Journal of Informatics and Web Engineering
Vol. 3 No. 1 (February 2024) eISSN: 2821-370X

# Vision-Based Gait Analysis for Neurodegenerative Disorders Detection

**Vincent Wei Sheng Tan[1], Wei Xiang Ooi[1], Yi Fan Tan[1], Tee Connie[1*], Michael Kah Ong Goh[1]**
[1]Faculty of Information Science and Technology, Multimedia University, Jalan Ayer Keroh Lama, 75450, Melaka, Malaysia
**Corresponding author: (tee.connie@mmu.edu.my, ORCiD: 0000-0002-0901-3831)*

*Abstract* - Parkinson's Disease (PD) is a debilitating neurodegenerative disorder that affects a significant portion of aging population. Early detection of PD symptoms is crucial to prevent the progression of the disease. Research has revealed that gait attributes can provide valuable insights into PD symptoms. The gait acquisition techniques used in current research can be broadly divided into two categories: vision-based and sensor-based. The markerless vision-based classification model has become a prominent research trend due to its simplicity, low cost and patient comfort. In this study, we propose a novel markerless vision-based approach to obtain gait features from participants' gait videos. A dataset containing gait videos from normal subjects and PD patients were collected, along with a control group of 25 healthy adults. The participants were requested to perform a Timed Up and Go (TUG) test, during which their walking sequences were recorded using two smartphones positioned at different angles, namely side and front. A multi-person pose estimator is used to estimate human skeletal joint points from the collected gait videos. Different gait features associated with PD including stride length, number of steps taken during turn, turning duration, speed and cadence are derived from these key point information to perform PD detection. Experimental results show that the proposed solution achieves an accuracy of 89.39%. The study's findings demonstrate the potential of markerless vision-based gait acquisition techniques for early detection of PD symptoms.

*Keywords— Parkinson's disease detection, gait analysis, timed up and go test, computer vision, pose estimation.*

Received: 06 July 2023; Accepted: 18 September 2023; Published: 16 February 2024

## I. INTRODUCTION

Parkinson's disease (PD) is a degenerative brain disorder characterized by motor symptoms such as bradykinesia, staggering, rigidity, walking difficulties, and imbalance [1]. It also presents various non-motor complications including cognitive impairment, mental issues, sleep disturbances, and pain disorders. Movement disorders like involuntary movements and dystonia can further limit speech and movement, leading to high levels of disability. Dementia often develops in individuals with PD. While it is the most common movement disorder, other disorders like multisystem atrophy, progressive supranuclear palsy, chorea, ataxia, and dystonia exist. These disorders share similar symptoms with PD and face similar challenges in terms of diagnosis and treatment access, particularly in low-income and middle-income countries.

The onset of PD is associated with several risk factors, including advancing age of individuals, albeit afflicting young adults as well. When compared to women, men are more likely to develop Parkinson's disease. The involvement of environmental determinants, such as noxious fumes, atmospheric contamination, and industrial solvents, has been

Journal of Informatics and Web Engineering
https://doi.org/10.33093/jiwe.2024.3.1.9

Published by MMU Press. URL: https://journals.mmupress.com/jiwe

posited as contributing to an elevated susceptibility to PD [2]. The precise cause of PD is unknown but is believed to result from a complex interaction between genetic and environmental factors, including lifelong exposure to toxic substances.

Dr. Parkinson, a London surgeon, provided the initial comprehensive clinical account of what is now recognized as PD. This condition, previously termed tremulous paralysis, can impact individuals irrespective of their race or gender [3]. Early detection is critical to mitigate disease progression, and artificial intelligence-based approaches, particularly machine learning, have been developed to enhance recognition and diagnosis. Machine learning techniques provide a compelling avenue for delivering rapid and precise diagnostic outcomes, with the potential to revolutionize clinical decision-making and the diagnostic process.

For this reason, we aim to develop a novel markerless vision-based classification model that can accurately detect PD by analyzing human gait features. To achieve this, participants were recruited and requested to perform a Timed Up and Go (TUG) test, a commonly used assessment to assess mobility and balance in individuals with PD. During the TUG test, the participants' walking sequences were recorded using two smartphones placed at different angles. A multi-person pose estimator was then used to identify and track the locations of key skeletal joint points in human bodies from the recorded videos. Various gait features associated with PD were extracted from these estimated skeletal joint points. These features include stride length (the distance covered during each step), the number of steps taken during turns, turning duration, walking speed, and cadence (the number of steps per minute). These gait features are considered important indicators of PD-related abnormalities in motor function and are used as inputs for the PD detection process. By training the model on a dataset that includes gait videos from both healthy individuals and PD patients, it learns to recognize patterns and identify characteristic differences in gait between the two groups.

