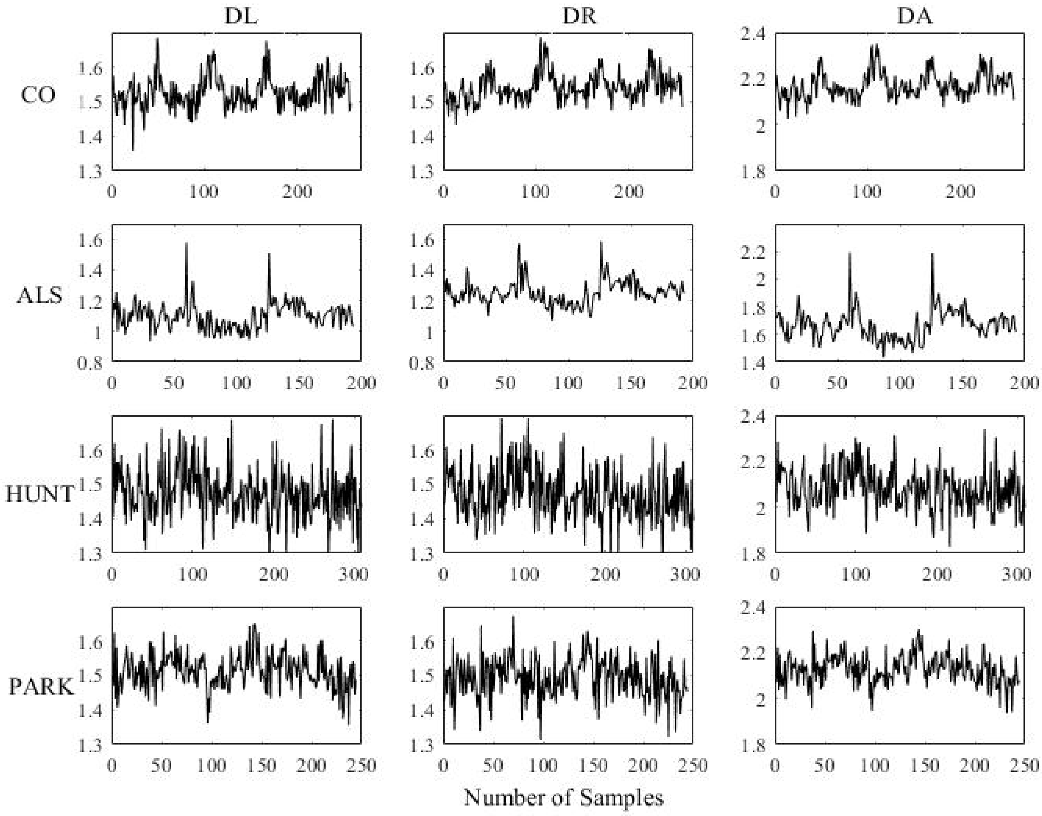

Figure 6. DL, DR, DA distance series obtained from state spaces of left foot, right foot and whole system, respectively.

In this way, the signals in the gait data set were converted into eight new series and some features were calculated from these new signals. The five statistical-based features, which are mean, standard deviation (SD), median, maximum (Max) and minimum (Min), were calculated for each distance series. Thus, 40 features were obtained from the data of each subject, totally. The mean and SDs of the features obtained from D1, D2, D3, D4, D5, DL, DR, DA series for the control and neurodegenerative disease (ND) groups are given in Table 3. These features were compared statistically, between the CO and ND groups, using the SPSS 17.0 program. The non-parametric Mann-Whitney U test was applied ($p<0.05$ level), considering the size of the dataset, and the p values are also given in Table 3. According to the test results, p values for all features were found as less than 0.05, so all features are significantly different between the two groups.

2.4. K Nearest Neighbor Classifier

K Nearest Neighbor (KNN) is a non-parametric algorithm and a type of instance-based learning (Cover and Hart, 1967; Zelula et al., 2006). KNN is one of the most known and frequently used algorithm among machine learning algorithms. This algorithm uses the benchmark learning rules, that is, it has not any assumptions relating to the distribution of data. The KNN directly uses the training data for classification. The algorithm is based on comparing the class memberships of the test vector with known vectors. The class value of the input vector is assigned by looking at the classes of the closest K neighboring vector by comparison. The Cosine, Euclidean, Manhattan, City-block, etc. distances can be selected for calculation of the distance between the neighbors. The success and performance of the KNN may vary depending on the distance methods and the number of K neighbors. When the number of neighbors is taken 1, the class of the undefined test vector is learned using a single training example that is closest to it. In cases where the number of neighbors is more than 1, the class of this vector is made by looking at the number of the majority. In order to avoid any equality situation, in practice, k value is usually taken as an odd number. This algorithm, due to its simple structure, is easy to implement, simple to calculate, and works well when there are few feature vectors.

