Semnan University

Journal of Modeling & Simulation in Electrical & Electronics Engineering (MSEEE)

Journal homepage: **https://mseee.semnan.ac.ir/**

ISSN: 2821-0786

Short-time Classification of Neurodegenerative Diseases Based on Cross-Recurrence Quantification Analysis and Statistical Features

Negar Monfared Jahromi [1] and Ali Maleki [2]*

***Abstract*— Neurodegenerative diseases refer to progressive disorders of the nervous system that impair motor functions. Using machine learning techniques to analyze and classify gait data can lead to early diagnosis and better management of treatments. This paper aims to classify neurodegenerative patients and healthy individuals by analyzing a three-second gait signal using a combination of effective features from cross-recurrence quantification analysis (CRQA) and statistical analysis. The dataset includes force signals from the left and right feet of 16 healthy individuals (HC), 13 with amyotrophic lateral sclerosis (ALS), 15 with Parkinson's disease (PD), and 20 with Huntington's disease (HD). The CRQA features extracted include recurrence rate, determinism, averaged diagonal length, length of longest diagonal length, entropy of diagonal length, laminarity, trapping time, length of longest vertical line, recurrence time of 1st type, recurrence time of 2nd type, recurrence period density entropy, clustering coefficient, and transitivity. Statistical features include mean, variance, skewness, and kurtosis. A sequential feature selection algorithm was used to select effective features. The classification accuracy for the Ensemble (Bagged Trees) classifier was obtained using 10-fold cross-validation for the groups HC vs. PD, HC vs. HD, HC vs. ALS, ALS vs. PD, ALS vs. HD, PD vs. HD, NDD vs. HC, and ALS vs. PD vs. HD vs. HC, with the respective accuracy values of 98.3%, 94.8%, 97.7%, 98.2%, 98.4%, 95%, 94%, and 93.5%. The results indicate that the effective fusion of features and the ability of cross-recurrence quantification analysis to quantify the synergistic relationship of the dynamic movements of the left and right feet provide an effective means of diagnosing diseases during the short 3-second walking period.**

***Index Terms- Cross-Recurrence Quantification Analysis, Statistical Features, Force Signal, Gait, Classification*.**

I. INTRODUCTION

NEURODEGENERATIVE disease (NDDs) have affected over 57 million people worldwide, and this number is expected to increase significantly by 2050 [1]. NDDs refers to the process of damage or death of nerves in different areas of the nervous system, leading to the loss of function and structure of neurons. NDDs include amyotrophic lateral sclerosis (ALS), Parkinson's disease (PD), and Huntington's disease (HD). The symptoms of neurodegenerative diseases, which are directly related to neuron function, include changes in executive functions, speech, memory, pain, and muscle weakness [2,3]. The most common symptoms are changes and fluctuations in gait patterns, such as slow walking and a stooped posture [4], which result in motor dysfunction [5]. Healthy individuals (HC) typically exhibit a rhythmic pattern of alternating left and right leg movements in contact with the ground [6]. This process naturally occurs with coordination between the central nervous system and the body's muscles, such that step lengths are nearly constant and coordinated, with alternating movement of the left and right legs. In individuals with Parkinson's disease (PD), there are fundamental disturbances in muscular control and movement coordination, which, according to Hasdorff's theory [7], result in shorter and more variable step lengths. These disturbances manifest as a 'shuffling gait' (short, closely spaced steps), leading to

Received; 2025-05-11 Revised; 2025-09-15 Accepted; 2025-09-24

[1]. Department of Biotechnology, Faculty of New Sciences and Technologies, Semnan University, Semnan, Iran.
[2]. Biomedical Engineering Department, Semnan University, Semnan, Iran.
Corresponding author Email: amaleki@semnan.ac.ir.

Cite this article as:
Monfared Jahromi, N. and Maleki, A. (2025). Short-time Classification of Neurodegenerative Diseases Based on Cross-Recurrence Quantification Analysis and Statistical Features. *Journal of Modeling & Simulation in Electrical & Electronics Engineering (MSEEE), 5*(2), 0-0.

DOI:https://doi.org/10.22075/MSEEE.2025.37723.1213

© 2025 The Author(s). Journal of Modeling & Simulation in Electrical & Electronics Engineering is published by Semnan University Press. This is an open-access article under the CC-BY 4.0 license. (https://creativecommons.org/licenses/by/4.0/)

reduced gait speed, shorter step length, and increased difficulty in walking [8]. In comparison to Parkinson's disease, walking in individuals with Huntington's disease (HD) is characterized by involuntary fluctuations and unpredictable movements during the step cycle. In HD, the abnormal accumulation of proteins in the brain causes motor impairments, leading to decreased stability and rhythm in walking [9]. Patients with HD often take non-alternating and variable steps, which may result in either shorter or longer step lengths. These fluctuations make movement control more difficult for them [10]. Walking in ALS patients is affected by severe muscular weakness, leading to a reduced ability to take natural and coordinated steps. In these patients, step lengths are typically longer than usual, but each step requires more time, and the movements are generally less coordinated and less uniform [11,12].

The differences in gait patterns across various diseases can serve as a criterion for disease classification. In many neurological diseases, early diagnosis not only helps accelerate treatment but may also have a significant impact on altering the treatment course [13-14]. However, in cases with mild disease symptoms, clinical specialists may be unable to make an accurate diagnosis. In such cases, using machine learning methods for timely diagnosis, disease classification, and precise assessment of its progression can play a vital role in improving the treatment process [17]. The following section reviews recent research on the analysis of motor signals for diagnosing and classifying neurodegenerative diseases, focusing on applying recurrence quantification analysis (RQA) features.

Faisal et al. [18] designed the NDDNet model for identifying neurodegenerative diseases by processing force signals and gait pattern features. This method was able to detect abnormal gait patterns with an accuracy of 96.75%. Zhou et al. [19] compared seven deep learning architectures for classifying NDD patients using gait data. Results showed that the ResNet model performed better in distinguishing healthy individuals from neurodegenerative and Parkinson's patients, while the TST model achieved the highest accuracy in detecting ALS and HD from the healthy group. Visvanathan et al. [20] extracted features using switching state-space decomposition and Shannon entropy. Classification was performed by a genetic algorithm-optimized perceptron, achieving an accuracy of 98.4%. Zhou et al. [21] designed a dual-channel LSTM model to combine time-series features and the recorded force signal from NDD patients. The results of these studies are reported in Table 3.