## II. RELATED WORK

### *A. Conventional machine learning methods*

Several studies had used machine learning techniques to analyse gait data for the detection and classification of PD. In one study [4], a random forest algorithm achieved 93.33% accuracy in stride length prediction. Another study [5] applied logistic regression on acceleration signals to detect freezing of gait (FoG) in PD patients, achieving a classification accuracy of 81.3%.

Frequency analysis of gait signals has also been explored for the detection of neurodegenerative diseases. One study [6] applied artificial neural networks (ANN), support vector machines (SVM), and Naïve Bayes classifiers on gait data from patients with amyotrophic lateral sclerosis, Huntington's disease, PD, and healthy control subjects. The ANN classifier achieved the highest accuracy of 90.6%, followed by SVM (64.00%) and Naïve Bayes (44.44%).

Vision-based and sensor-based models have been used for gait data acquisition in PD. A previous study [7] applied principal component analysis (PCA) and linear discriminant analysis (LDA) to videos of PD patients' gait, achieving 95.49% accuracy with LDA-MDC, outperforming PCA-MDC. Another study [8] implemented SVM with a radial basis function kernel to classify PD, ALS, and Huntington's disease using force-sensitive resistors, achieving an accuracy of 83.33%.

Vertical ground reaction force (VGRF) data has been utilized to distinguish PD patients from healthy controls. One study [9] compared SVM algorithms on VGRF data, with SVM achieving the highest accuracy of 91.6% using a linear kernel.

In an investigation using the Microsoft Kinect V2 vision system, gait data was collected from PD patients and healthy subjects. ANN was applied for classification, with ANN, achieving accuracy rates of 89.4%, respectively [10]. Another study [11] compared SVM algorithms for classifying PD patients from healthy subjects based on sensor-based gait data, with SVM achieving the highest average accuracy of 86%.

Overall, the potential of machine learning in gait data analysis for PD detection and classification was demonstrated in these studies. These methods have shown promising results in accurately distinguishing PD patients from healthy controls and detecting specific gait abnormalities associated with the disease. However, more research and validation are necessary to confirm their effectiveness in real clinical scenarios. A summary of the existing conventional machine learning approaches is presented in Table 1.

Table 1. Summary of the existing conventional machine learning approaches.

| Authors | Classification technique | Dataset | Accuracy |
|---|---|---|---|
| Soltaninejad et al. (2018) [4] | Random Forest | 15 subjects in the control group and 15 in the PD group | 93.33% |
| Polat (2019) [5] | Logistic Regression | Daphnet Freezing of Gait Data Set | 81.3% |
| Das et al. (2017) [6] | ANN, SVM, Naïve Bayes | Physionet database | ANN: 90.6%<br>SVM: 64.00%<br>Naïve Bayes: 44.44% |
| Cho et al. (2009) [7] | PCA, LDA, MDC | Self-collected | LDA-MDC: 95.49%<br>PCA-MDC: 77.18% |
| Shetty & Rao (2016) [8] | SVM | Gaitpdb database | 83.33% |
| Alam et al. (2017) [9] | SVM | Gaitpdb database | 91.6% |
| Buongiorno et al. (2019) [10] | ANN | Self-collected | 89.4% |
| Trabassi et al. (2022) [11] | SVM | Self-collected | 86% |

*B. Deep learning methods*

In recent years, researchers have carried out several studies using deep learning (DL) techniques to aid in the early diagnosis and classification of PD. These studies have explored approaches such as EEG analysis, gait analysis, sensor data, and wearable devices. One particular study [12] focused on using an Artificial Neural Network (ANN) model to analyze EEG samples from a dataset of 110 subjects. Impressively this model achieved an accuracy rate of 98% with a sensitivity of 97% and specificity of 100% in distinguishing subjects with PD from those without it. Another study [13] combined LeNet and Long Short-Term Memory (LSTM) models to enhance PD diagnosis and understand the progression of the disease. The models used a dataset consisting of 102 spiral drawings.