Table 3. The mean±SD values of the features calculated from new series for CO and ND groups.

		Mean	SD	Median	Min	Max
D1	CO	1.28±0.04	0.03±0.01	1.27±0.04	1.20±0.05	1.40±0.01
	ND	1.10±0.23	0.07±0.03	1.09±0.23	0.95±0.26	1.36±0.08
		p <0.001	p <0.001	p <0.001	p <0.001	p <0.05
D2	CO	1.26±0.04	0.04±0.01	1.25±0.04	1.15±0.06	1.38±0.02
	ND	1.08±0.15	0.08±0.04	1.08±0.15	0.84±0.25	1.33±0.06
		p <0.001	p <0.001	p <0.001	p <0.001	p <0.001
D3	CO	1.30±0.03	0.03±0.01	1.92±0.19	1.94±0.20	1.95±0.21
	ND	1.16±0.06	0.06±0.04	1.80±0.65	1.76±0.67	1.84±0.56
		p <0.001	p <0.001	p <0.05	p <0.05	p <0.05
D4	CO	2.02±0.24	0.04±0.01	1.24±0.06	1.15±0.07	1.38±0.03
	ND	1.78±0.58	0.07±0.03	1.03±0.26	0.89±0.27	1.34±0.09
		p <0.05	p <0.001	p <0.001	p <0.001	p <0.05
D5	CO	1.34±0.02	0.02±0.00	1.34±0.02	1.30±0.03	1.40±0.01
	ND	1.26±0.07	0.03±0.02	1.26±0.07	1.17±0.12	1.38±0.02
		p <0.001	p <0.001	p <0.001	p <0.001	p <0.001
DL	CO	1.54±0.06	0.04±0.01	1.54±0.06	1.44±0.08	1.70±0.02
	ND	1.32±0.22	0.09±0.04	1.31±0.30	1.64±0.06	1.33±0.21
		p <0.001	p <0.001	p <0.001	p <0.001	p <0.001
DR	CO	1.54±0.05	0.04±0.01	1.54±0.05	1.44±0.06	1.69±0.02
	ND	1.33±0.21	0.08±0.04	1.32±0.22	1.12±0.30	1.64±0.07
		p <0.001	p <0.001	p <0.001	p <0.001	p <0.001
DA	CO	2.18±0.07	0.05±0.01	2.17±0.08	2.05±0.09	2.37±0.03
	ND	1.87±0.30	0.11±0.05	1.86±0.30	1.62±0.40	2.24±0.16
		p <0.001	p <0.001	p <0.001	p <0.001	p <0.001

3. RESULTS and CONCLUSION

In this study, in order to evaluate the potential of the proposed state vector approach, its success in discriminating healthy and diseased groups with the calculated features was evaluated by using the KNN classifier. To validate the performance of the classifier, the k-fold cross-correlation validation method was used (k=10). Firstly, the comparisons were achieved between each ND group and healthy group, and then the group of ND (ALS, HUNT, PARK) against the healthy group was classified. The classification results were presented in Table 4.

Table 4. Classification performance of the proposed study.

	Specificity (%)	**Sensitivity (%)**	**Accuracy (%)**	**AUC**
CO vs ALS	84.21	100	89.70	0.79
CO vs HUNT	88.89	100	94.30	0.97
CO vs PARK	88.89	100	93.50	0.98
CO vs ND	75.00	97.67	90.50	0.92

AUC: Area Under Curve

The results of some studies in the literature and the results obtained in this study are presented in Tables 5-8. In Table 5, the performance results of the comparisons for CO vs ND groups are given. For CO vs ND comparison, Ye et al., (2018) found the values of specificity and sensitivity as 87.50% and 91.67%, respectively. In the study of Daliri (2012)'s, these values obtained as 92.87% specificity and 90.65% sensitivity. Baratin et al. (2015) reported specificity of 79% and sensitivity of 92.2% for CO vs ND. Xia et al. (2015) noticed 0.94 of specificity and 0.98 of sensitivity as the best values they get. When the results in Table 4 are compared with the results of these previous studies, it can be said that the sensitivity values obtained in this study were found higher while the specificities were found lower. According to Table 5, Xia et al. (2015)'s accuracy is highest, but the accuracy rate achieved in this study is as high as the others.