Recurrence quantification analysis (RQA) is a suitable tool for processing nonlinear and dynamic data, which is used in areas such as detecting abnormal heartbeats [22]. Subsequently, recent studies on NDD classification using RQA in gait patterns have been reviewed. Prabhu et al. [23] extracted eight features from RQA for the time series gait data of Huntington's, Alzheimer's, Parkinson's patients, and healthy individuals, evaluating the efficacy of RQA features such as stride, stance, and swing interval for classifying these patients. Prabhu [24] extracted RQA-based features along with statistically significant features determined by the Mann-Whitney test, including minimum, mean, maximum, skewness, standard deviation, and kurtosis from the gait signals of healthy individuals and neurodegenerative disease patients over a five-minute period. Then, they employed the Hill Climbing feature selection method to choose an effective combination of features. Subsequently, they used probabilistic neural network (PNN) and support vector machine (SVM) classifiers, along with leave-one-out cross-validation (LOOCV), to classify the two classes of NDD diseases.

Fam [25], using a qualitative approach, transformed the time series of walking into texture images using fuzzy recurrence plots to gain a deeper intuitive understanding of disease patterns. Goshvarpour et al. [26] reconstructed the delayed Poincaré map from the gait patterns of the right and left foot. Then, using polar indices, they quantified these maps and employed a neural network classifier for binary classification. Azleglu et al. [27] conducted a multi-class classification of these diseases to diagnose PD, ALS, and HD diseases from groups of NDD patients and healthy individuals. Methods such as detrended fluctuation analysis (DFA), dynamic time warping (DTW), and autocorrelation coefficient (AC) were used to extract features from the ground reaction force signal. Finally, SVM, KNN, and neural networks were also applied to compare the performance of the classifiers. The results indicated that DFA performed better in diagnosing ALS, DTW in diagnosing PD, and AC in diagnosing HD.

Despite extensive research conducted in recent years on gait pattern analysis, a fundamental challenge remains in the need for relatively long signals for accurate analysis. However, reducing the signal recording duration has gained significant attention, since recording long signals not only causes user fatigue but can also reduce the user-friendliness of analytical methods. Therefore, it is necessary to develop an efficient method capable of accurately recognizing gait patterns within shorter time intervals.

The main objective of this study is to present an innovative approach for classifying gait patterns of individuals with neurological disorders compared to healthy individuals, based solely on a short three-second signal segment. To this end, the proposed method utilizes features from Cross Recurrence Quantification Analysis (CRQA) as an effective tool for extracting dynamic and synergistic information from the right and left foot force signals during walking.

II. METHODS

Fig. 1shows the overall framework of the paper, which includes the database and the steps involved, such as data preprocessing, extraction of CRQA and statistical features, feature selection, and classification using machine learning algorithms. In this process, the right and left foot ground reaction force (VGRF) signals were initially used as the input data. In the data preprocessing stage, all input data were thoroughly and visually examined, and those affected by external factors, which caused changes in the signal shape, were removed from the dataset. Then, the signals were divided into 100 three-second windows. In the feature extraction stage, since gait signals are nonlinear, non-stationary, quasi-periodic, and noisy, 13 features from cross-recurrence quantification analysis (CRQA), along with statistical features such as mean, variance, skewness, and kurtosis, were calculated for each input window. These features were calculated and extracted separately for both the left and right feet. Afterward, all statistical and CRQA

features were combined to create a comprehensive set. To reduce the dimensionality of the extracted features and select an effective feature combination, the sequential feature selection (SFS) algorithm was used. In this stage, the features that had the most significant impact on the model's performance were selected without reducing the model's accuracy [28, 29]. Finally, the features chosen by SFS were input into the Ensemble (Bagged Trees) machine learning model for binary and multi-class classification. Additionally, cross-validation techniques were employed to evaluate the accuracy of the classifiers and the model's performance. Each of these steps will be discussed in detail in the following sections.