In addition, researchers [14] have optimized Deep Neural Network (DNN) models to outperform traditional algorithms like random forests, SVM, XG Boost, and KNN in PD classification tasks. These optimized DNN models achieved high accuracy, precision, and F1 scores. For example, one study used a DNN model with three hidden layers trained on a dataset of 500 PD patients and 500 healthy controls, achieving an accuracy of 98%.

In a 2020 study [15], introduced a novel architecture called Bidirectional Long Short-Term Neural Network (BLSTM) for PD classification. The BLSTM architecture included 5 layers and incorporated both forward and backward time series information. The study involved 64 PD patients and 50 healthy controls who performed 13 motor exercises using wearable sensors. The proposed BLSTM model achieved a maximum accuracy of 82.94%, sensitivity of 92.3%, and specificity of 76.2%. Although the specificity of the model is relatively low due to the imbalanced disease staging in PD subjects, it still shows potential for classifying PD.

These studies also highlighted the importance of dataset selection and preprocessing in obtaining accurate results. For instance, a study [16] using a dataset of 33 participants, including MS patients, PD patients, and healthy older adults, applied multi-view vision-based techniques and DL models. The MSResNet model achieved a 100% accuracy in classifying the walking dataset. Another study [17] used the Gait in Parkinson's Disease Database (Gaitpdb) and achieved a high accuracy of 99.5%, sensitivity of 98.7%, and specificity of 99.1% in PD classification.

A recent study [18], the application of DL in combination with wearable sensors was examined for early detection of PD outside the clinic. They used data from the Parkinson's Progression Markers Initiative (PPMI) study, focusing on walk-like events. Their approach involved two main components, which are a dynamic activity detection module and a walk-like detection module. The data were fed into a 1D convolutional neural network (CNN) for PD classification. The results showed 90% accuracy for single walk-events and perfect accuracy with multiple events. This suggests DL's potential for PD detection.

In summary, these studies emphasize the effectiveness of DL techniques in PD diagnosis and classification. These approaches have the potential to enhance early detection, provide insights into disease progression, and contribute to personalized treatment strategies for PD. A summary of the DL methods is provided in Table 2. These studies

showcase the effectiveness of DL techniques, such as ANN, LeNet, LSTM, DNN, CNN, RNN, and hybrid models, in diagnosing and classifying PD. The accuracy, sensitivity, and specificity values vary across different datasets and approaches, demonstrating the potential of DL in improving early detection and understanding the progression of the disease.

Table 2. A summary highlighting the state-of-the-art deep learning methods. Note that the "-" entries indicate missing information or unavailable results for certain studies.

| Study | Deep Learning Technique | Dataset | Accuracy | Sensitivity | Specificity |
|---|---|---|---|---|---|
| M. Shaban (2021) [12] | Artificial Neural Network (ANN) | EEG Samples (110 subjects) | 98% | 97% | 100% |
| M. Sivakumar, A. H. Christinal and S. Jebasingh (2021) [13] | LeNet and Long Short-Term Memory (LSTM) | 102 spiral drawings | - | - | - |
| Savitha S. Upadhya et al. (2018) [14] | Deep Neural Network (DNN) | PD patients and controls (500 each) | 98% | | |
| Butt et al. (2020) [15] | BLSTM | Self-collected | 82.94% | 92.3% | 76.2% |
| Kaur et al. (2022) [16] | Multi-view vision-based DL models | MS patients, PD patients, and healthy older adults (33 participants) | MSResNet: 100% | - | - |
| Aşuroğlu and Oğul (2022) [17] | Hybrid DL model (CNN and Locally Weighted Random Forest) | Gaitpdb (Parkinson's Disease Database) | 99.5% | 98.7% | 99.1% |
| Atri et al. (2022) [18] | DL using sensor data | PPMI dataset (7 PD patients, 4 HCs) | 90% | - | - |

## III. PROPOSED SOLUTION

In this study, we propose a markerless vision-based gait analysis system which is based on human pose estimation (HPE). The overall process flow of the proposed approach is illustrated in Figure 1.