Table 5. The performance results of the previous studies for discriminating CO vs ND.

Study	Classifier	Validation Method	Accuracy (%)
Daliri (2012)	SVM	Random sub-sampling	90.63
Baratin et al. (2015)	LDA	Leave-one-out Resubstitution	80.4 80.4
Xia et al. (2015)	SVM	Random sub-sampling Leave-one-out	93.67 96.83
Aydın and Aslan (2017)	LogitBoost	10 fold cross validation	92.18
Ye et al. (2018)	ANFIS	Leave-one-out	90.63
Proposed study	KNN	10 fold cross correlation	90.50

LDA: Linear Discriminant Analysis.

The performance results of the comparisons between ALS, HUNT, and PARK groups and CO, obtained from previous studies and the proposed study, are also given in Tables 6, 7, and 8, respectively. Daliri, (2012) reported the classification results between the ALS, HUNT, PARK, and CO groups as 96.79%, 90.28%, and 89.33%, respectively. Ye et al. (2018) classified the ALS, HUNT, and PARK groups with the accuracies of 93.10%, 94.44%, and 90.32%, respectively. Aydın and Aslan (2017) achieved the accuracies of 93.1%, 97.22%, and 83.87% for ALS, HUNT, and PARK, respectively. These results are similar to the results of the proposed study. The CO-PARK classification accuracy obtained in the proposed study was found higher than the results of Ye et al. (2018), Aydın and Aslan (2017), and some others. The study of Xia et al. (2015), achieved the highest accuracy compared to other studies for discriminating ALS, HUNT, and PARK groups from CO. Although the accuracy result for CO-ALS in the proposed study is slightly lower than those found for CO-HUNT and CO-PARK, the accuracy performance obtained especially for CO-PARK was evaluated as successful compared to the other studies.

Table 6. The performance results of the previous studies for discriminating CO vs ALS.

Study	Classifier	Validation Method	Accuracy (%)
Wu and Krishnan (2009)	SVM	Leave-one-out.	89.66
Daliri (2012)	SVM	Random sub-sampling	96.79
Baratin et al. (2015)	LDA	Leave-one-out Resubstitution	86.2 89.7
Xia et al. (2015)	SVM	Random sub-sampling Leave-one-out	81.5 96.55
Kamath (2017)	ROC Analysis	-	100
Aydın and Aslan (2017)	RBF Network	10 fold cross validation	93.1
Ye et al. (2018)	ANFIS	Leave-one-out	93.10
Proposed study	KNN	10 fold cross correlation	89.70

LDA: Linear Discriminant Analysis.

Table 7. The performance results of the previous studies for discriminating CO vs HUNT.

Study	Classifier	Validation Method	Accuracy (%)
Daliri (2012)	SVM	Random sub-sampling	90.28
Baratin et al. (2015)	LDA	Leave-one-out Resubstitution	86.1 96.7
Xia et al. (2015)	SVM	Random sub-sampling Leave-one-out	90.67 100
Kamath (2017)	ROC Analysis	-	94.4
Aydın and Aslan (2017)	AdaBoost	10 fold cross validation	97.22
Ye et al. (2018)	ANFIS	Leave-one-out	94.44
Proposed study	KNN	10 fold cross correlation	94.30

LDA: Linear Discriminant Analysis.

Table 8. The performance results of the previous studies for discriminating CO vs PARK.

Study	Classifier	Validation Method	Accuracy (%)
Wu and Krishnan (2010)	SVM	Leave-one-out	90.32
Daliri (2012)	SVM	Random sub-sampling	89.33
Baratin et al. (2015)	LDA	Leave-one-out Resubstitution	87.1 96.3
Xia et al. (2015)	SVM	Random sub-sampling Leave-one-out	84.67 100
Kamath (2017)	ROC Analysis	-	87.5
Aydın and Aslan (2017)	AdaBoost	10 fold cross validation	83.87
Ye et al. (2018)	ANFIS	Leave-one-out	90.32
Proposed study	KNN	10 fold cross correlation	93.50

LDA: Linear Discriminant Analysis.

As a result, when these studies are evaluated together, it can be said that the classification accuracies are sensitive to the selected classifier types and especially validation methods since the small size of the data set. The small size of the gait dataset can be considered as the reason for this situation. In the proposed study, unlike the previous studies, state vector sizes were used as a new series to investigate gait behavior. Additionally, 10-fold cross-correlation validation and KNN classifier were presented in this study. The obtained results indicated that the state space approach produces effective findings for detecting NDs and it can be used for other studies of biomedical signal processing. The classification performance of the proposed approach can be increased by using feature selection algorithms and trying other classifiers and validation methods.