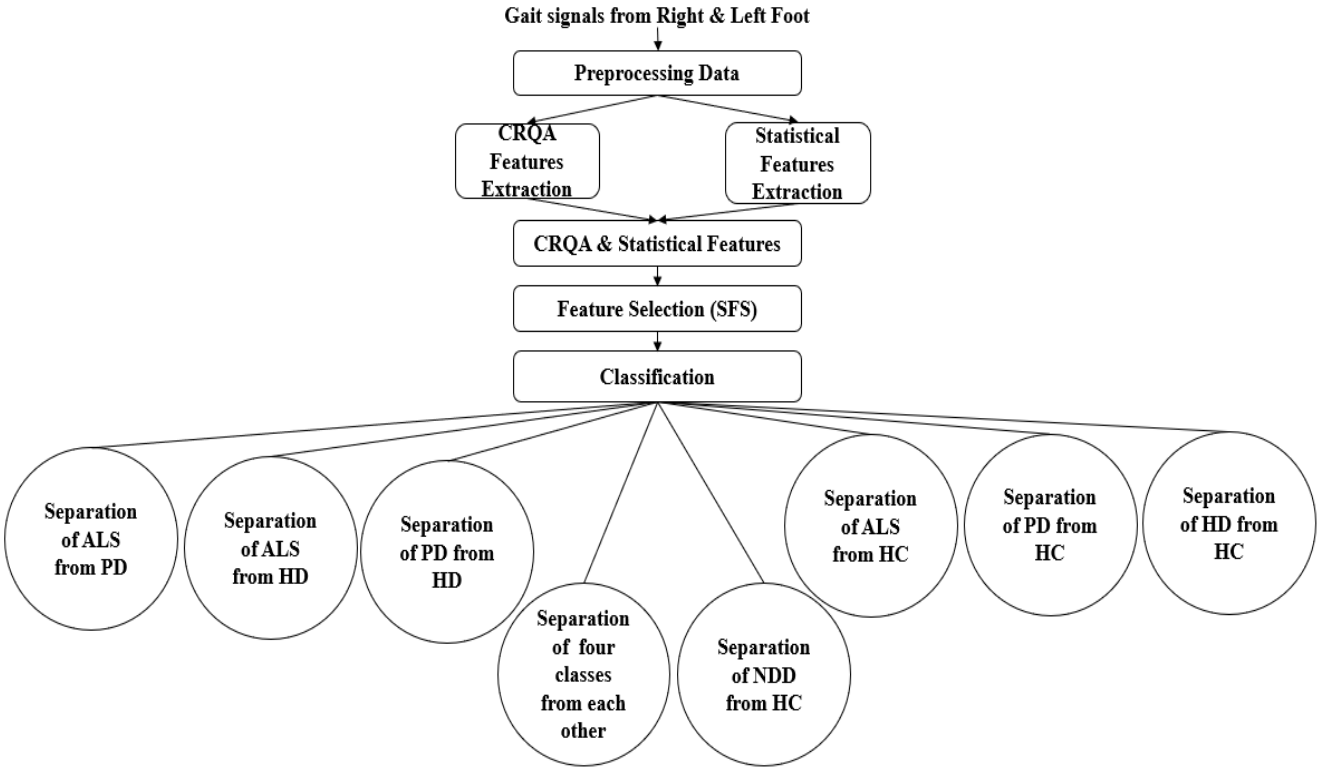

Fig. 1. Overall diagram of the proposed method

A. Database

Hasdorff and colleagues presented the VGRF database for neurodegenerative diseases used in this study [30]. The data were collected using eight sensors that separately record the distributed force under each foot. A total of 64 recordings from 48 participants were available, including 20 patients with Huntington's disease (13 females and 7 males; 29–71 years), 15 patients with Parkinson's disease (5 females and 10 males; 44–80 years), 13 patients with amyotrophic lateral sclerosis (3 females and 10 males; 36–70 years), and 16 healthy controls (14 females and 2 males; 20–74 years). On average, ALS and Parkinson's patients were older than participants in the other two groups. The healthy controls and Huntington's patients were predominantly female, whereas the Parkinson's and ALS groups were predominantly male. The database includes two types of data: force signal data and time-series features derived from the force signal. In this study, only the force signals of the left and right feet of the participants were used. The force signals were recorded for five minutes at a sampling frequency of 300 Hz. Table 1 provides a summary of the demographic information and features of the individuals, including weight (kg), number, gait speed (m/s), and age (years) of the participants.

TABLE I
Demographics of the Subjects in the Neurodegenerative Disease Database

Class	Number	Ages (Year)	Weight (kg)	Gait Speed(m/s)
PD	15	46.65 ±12.6	75.07 ± 16.9	1.0 ± 0.2
HD	20	55.62 ±12.83	73.47 ± 16.23	1.15 ± 0.35
ALS	13	66.8 ±10.85	77.11 ± 21.15	1.05 ± 0.22
HC	16	39.31 ±18.51	66.81 ± 11.08	1.35 ± 0.16

B. Data Preprocessing

In the first step, it should be noted that the number of individuals in different classes is not the same. To standardize the number of participants in each group, 20 individuals were selected to match the highest number of individuals in the HD group for the PD, HC, and ALS groups, with fewer participants. Participants were recruited from the beginning based on the shortage of people in each group until the number reached 20.

The five-minute recorded gait signals were divided into 100 three-second windows to increase the number of available samples, resulting in 900 non-overlapping samples. In a five-minute recording of individuals walking down a corridor, they were instructed to continue walking in the opposite direction after reaching the end of the path. During this process, a set of noisy signals were generated at the beginning, middle, and end of the samples. After reviewing the samples, the changes in the noisy and healthy windows were compared, and a threshold was set for these windows. If the changes in the 900 sample windows were below the threshold, the window was identified as noisy and eventually removed.

C. Feature Extraction

Considering the gait characteristics of patients with Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and the healthy control group presented in Table 2, disease diagnosis based solely on gait parameters is challenging and may lead to errors. Therefore, it is essential to investigate the hidden relationships among gait signals associated with these three neurodegenerative diseases. Such relationships, in turn, enable the automation of patient classification based on gait signals in neurodegenerative disorders [31].

TABLE II
Symptoms and Gait Characteristics in Progressive Neurological Disorders: Parkinson's Disease, Huntington's Disease, and Amyotrophic Lateral Sclerosis [31].

NDD	Symptoms	Gait Characteristics
Parkinson Disease	Hyperkinetic movement, Bradykinesia, Hypertonia, Tremor, Flexed posture, Festination, Loss of postural reflexes, and freezing	Decreased walking speed Increased cadence Reduced stride length Reduced swing time Higher double support time
Huntington Disease	Uncontrolled movements Emotional problems Psychiatric disorders Loss of thinking abilities	Decrease walking speed Decrease step/stride length Increase stance-swing phase Decrease single support time
Amyotrophic Lateral Sclerosis	Perturbations in the fluctuation dynamics, , altered gait rhythm, weakness in legs, feet, or ankles.	Decreased walking speed Increased stride time variability Increased stride time

In the early stages of neurodegenerative diseases, subtle changes in motor signals can be observed that time-domain features cannot detect. It seems that feature extraction from the force signals of both feet is a suitable choice for analyzing gait motor signals, as these signals exhibit nonlinear, non-stationary, quasi-periodic, and noisy oscillations in neurodegenerative disorders [24]. According to previous studies [23], the RQA technique has been introduced as one of the most effective methods for representing nonlinear dynamics patterns in individuals' walking. The feature extraction process used in recent papers reduces real-time limitations such as noise and non-stationarity of motor signals. In this study, cross-recurrence quantification analysis is used to take advantage of the simultaneous information from the left and right foot gait signals and calculate the interrelations between them. Therefore, the CRQA technique is proposed as an efficient approach for analyzing nonlinear and non-stationary dynamic patterns in individuals with neurodegenerative disorders. This approach can also clearly reveal synergistic dynamic relationships between the lower limbs.

To this end, 13 effective CRQA features were extracted, including recurrence rate, determinism, averaged diagonal length, length of longest diagonal length, entropy of diagonal length, laminarity, trapping time, length of longest vertical line, recurrence time of 1st type, recurrence time of 2nd type, recurrence period density entropy, clustering coefficient, and transitivity. Based on the embedding theorem, CRQA examines the cross-recurrent nature of force signals in the reconstructed phase space [32]. This theorem helps in better understanding the dynamics and gait patterns.

To determine the recurrence plot and chaotic features of CRQA, the optimal embedding dimension and time delay parameters were estimated using the false nearest neighbor algorithm and mutual information. The 100 windows were used with 900 samples, each considered an independent signal sample for extracting CRQA features. Then, to adjust the embedding dimension and time delay, the values of these parameters were determined for all individuals from different groups, and their median value was selected as the criterion [33, 34]. In total, 39 effective CRQA features were extracted from each window, calculated three times separately with different time delays, embedding dimensions, and thresholds. Additionally, statistical features were used to obtain meaningful quantitative gait parameters. The CRQA features, along with statistical features such as mean, variance, skewness, and kurtosis from the signals of both feet, were combined, resulting in a total of 47 extracted features. All these features were independently calculated from each signal (including both the left and right feet). The CRP toolbox in MATLAB was used for the visual representation of the force signal and the quantitative calculations of CRQA features. Further explanations of the features derived from CRQA and statistical features are provided below.

Cross-Recurrence Quantification Analysis Features

Gait signals are inherently nonlinear and are deeply coordinated with the brain's complex activities. CRQA is a suitable tool for processing nonlinear and dynamic data that continuously change. In this study, CRQA was used to investigate the dynamic structures of stepping by quantifying

the recurrence plots (RP). The RP visually illustrates the spatial and temporal correlations between the features related to the left and right foot VGRF and is presented as a matrix. The mathematical relationships for the recurrence plot are as follows:

$$R_{i,j}(\varepsilon) = \Theta\left(\varepsilon - \left\|\vec{x}_i - \vec{y}_j\right\|\right), i,j = 1,\dots,N \quad (1)$$

In this case, RP is the recurrence plot matrix, $\left\|\vec{x}_i - \vec{y}_j\right\|$ represents the difference between the step distances of the left and right feet, ε is the threshold value, and Θ is the Heaviside function. The 13 features derived from CRQA are as follows:

- Recurrence Rate

The recurrence rate is a measure of the density of recurrence points in the recurrence plot, calculated as the ratio of the number of recurrence points to the total number of points in the recurrence plot (N), as calculated in equation (2).