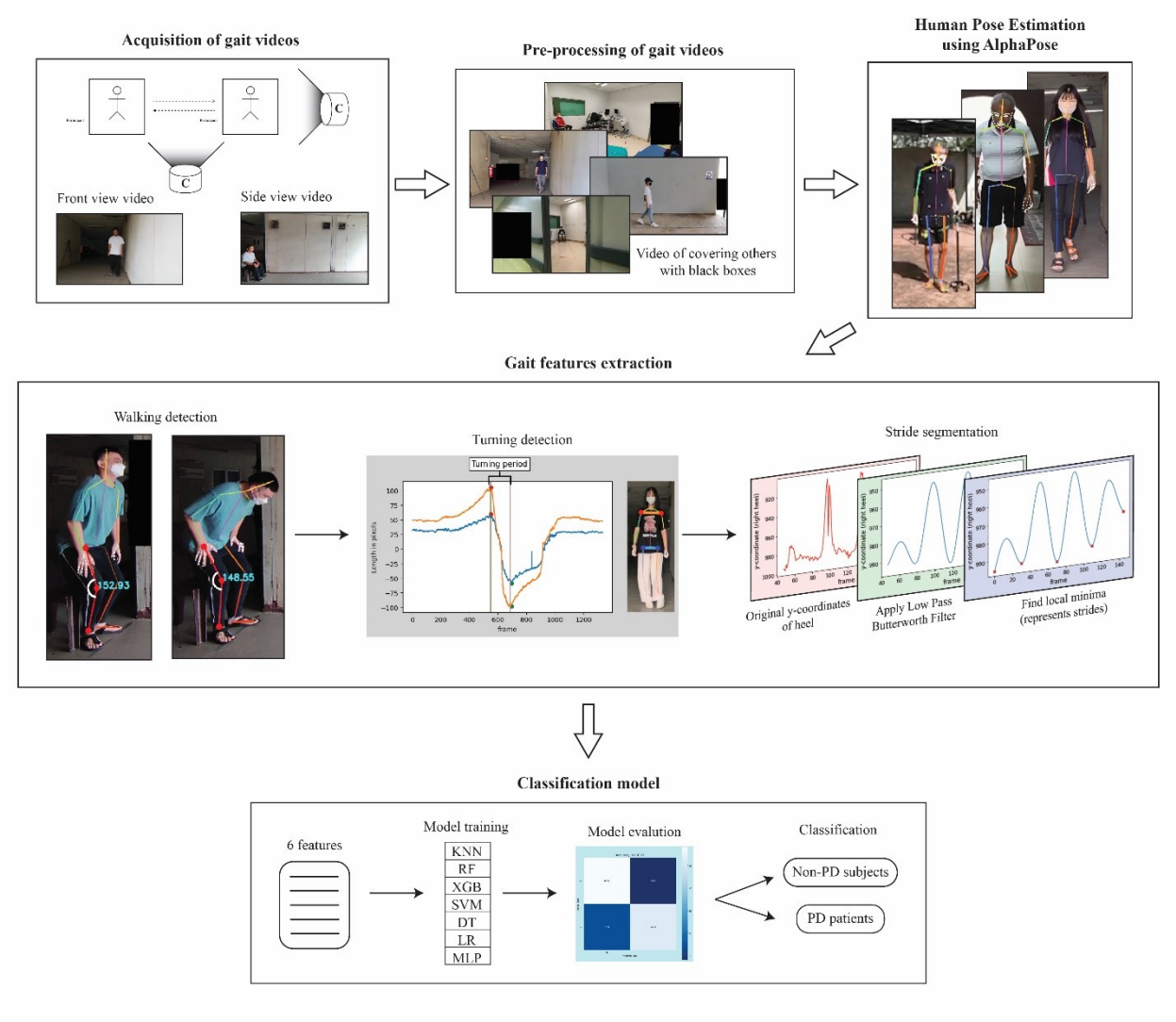



Figure 1. Overall process flow of proposed solution.

### *A. Data Collection*

#### *i. Subjects and Locations*

Gait videos were recorded from a total of 38 subjects in this study. Among them, 25 were healthy adults, 11 were normal patients, and the remaining subjects were diagnosed with Parkinson's disease (PD). The participant group consisted of 18 females and 20 males, with an average age of 40.98 ± 26.51 years. The gait video capture process for the healthy adult participants took place at Multimedia University Melaka campus (MMU). These individuals were either students or janitors affiliated with MMU. To ensure convenience and accessibility, the gait video collection was conducted within the premises of MMU.