REFERENCES

Aydın, F., Aslan, Z. (2017) Diagnosis of neuro degenerative diseases using machine learning methods and wavelet transform. Journal of the Faculty of Engineering and Architecture of Gazi University. 32:3, 749-766.

Baratin, E., Sugavaneswaran, L., Umapathy, K., Ioana, C., Krishnan, S. (2015). Wavelet-based characterization of gait signal for neurological abnormalities. Gait Posture. 41(2), 634-639.

Carletti T., Fanelli D., Guarino A. (2006). A new route to non invasive diagnosis in neurodegenerative diseases? Neuroscience Letters. 394, 252–255.

Cover, T., and Hart, P. (1967). Nearest neighbor pattern classification. IEEE Transactions on Information Theory. 13(1), 21–27.

Daliri M.R. (2012). Automatic diagnosis of neuro-degenerative diseases using gait dynamics, Measurement. 45, 1729-1734.

Dutta S., Chatterjee A., Munshi S. (2009). An automated hierarchical gait pattern identification tool employing cross-correlation-based feature extraction and recurrent neural network based classification. Expert Systems. 26 (2), 202-217.

Goldberger, A.L., Amaral, L.A.N., Glass, L, Hausdorff, J.M., Ivanov, P.Ch., Mark, R.G., Mietus, J.E., Moody, G.B., Peng, C.-K., Stanley, H.E. (2003). PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. Circulation. 101(23), e215-e220.

Hausdorff, J.M., Purdon, P.L., Peng, C.-K., Ladin, Z., Wei, J.Y., Goldberger, A.L. (1995). Is walking a random walk? Evidence for long-range correlations in the stride interval of human gait. Journal of Applied Physiology. 78, 349-358.

Hausdorff, J.M., Cudkowicz, M.E., Firtion, R., Wei, J.Y., Goldberger, A.L. (1998). Gait variability and basal ganglia disorders: stride-to-stride variations in gait cycle timing in Parkinson's and Huntington's disease. Movement Disorder. 13, 428-437.

Hausdorff, J.M., Lertratanakul, A., Cudkowicz, M.E., Peterson, A.L., Kaliton, D., Goldberger, A.L. (2000). Dynamic markers of altered gait rhythm in amyotrophic lateral sclerosis. Journal of Applied Physiology. 88(6), 2045-2053.

Jin, H., Li, C., Xu, J. (2018). Pilot Study on Gait Classification Using fNIRS Signals. Computational Intelligence and Neuroscience. 2018, Article ID 7403471, 9 pages.

Kamath, C. (2017). Influence of neurodegenerative disorders on gait dynamics using poincaré symbolic measures. MOJ Gerontol Geriatrics. 1(6), 164–173.

Lee S., Lim J.S. (2012) Parkinson's disease classification using gait characteristics and wavelet-based feature extraction. Expert Systems with Applications. 39, 7338-7344.

Pailhous, M. B. (1992). Steady-state fluctuations of human walking. Behavioral Brain Research. 47, 181–190.

Spiegel, M.R., Lipschutz, S., Spellman, D. (2009). Vector Analysis. 2. ed. McGraw Hill, USA.

Williams, G.P. (1997). Chaos Theory Tamed. 1. ed. London:CRC press.

Wu Y., Krishnan S. (2009). Computer-aided analysis of gait rhythm fluctuations in amyotrophic lateral sclerosis. Medical & Biological Engineering & Computing. 47, 1165-1171.

Wu Y., Krishnan S. (2010). Statistical analysis of gait rhythm inpatients with Parkinson's disease. IEEE Transactions on Neural Systems and Rehabilitation Engineering. 18(2), 150-158.

Xia Y., Gao Q., Ye Q. (2015). Classification of gait rhythm signals between patients with neuro-degenerative diseases and normal subjects: Experiments with statistical features and different classification models. Biomedical Signal Processing and Control. 18, 254-262.

Ye, Q., Xia, Y., Yao, Z. (2018). Classification of gait patterns in patients with neurodegenerative disease using adaptive neuro-fuzzy inference system. Computational and Mathematical Methods in Medicine. 2018, Article ID 9831252, 8 pages.

Zelula, P., Amato, G., Dohnal, V., Batko, M. (2006). Similarity Search the Metric Space Approach. Springer, US.