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

In this regard, the expression $\sum_{i,j=1}^{N} R_{i,j}(\varepsilon)$ represents the sum of the recurrence matrix, or in other words, the number of recurrence points, where the variable N is the length of the time series and the term N^2 represents the total number of points in the plot.

- Determinism

Determinism, or predictability, is one of the features based on the histogram of diagonal lines and is defined as the ratio of the recurrence points that form diagonal structures with a minimum length to all recurrence points, as given by equation (3).

$$DET = \frac{\sum_{l=l_{min}}^{N} l\,P(l)}{\sum_{l=1}^{N} l\,P(l)} \quad (3)$$

In this regard, l represents the length of the diagonal structure, l_{min} is the minimum length of the diagonal structures, N is the length of the time series, and P is the histogram of the diagonal line lengths. The histogram of the diagonal lines is calculated according to equation (4), where the symbol ε has been omitted for simplicity in the CRQA measurements.

$$P(\varepsilon,l) = -\sum_{i,j=1}^{N}(1 - R_{i-1,j-1}(\varepsilon))(1 - R_{i+1,j+1}(\varepsilon))\prod_{k=0}^{l-1} R_{i+k,j+k}(\varepsilon)$$

$$P(\varepsilon,l) = P(l) \quad (4)$$

- Averaged Diagonal Length

The averaged diagonal length is obtained based on the histogram of diagonal lines. The averaged diagonal length represents the average time two path sections are close to each other and can indicate prediction time. This feature is obtained by equation (5):

$$L = \frac{\sum_{l=l_{min}}^{N} l\,P(l)}{\sum_{l=l_{min}}^{N} P(l)} \quad (5)$$

- Length of the longest diagonal length

The length of the longest diagonal line and its inverse are recognized as divergence features related to the diagonal lines, which are obtained by equations (6) and (7).

$$L_{max} = max(\{l_i; i = 1,\dots,N_l\}) \quad (6)$$

$$DIV = \frac{1}{L_{max}} \quad (7)$$

In these relations, the parameter N_l represents the total number of diagonal lines, which is calculated by equation (8):

$$N_l = \sum_{l \geq l_{min}} P(l) \quad (8)$$

- Entropy of Diagonal Line

The entropy of the diagonal line measures the complexity and randomness of the diagonal lines in the recurrence plot, indicating nonlinear behavior and irregularity in the system. It is calculated by equation (9).

$$ENT = -\sum_{l=line}^{N} p(l)\,ln\left(p(l)\right) \quad (9)$$

In this relation, $p(l)$ is the probability of finding a diagonal line of a specific length l, which is calculated by equation (10).

$$p(l) = \frac{P(l)}{N_l} \quad (10)$$

- Laminarity

The laminarity feature indicates the occurrence of layered states in the system without describing the duration of these layered stages. This feature is calculated as the ratio of the recurrence points that form vertical structures to the total set of recurrence points, as shown in equation (11).

$$LAM = \frac{\sum_{v=v_{min}}^{N} v\,P(v)}{\sum_{v=1}^{N} v\,P(v)} \quad (11)$$

In this regard, v is the length of the vertical structure, v_{min} is the minimum length of the vertical structures, N is the length of the time series, and P is the histogram of the vertical structure lengths. The total number of vertical lines with length v in the recurrence plot is calculated as the histogram of the vertical lines, according to equation (12).

$$P(v) = -\sum_{i,j=1}^{N}(1 - R_{i,j-1})(1 - R_{i,j+v})\prod_{k=0}^{v-1} R_{i,j+k} \quad (12)$$

- Trapping Time

Trapping time refers to the duration for which the system remains in a specific repetitive state. This feature is calculated by considering the minimum length of the vertical structures and can assist in identifying the steady-state conditions of the system. It is calculated by equation (13).

$$TT = \frac{\sum_{v=v_{min}}^{N} v\,P(v)}{\sum_{v=v_{min}}^{N} P(v)} \quad (13)$$

- Length of Longest Vertical Line

The length of the longest vertical line, similar to the length of the longest diagonal line, is obtained using equation (14). The length of the longest vertical line in the recurrence plot indicates the maximum period of invariance in the system.

$$V_{max} = max\left(\{v_i\}_{i=1}^{N_v}\right) \quad (14)$$

- Recurrence time of 1st type

The time or number of steps for the first recurrence in the data, which indicates the time when the repetition of patterns begins. This feature is obtained according to equation (15).

$$\left\{T_k^{(1)} = j_{k+1} - j_k\right\}_{k\in N} \quad (15)$$

Due to possible tangential movements, some recurrence plots in R_i may be associated with first recurrence times $T_k^{(1)} = 1$. To obtain the true recurrence points, all consecutive recurrence points with $T_k^{(1)} = 1$ are removed from the set R_i, and the new set is obtained according to equation (16):

$$R_i^{'} = \left\{\vec{x}_{j^{'}1}, \vec{x}_{j^{'}1} \ldots\right\} \quad (16)$$

- Recurrence time of the 2nd type

The recurrence time of the 2nd type refers to the time interval between the start of consecutive vertical recurrence structures in the recurrence plot. It is obtained through equation (17) or equation (18).

$$\{Tk(2) = jk + 1' - jk'\}k \in N \quad (17)$$

$$\left\{T_k^{(2)} = j_{k+1}^{'} - j_k^{'}\right\}_{k\in N} \quad (18)$$

- Recurrence period density entropy

The entropy of the recurrence period density represents the normalized entropy of the distribution of recurrence times in the time series, indicating the complexity level of a system. This parameter is calculated by equation (19).

$$H_{norm} = \frac{-\sum_{i=1}^{T_{max}} P(i)\, ln\, P(i)}{lnT_{max}} \quad (19)$$

In this context, T_{max} is the maximum recurrence time, and P is the probability density of recurrence times, which is calculated by equation (20).

$$P(T) = \frac{R(T)}{\sum_{i=1}^{T_{max}} R(i)} \quad (20)$$

- Clustering Coefficient

The clustering coefficient is a feature that describes the complex recurrence network. This network consists of nodes and edges, where the nodes represent the state-space vectors, and the edges represent the recurrences.

$$R_{ij} = R_{ji} = \begin{cases} 1 & (i,j) \in E \\ 0 & (i,j) \notin E \end{cases} \quad (21)$$