For the normal patients, the gait videos were collected at University Malaya Medical Centre (UMMC) in collaboration with a professional specialist at UMMC. With the help of the specialist, the gait videos capture process went smoothly. In the case of PD patients, we conducted the video recordings in the comfort of their own homes. This decision was made to address the challenges and limitations faced by PD patients, especially in terms of mobility.

All participants were provided with comprehensive information about the nature of the study, including the methods used, potential risks, adverse effects, and possible complications, before participating in the study. They were fully informed about the purpose and objectives of the study. Each participant was required to complete an online consent form, explicitly stating their willingness to take part in this study and their acknowledgment and acceptance of the

study's terms and conditions. Additionally, we requested some basic personal information, such as identity card numbers, as well as details about their height and weight.

*ii. Gait Test Task*

In this study, Timed Up and Go test (TUG) [19] was selected for participants to perform during gait video capturing to assess their mobility and functional ability. TUG incorporates different gait activities such as standing up from a chair, turning around and walking which is more comprehensive compared to other gait tests. A typical TUG configuration was used in which subjects were requested to stand up from their chair, walk 3 meters, turn around, walk back to their seat, and sit down.

In order for participants to have a clearer understanding of the proper ways to perform TUG, we prepared a demonstration video with detailed instructions clearly written on the screen. In addition, the start and end points of the TUG were marked with masking tape to provide visual cues for participants. The following verbal instructions were also directed to them:

1. "When you hear the words "1 2 3 go", please stand up from your chair."
2. "After that, please walk to the marked line on the floor at your normal pace."
3. "When you have crossed that line, please turn around and walk back to your chair."

*iii. Shooting Site Configuration*

A shooting site configuration plan (refer to Figure 2) was outlined to maintain consistency in the captured gait videos. For our data collection, two tripod-mounted smartphone cameras were placed in front and the right side of the participants. The masking tape used to mark the start and end points is used as a reference to measure the distance between the two smartphone cameras and the walking path. During each data collection setup, the length of the masking tape will be fixed at 0.6 meters. Then, 1.7 meters is measured from the starting point and the result will be used to measure 1.65 meters laterally to place the camera for side angle capturing. While the front view camera will be placed at a distance of 2.2 meters from the center of the ending point. The complete shooting site configuration plan is sketched in the figure below.

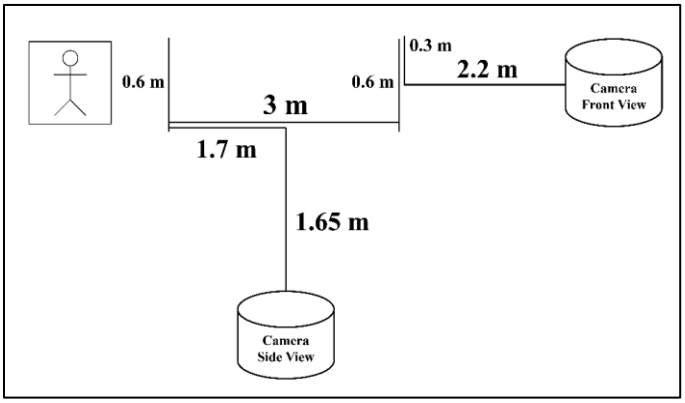



Figure 2. Shooting site configuration plan.

(a) Side-view video

(b) Front-view video

Figure 3. Example of captured gait videos

However, due to the insufficient space, the distance between the camera and the walking path will vary to ensure that the complete gait activities of the participants are captured. This occurs mainly when recording gait videos for normal patients in the consultation room at UMMC. In this case, the walking path of the TUG is forced to be oriented diagonally such that there is enough distance for the camera to be placed in front of the walking path. The camera for the side angle will be placed in the corner of the room and adjusted to be able to capture the complete gait cycle for the entire gait test (refer to Figure 3).

*iv. Online Dataset*

Due to the insufficient number of gait videos collected in our dataset, we expanded our collection efforts by including gait videos sourced from popular video sharing platforms, which are YouTube and Bilibili. We eventually collected 142 videos of PD patients and 150 videos of normal people. Consequently, our final dataset is composed of gait videos of 144 subjects with PD and 186 subjects without PD.

*C. Video Pre-Processing*

As the gait video shooting process was conducted in public places, the presence of passersby in the video was unavoidable. The walking pattern and characteristics of these passersby will interfere with each other resulting in anomalies in the data. Therefore, we pre-processed the original video to ensure the video consisted of only one subject. We achieved this by covering all the passersby in each frame with black boxes using video editing software.

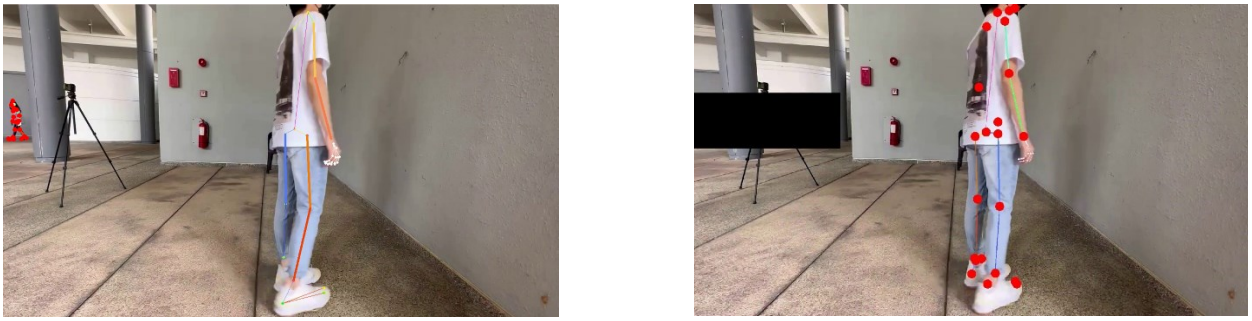


Figure 4. Presence of passer-by causing data anomalies and results after black box coverage of passer-by.
(Red points indicate the coordinates of the key points selected in the feature extraction)

For the online video, the presence of multiple people in the video occurs more frequently and the process of covering them with a black box becomes trickier. This is due to the uncontrollable nature of the recording conditions of online video, which results in the presence of people walking very close to the subject or passing by the camera and blocking the view of the camera. In this case, we blocked certain parts of the person who was very close to the subject. In this way, although the HPE will still be able to identify the person, the confidence score will be relatively low and can be filtered out in comparison with the confidence score of the subject (refer to Figure 4).

*D. Human Pose Estimation*

Although the marker-based gait analysis system could reliably determine the position of markers on the body with a precision of approximately 1 millimeter [20] the attached markers on the body often caused discomfort for the subjects,

thereby resulting in unnatural movements. The markerless approach has become an attractive trend in gait analysis by eliminating the time-consuming subject preparation often associated with marker-based systems, reducing the cost of expensive equipment or laborious manual processing of the data involved, improving subject comfort, and promoting more natural movements during analysis [21]. To achieve markerless motion capture, we perform human pose estimation by using AlphaPose, a real-time multi-person pose estimation software that detects and tracks the poses of multiple people in images and videos.

AlphaPose adopted a top-down approach to perform accurate human pose estimation. Typically, AlphaPose starts by detecting individuals in an image or video frame and generating bounding boxes around them. Then, it proceeds to estimate their body joint position and pose. For each gait video we collected, AlphaPose outputted a video with pose tracking of the person and a JavaScript Object Notation (JSON) file. The JSON file consisted of the coordinates of key points in each frame that represent the skeleton joints of the examined persons. A total of 136 key points were generated by the AlphPose, 26 of which are related to the human body, 68 to the face and 42 to the hands. Each key point consists of three elements, which were in the order of the coordinate on the x-axis, the coordinate on the y-axis, and its confidence score.

*E. Features Extraction*

As gait difficulties were one of the main motor symptoms of PD, we extracted five gait-related features from the coordinates of the obtained skeletal key points. They were the stride length of the left and right foot, cadence, gait speed, duration of turning, and steps taken during turn. We divided the feature extraction process into three consecutive parts, which were walking detection, followed by turning detection, and eventually stride segmentation.

Since the TUG task involved the subject's transition from sitting to walking and back to sitting, it was crucial to determine the specific frame numbers in which the subject walked before the stride segmentation process to ensure accurate detection of strides. This was because by focusing on the periods in which the subject walked and excluding the frames in which the subject sat, we can minimize the detection of spurious or noisy strides that might occur when the subject was at rest. In addition, the implementation of walking detection provided the advantage of achieving fully automated feature extraction in subject-only shooting environments. This eliminated the need to manually remove the initial and final segments of the video because the walking detection algorithm can accurately identify specific periods of walking.