Here, R is the adjacency matrix, i and j are the nodes of the set V of the network vertices, and E is the set of edges of the network. The clustering coefficient represents the likelihood of adjacency between any two occurrences of each state and is calculated using equation (22).

$$C = \sum_{i=1}^{N} \frac{\sum_{j,k=1}^{N} R_{i,j}^{m,\varepsilon} R_{j,k}^{m,\varepsilon} R_{k,i}^{m,\varepsilon}}{RR_i} \quad (22)$$

- Transitivity

Transitivity is the probability that two neighbors of a given state are also neighbors of each other, and the relationship among three nodes is expressed in equation (23).

$$A_{ij} = A_{ik} = 1, A_{jk} = 1 \quad (23)$$

Here, i , j and k are nodes from the set V of vertices. The transitivity feature is calculated using equation (24).

$$C = \frac{\sum_{i,j,k=1}^{N} R_{i,j}^{m,\varepsilon} R_{j,k}^{m,\varepsilon} R_{k,i}^{m,\varepsilon}}{\sum_{i,j,k=1}^{N} R_{i,j}^{m,\varepsilon} R_{k,i}^{m,\varepsilon}} \quad (24)$$

Statistical Features

The computed statistical features include characteristics such as mean, variance, kurtosis, and skewness of the VGRF signals from the left and right foot, measured during only three seconds of walking in the shortest possible duration.

- Mean

The mean is defined as the average value of the stepping signal. It is calculated by summing the signal values x_i and dividing by the total number of samples N:

$$\mu = \frac{\sum_{i=1}^{N} x_i}{N} \quad (25)$$

- Variance

Variance is a squared measure of the standard deviation. It is calculated based on the sum of the squared differences between each element x_i and the mean μ, divided by the total number of samples. Variance is computed using equation (26):

$$Var = \frac{\sum_{i=1}^{N} (x_i - \mu)^2}{N - 1} \quad (26)$$

- Skewness

Skewness is used as a measure of asymmetry in the amplitude distribution relative to the mean and can be calculated as follows:

$$Sk = \frac{\frac{1}{N}\sum_{i=1}^{N} (Ai - \mu)^2}{Var^{\frac{3}{2}}} \quad (27)$$

- Kurtosis

Kurtosis measures the tendency of a signal's distribution to produce outlier values and is calculated as:

$$Ku = \frac{\frac{1}{N}\sum_{i=1}^{N} (Ai - \mu)^4}{Var^2} \quad (28)$$

The above statistical features were independently extracted from each signal sample for both the left and right foot. Finally, in this study, gait parameters were analyzed by combining effective CRQA and statistical features derived from a short 3-second gait period across both pathological and healthy individuals.

D. Feature Selection

The Sequential Feature Selection (SFS) algorithm is a feature search technique used to create a subset for model design and reduce the dimensionality of the original features. This technique is employed to minimize the mean squared error and maximize the accuracy percentage. Let S be the set of all features, N the number of features to be selected, S_N the set of selected features, and $acc(S_N)$ the classification accuracy for this set. Initially, S_N is considered an empty set. Then, each feature is evaluated by the classifier individually, and the feature with the highest accuracy is added to S_N. In

the next step, pairwise combinations of features, including the first selected feature and the other features, are examined to identify the best combination, which is then chosen as the new S_N. Subsequently, if adding a new feature improves performance, it is added to the set; otherwise, the weakest feature is removed [35-40]. This process continues until the desired number of features is reached. As an efficient tool in machine learning and analysis of complex data, the SFS algorithm plays a significant role in optimizing models (reducing classifier computations), improving prediction accuracy, and selecting the most effective features [32, 33].

E. Classification

The Ensemble method using Bagged Trees was employed as a powerful and effective approach for classifying. This method is considered suitable for this research due to its outstanding ability to analyze and classify data, as well as its capacity to distinguish between different groups. The use of the Ensemble Bagged Trees technique enables the modeling of complex nonlinear relationships, effectively extending the class separation margins and providing significant resistance to noisy data.

F. Cross Validation

A 10-fold cross-validation technique was used to evaluate the classifiers' performance. In this technique, the data is randomly divided into 10 parts, and the classifiers are iteratively evaluated on the data of each part. In each cross-validation iteration, nine parts of the data are used for training the model, and the remaining part is used as the test data to assess the model's performance. Finally, the results from the 10 iterations are combined to provide a final evaluation of the classifier's performance. Cross-validation prevents errors caused by random data splitting and overfitting, enabling a more accurate evaluation of the classifiers.

G. Evaluation Metric

The accuracy metric was used to evaluate the performance of the model. Accuracy is calculated as the ratio of the number of correct predictions to the total number of samples. This metric widely facilitates the simulation and evaluation of the model's performance under various conditions. Accuracy is computed using the numbers from the confusion matrix according to equation (29), where TP, TN, FP, and FN are defined as follows:

TP (True Positives): The number of samples correctly identified as positive.

TN (True Negatives): The number of samples correctly identified as negative.

FP (False Positives): The number of samples incorrectly identified as positive.

FN (False Negatives): The number of samples incorrectly identified as negative.

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{29}$$

III. Results

This study provides valuable information regarding the effectiveness of combining the relevant CRQA features and statistical measures that describe the changes in nonlinear and non-stationary gait signals for the quantitative and qualitative assessment of NDD patients. Therefore, by implementing CRQA on the synchronized left and right foot gait signals of individuals, different groups are classified. Cross-recurrence plots derived from the CRQA technique are used to represent the texture patterns of physiological gait signals visually, and their parameters are measured through quantitative cross-recurrence analysis. These plots contain structures such as diagonal lines, vertical lines, and isolated points. Fig. 2 shows the distinction between different groups of NDDs using these plots. The texture patterns in the recurrence plots of healthy individuals are notably regular and repetitive. In this group, all patterns appear repeatedly and with a specific order over three seconds, indicating coordination and stability in the stance and swing intervals of the main signal. These features reflect the specific gait cycle characteristics of healthy individuals, where each stance and swing interval repeats regularly and consistently. In contrast, the texture patterns in the recurrence plots of individuals with ALS are significantly more complex. In these plots, two types of black squares are observed, one larger and the other smaller. These squares represent the stance phase and swing phase, respectively. It is noteworthy that the size of the larger squares is larger compared to the healthy and other patient groups, indicating increased stance intervals in individuals with ALS. The recurrence plots of individuals with PD show more irregular features. Specifically, the black squares in these plots are irregular and scattered, clearly indicating instability in step and swing intervals. These irregular changes are notably more apparent when compared to healthy individuals. Finally, the recurrence plots of individuals with HD have the most irregular yet distinct features among other neurological disorders. In this group, the size of the two types of squares is almost identical, indicating no significant difference between the stance and swing intervals in these individuals. These features represent a more complex and irregular pattern compared to other patient groups. These analyses demonstrate the effectiveness of the CRQA method in describing the nonlinear and non-stationary dynamic changes in walking and distinguishing between different groups of neurodegenerative diseases.

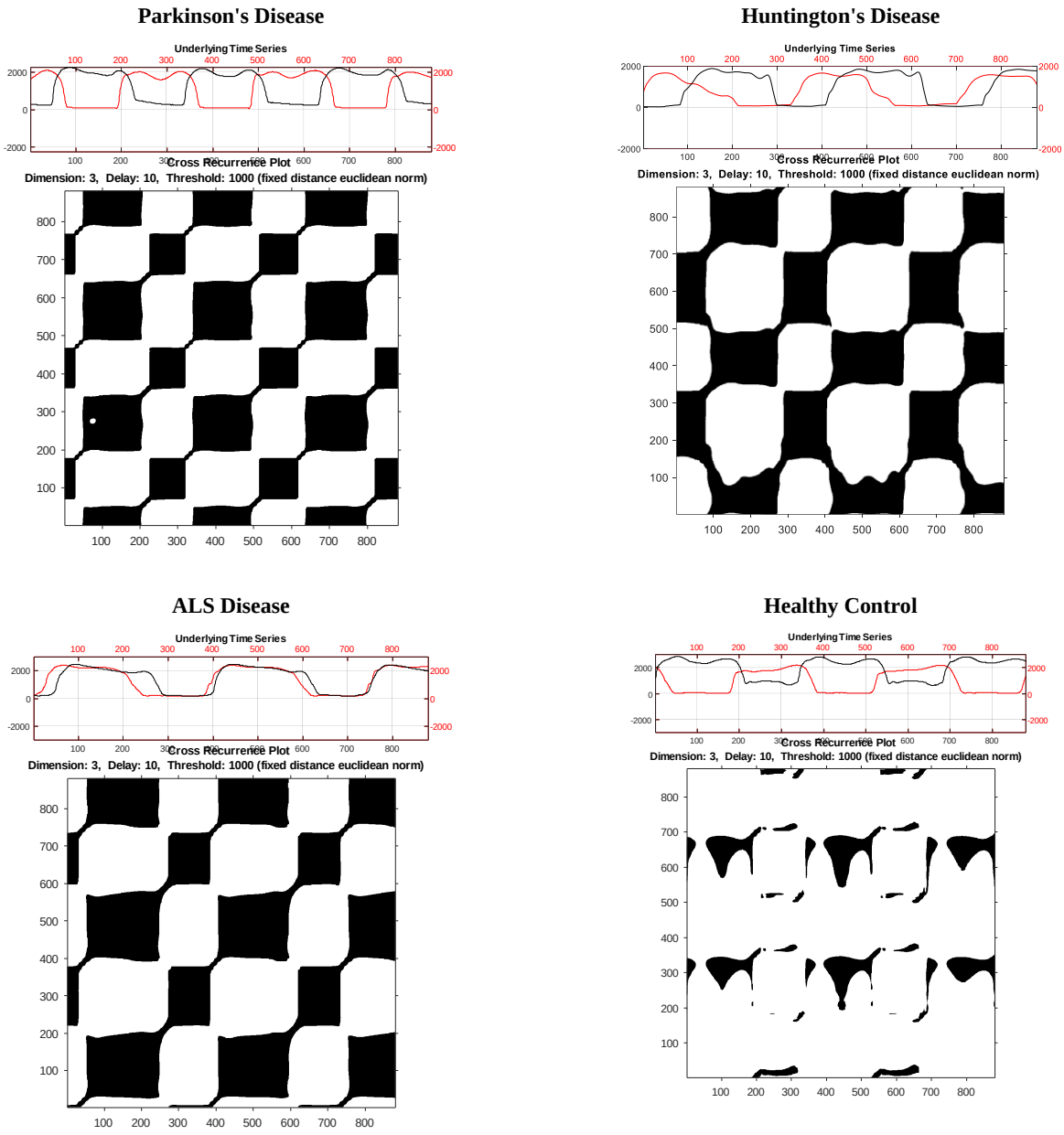

Fig. 2. Quantitative cross recurrence analysis of four classes of neurodegenerative diseases and healthy individuals. The ambiguous sections of the RP diagrams indicate the presence of more frequent data points. The vertical lines represent the stability of the system's state and its lack of change in the current state for a short period.

TABLE III

Comparison of the Accuracy Percentage Between the Suggested Method and the Recent Literature

Classification Task	[21]	[26]	[27]	Proposed Algorithm
HC vs PD	94.96	91.93	-	98.3
HC vs HD	97.33	92.92	-	94.8
HC vs ALS	97.43	91.13	-	98.7
PD vs HD	-	-	-	95
PD vs ALS	-	-	-	98.2
ALS vs HD	-	-	-	98.4
NDD vs HC	96.42	-	-	94
PD vs HD vs ALS vs HC	-	-	93.5	93.5

This study aims to classify short-time two-class and four-class data using Ensemble (Bagged Trees) with ground reaction force signals as the input. In machine learning, one of the main challenges is creating extensive changes in the gait signal datasets to provide the best training set for the algorithms. These changes include signal enhancement and feature extraction to optimize classifier performance. Initially, a set of noisy signals located at the beginning, middle, and end of the samples was generated. Then, the noise was removed using the aforementioned preprocessing methods, and the enhanced signals were prepared for subsequent analyses. The feature extraction phase is one of the most crucial steps in improving classifier performance, as effective and meaningful features can provide a better representation of signal patterns and increase classifier accuracy. In this study, effective statistical and CRQA features were extracted from three seconds of gait force signals of both legs from the patient groups and healthy individuals. The SFS algorithm was used to select the optimal features. A set of features was selected, including recurrence rate, entropy, average diagonal length, trapping time, recurrence period density entropy, recurrence time of 1st type, transitivity from CRQA, skewness, and kurtosis from statistical features. Ultimately, 17 features out of 47 features were chosen. In this study, eight binary classifications were considered, including ALS vs. PD, ALS vs. HD, PD vs. HD, ALS vs. HC, PD vs. HC, HD vs. HC, NDD vs. HC, and the four-class classification ALS vs. PD vs. HD vs. HC. The results obtained are presented in Table 3. After evaluation with various classifiers, the best accuracy for disease and control groups, such as HC vs. ALS, HC vs. PD, and HC vs. HD, were 98.7%, 98.3%, and 94.8%, respectively, using Ensemble (Bagged Trees). The accuracy for two disease groups, such as HD vs. PD, PD vs. ALS, and HD vs. ALS were 95%, 98.2%, and 98.4%, respectively. The accuracy of separating disease classes from the healthy group (NDD vs. HC) was 94%. Up to this point, all separations were performed in a binary classification. In the four-class classification based on ALS vs. PD vs. HD vs. HC, an accuracy of 93.5% was achieved. These results demonstrate that the proposed algorithm, based on the combination of CRQA and statistical features of the three-second time window, effectively improves the performance of Ensemble (Bagged Trees) in classifying different groups of neurodegenerative patients and healthy individuals.

TABLE IV
Four-class Classification Accuracy of the Ensemble Classifier Using CRQA Features, Statistical Features, and their Combination

Feature Set	Accuracy of the four-class classification
CRQA Features	88.74
Statistical Features	83.69
CRQA Features + Statistical Features	93.55

Table IV presents the four-class classification accuracy for each feature set to compare the effectiveness of CRQA features, statistical features, and their combination. The results indicate a complementary effect of the two feature sets for classifying neurodegenerative diseases, such that their combination achieves the highest classification accuracy.

IV. Discussion and Conclusion

Gait analysis and extracting relevant features are recognized as primary tools for assessing the status of neurodegenerative diseases and supporting therapeutic processes. In this study, fluctuations and dynamic distinctions in the gait of healthy and sick individuals were extracted from just three seconds of ground reaction force signals. One of the key points of this research is the simultaneous use of information from both the left and right feet, which leads to information synergy and improves accuracy in identifying different states. Feature extraction from the force data represents the synergistic dynamic relationship between the left and right feet and more accurately reflects the complex interactions among the lower limbs. Additionally, it has a positive effect on identifying different patterns and is capable of detecting characteristic changes between different disease groups. The Ensemble (Bagged Trees) model, along with 10-fold cross-validation, was used to classify various groups to improve the classification system's performance and achieve higher accuracy.

In previous studies [18–27], the recording duration of signals was five minutes, and the long recording time represented a practical limitation. In contrast, the present study considered a much shorter recording duration of 3 seconds. In study [26], using five-minute data, the average accuracy for three disease classes versus healthy subjects reached 91.9%, whereas the present study, using three-second data, achieved an average accuracy of 97.26%. Furthermore, the results of the study [21], which utilized both force signals and time-series features over a five-minute walking period, showed lower accuracy in distinguishing Parkinson’s disease from healthy subjects, as well as ALS from healthy subjects, compared to the proposed method.

To further strengthen the clinical applicability of the proposed method, it is recommended to evaluate its generalizability across diverse walking environments, such as inclined, uneven, or obstacle-laden surfaces. Numerous studies have shown that gait parameters—including walking speed, step length, symmetry, and plantar pressure distribution—are influenced by environmental conditions and surface type. For instance, study [41] demonstrated that walking speed, step length, and gait symmetry vary significantly in outdoor paths, such as grass or sidewalks, compared to flat indoor surfaces. Additionally, another study [42], using pressure-sensing insoles, reported that inclined, flexible, or uneven surfaces substantially affect vertical ground reaction forces and center of pressure location. These findings highlight the limitations of studies conducted solely on flat and straight paths and indicate that gait assessments should be performed under varied environmental conditions for true generalizability and clinical relevance. Expanding

testing in this manner, particularly for daily-life scenarios such as walking at home or on urban pathways, is expected to enhance the accuracy, stability, and reliability of gait analysis and recognition models.

This research aims to provide an efficient computational approach for identifying the dynamics of gait in healthy individuals compared to various conditions of patients with neurodegenerative disorders by extracting precise features and using a short time window. The advantages of using a short time in medical processes significantly impact the efficiency and comfort of both users and specialists. Short time not only prevents user fatigue but also helps doctors to engage in the diagnosis and treatment process quickly and more accurately. Additionally, using a short time allows researchers to use sensors like Kinect for data recording. For example, when a patient enters a doctor's office, the gait signals of the patient are recorded by Kinect sensors during a 3 to 4-second movement (depending on the room size) toward the doctor. Then, the data is analyzed using the proposed method, and the type and condition of the disease are diagnosed. Non-invasive and low-cost sensors like Kinect are considered effective tools for diagnosing neurodegenerative disorders. The results obtained from this research successfully develop a novel approach for gait analysis in NDD patients using diverse feature extraction and machine learning-based classification algorithms.

REFERENCES

[1] F. Imam et al., "The Global Neurodegeneration Proteomics Consortium: biomarker and drug target discovery for common neurodegenerative diseases and aging," Nature Medicine, pp. 1-11, 2025.

[2] A. Maleki and N. Monfared Jahromi, "Classification of Neurodegenerative Diseases Using Recurrence Quantification Analysis of Gait Signal," Iranian Journal of Biomedical Engineering, vol. 18, no. 3, pp. 221-230, 2024.

[3] X. Du, R. Vasudevan, and M. Johnson-Roberson, "Bio-lstm: A biomechanically inspired recurrent neural network for 3-d pedestrian pose and gait prediction," IEEE Robotics and Automation Letters, vol. 4, no. 2, pp. 1501-1508, 2019.

[4] S. Campbell, K. Dholakia, L. Sturgill, and R. Wellmon, "Preparing students for the national physical therapist examination (NPTE): One program's experience with using the practice exam and assessment tool (PEAT) as a graduation requirement," Internet Journal of Allied Health Sciences and Practice, vol. 21, no. 4, p. 22, 2023.

[5] G. G. Kovacs, "Concepts and classification of neurodegenerative diseases," in Handbook of clinical neurology, vol. 145: Elsevier, 2018, pp. 301-307.

[6] R. Baker, "Measuring walking: a handbook of clinical gait analysis," (No Title), 2013.

[7] M. Proudfoot et al., "Altered cortical beta-band oscillations reflect motor system degeneration in amyotrophic lateral sclerosis," Human brain mapping, vol. 38, no. 1, pp. 237-254, 2017.

[8] J. Hoff, A. v/d Plas, E. Wagemans, and J. Van Hilten, "Accelerometric assessment of levodopa-induced dyskinesias in Parkinson's disease," Movement disorders: official journal of the Movement Disorder Society, vol. 16, no. 1, pp. 58-61, 2001.

[9] M. Banaie, Y. Sarbaz, S. Gharibzadeh, and F. Towhidkhah, "Huntington's disease: modeling the gait disorder and proposing novel treatments," Journal of Theoretical Biology, vol. 254, no. 2, pp. 361-367, 2008.

[10] N. Rana, L. Kapil, C. Singh, and A. Singh, "Modeling Huntington's disease: An insight on in-vitro and in-vivo models," Behavioural Brain Research, vol. 459, p. 114757, 2024.

[11] J. Chaki and M. Woźniak, "Deep learning for neurodegenerative disorder (2016 to 2022): A systematic review," Biomedical Signal Processing and Control, vol. 80, p. 104223, 2023.

[12] B. M. Kelser, E. M. Teichner, R. C. Subtirelu, and K. N. Hoss, "A review of proposed mechanisms for neurodegenerative disease," Frontiers in Aging Neuroscience, vol. 16, p. 1370580, 2024.

[13] P. Rizek, N. Kumar, and M. S. Jog, "An update on the diagnosis and treatment of Parkinson disease," Cmaj, vol. 188, no. 16, pp. 1157-1165, 2016.

[14] M. Kogan, M. McGuire, and J. Riley, "Deep brain stimulation for Parkinson disease," Neurosurgery Clinics, vol. 30, no. 2, pp. 137-146, 2019.

[15] E. M. Coppen and R. A. Roos, "Current pharmacological approaches to reduce chorea in Huntington's disease," Drugs, vol. 77, pp. 29-46, 2017.

[16] S. Martin, A. Al Khleifat, and A. Al-Chalabi, "What causes amyotrophic lateral sclerosis?" F1000Research, vol. 6, p. 371, 2017.

[17] D. G. Gadhave et al., "Neurodegenerative disorders: Mechanisms of degeneration and therapeutic approaches with their clinical relevance," Ageing research reviews, p. 102357, 2024.

[18] Faisal, M. A. A., et al. (2023). "NDDNet: A deep learning model for predicting neurodegenerative diseases from gait pattern." Applied Intelligence 53(17): 20034-20046.

[19] Zhou, Z., et al. (2023). Deep learning-based classification of neurodegenerative diseases using gait dataset: A comparative study. Proceedings of the 2023 International Conference on Robotics, Control, and Vision Engineering.

[20] Visvanathan, P. and P. D. R. Vincent (2024). "Prediction of Gait Neurodegenerative Diseases by Variational Mode Decomposition Using Machine Learning Algorithms." Applied Artificial Intelligence 38(1): 2389375.

[21] A. Zhao, L. Qi, J. Dong, and H. Yu, "Dual channel LSTM based multi-feature extraction in gait for diagnosis of Neurodegenerative diseases," Knowledge-Based Systems, vol. 145, pp. 91-97, 2018.

[22] N. Marwan, C. L. Webber, E. E. Macau, and R. L. Viana, "Introduction to focus issue: Recurrence quantification analysis for understanding complex systems," Chaos: An Interdisciplinary Journal of Nonlinear Science, vol. 28, no. 8, 2018.

[23] P. Prabhu and N. Pradhan, "Recurrence quantification analysis of human gait in neurological movement disorders," Int. J. Eng. Res., vol. 5, no. 03, pp. 1-6, 2016.

[24] P. Prabhu, A. K. Karunakar, H. Anitha, and N. Pradhan, "Classification of gait signals into different neurodegenerative diseases using statistical analysis and recurrence quantification analysis," Pattern Recognition Letters, vol. 139, pp. 10-16, 2020.

[25] T. D. Pham, "Texture classification and visualization of time series of gait dynamics in patients with neuro-degenerative diseases," IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 26, no. 1, pp. 188-196, 2017.

[26] A. Goshvarpour and A. Goshvarpour, "Differentiation of Neurodegenerative Diseases by Dynamic Analysis of Gait Pattern and Feature-level Fusion Approaches," Journal of Paramedical Sciences & Rehabilitation, vol. 10, no. 1, pp. 31-45, 2021, doi: 10.22038/jpsr.2021.51058.2149.

[27] I. G. Ozeloglu and E. A. Aydin, "Combining features on vertical ground reaction force signal analysis for multiclass diagnosing neurodegenerative diseases," International Journal of Medical Informatics, vol. 191, p. 105542, 2024.

[28] C. L. Webber and N. Marwan, "Recurrence quantification analysis," Theory and best practices, vol. 426, 2015.

[29] N. Marwan, "How to avoid potential pitfalls in recurrence plot-based data analysis," International Journal of Bifurcation and Chaos, vol. 21, no. 04, pp. 1003-1017, 2011.

[30] J. M. Hausdorff, A. Lertratanakul, M. E. Cudkowicz, A. L. Peterson, D. Kaliton, and A. L. Goldberger, "Dynamic markers of altered gait rhythm in amyotrophic lateral sclerosis," Journal of Applied Physiology, 2000.

[31] G. Cicirelli, D. Impedovo, V. Dentamaro, R. Marani, G. Pirlo, and T. R. D'Orazio, "Human gait analysis in neurodegenerative diseases: A review," IEEE journal of biomedical and health informatics, vol. 26, no. 1, pp. 229-242, 2021.

[32] D. P. Muni, N. R. Pal, and J. Das, "Genetic programming for simultaneous feature selection and classifier design," IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), vol. 36, no. 1, pp. 106-117, 2006.
[33] C. L. Webber and N. Marwan, "Recurrence quantification analysis," Theory and best practices, vol. 426, 2015.
[34] N. Marwan, "How to avoid potential pitfalls in recurrence plot-based data analysis," International Journal of Bifurcation and Chaos, vol. 21, no. 04, pp. 1003-1017, 2011.
[35] D. P. Muni, N. R. Pal, and J. Das, "Genetic programming for simultaneous feature selection and classifier design," IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), vol. 36, no. 1, pp. 106-117, 2006.
[36] M. Last, A. Kandel, and O. Maimon, "Information-theoretic algorithm for feature selection," Pattern Recognition Letters, vol. 22, no. 6-7, pp. 799-811, 2001.
[37] S. Nakariyakul and D. P. Casasent, "An improvement on floating search algorithms for feature subset selection," Pattern Recognition, vol. 42, no. 9, pp. 1932-1940, 2009.
[38] J. Schenk, M. Kaiser, and G. Rigoll, "Selecting features in on-line handwritten whiteboard note recognition: SFS or SFFS? " in 2009 10th international conference on document analysis and recognition, 2009: IEEE, pp. 1251-1254.
[39] A. Marcano-Cedeño, J. Quintanilla-Domínguez, M. Cortina-Januchs, and D. Andina, "Feature selection using sequential forward selection and classification applying artificial metaplasticity neural network," in IECON 2010-36th annual conference on IEEE industrial electronics society, 2010: IEEE, pp. 2845-2850.
[40] L. Burrell, O. Smart, G. K. Georgoulas, E. D. Marsh, and G. J. Vachtsevanos, "Evaluation of Feature Selection Techniques for Analysis of Functional MRI and EEG," in DMIN, 2007, pp. 256-262.
[41] L. Brognara, A. Arceri, M. Zironi, F. Traina, C. Faldini, and A. Mazzotti, "Gait Spatio-Temporal Parameters Vary Significantly Between Indoor, Outdoor and Different Surfaces," Sensors, vol. 25, no. 5, p. 1314, 2025.
[42] E. Warmerdam, L.-M. Burger, D. F., Mergen, M., Orth, T., Pohlemann, and B. Ganse, "The walking surface influences vertical ground reaction force and center of pressure data obtained with pressure-sensing insoles," Frontiers in Digital Health, vol. 6, p. 1476335, 2024.