Turning while walking was a common motor symptom observed in patients with PD, characterized by difficulties in changing direction [22-24]. The extraction of information about the turning period could be valuable to distinguish PD patients from non-PD subjects. Therefore, after walking detection, we applied a turning detection algorithm to accurately recognize the frame number corresponding to the beginning and ending of the subject's turning period in order to be able to extract valuable features about their turns.

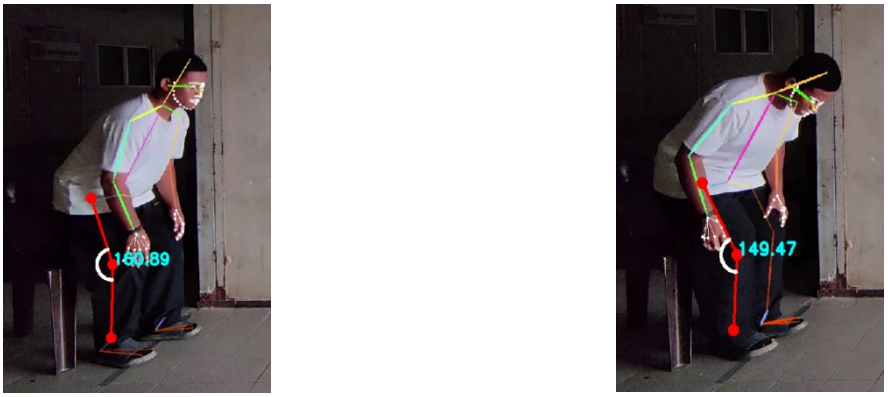
After determining the frame numbers of subject’s walking and turning, we obtained two walking sequences in opposite directions, which are forward and backward as observed from the front view. Specifically, the subject's walking sequence was categorized as forward after the starting of walking and before the turning period, while it was categorized as backward after the turning period and before sitting back into the chair. We then performed the stride segmentation separately on these two sequences. The detected strides were then further used to compute other gait features.

To facilitate walking detection and stride segmentation, we utilized the subject’s side-view video. The side angle provided a beneficial perspective for accurate identification and analysis of the subject's walking patterns and strides. However, to determine the turn period of the subject, we utilized the subject’s front-view video as the front angle provided clearer depiction of hip and shoulder orientation, which was important in turning detection. The detailed implementation of the detection and segmentation algorithms would be explained in the following subsections.

*i. Walking Detection*

As the camera for side angle is always positioned on the right side of the subject, we considered the bending degree of the subject's right foot to determine the sitting stage. The coordinates of the ankle, knee, and right hip key points estimated from the subject’s side-view video were taken and the angle in degrees between them was calculated. Then, we defined a threshold value of 150°. If the angle between the ankle, knee, and right hip was greater than the threshold for the first time, it indicated that the subject had started walking. Conversely, when the angle was less than the threshold, we considered that the subject had sat back down in the chair. Illustrations are given in Figure 5. To ensure

accurate detection, we further verified this by comparing the duration of the video with its' overall time. If the angle remained below the threshold and the video had exceeded 80% of its total duration, we interpreted that this indicated that the subject had indeed started to sit back in the chair.




(a) Angle greater than 150° for the first time (b) Angle less than 150° and time exceeded 80%

Figure 5. Result of walking detection algorithm.
(a) indicated the subject had started walking. (b) indicated the subject had started sitting back to the chair.

*ii. Turning Detection*

To detect the subject's turning period, the HPE results of the subject's front-view video were used. After analyzing the video, we observed that the distance between the subject and the camera was the lowest during the subject's turning period. This occurred because the subject was walking toward the camera before they started to turn around. In other words, we observed that at the beginning and the end of the turning period, the subjects appeared at their closest proximity to the camera, resulting in the largest visual presence. Thus, we exploited this observation to determine the subject's turning period.

We first computed the differences in x-coordinates of two key point pairs, namely right and left hips as well as the right and left shoulders. The selection of these key point pairs was based on the observation that when a person intended to turn, the movement of the hips and shoulders tended to precede and initiate the turning movement. Then, we determined the maximum and minimum differences for the two key point pairs. The maximum difference indicated the widest length of the hips and shoulders, while the minimum difference also indicated the widest length due to the swapping of the left and right sides of the body during the turn. We took the average of the maximum and minimum hip and shoulder length to determine the optimal starting and ending frames of the subject's turning period, respectively.

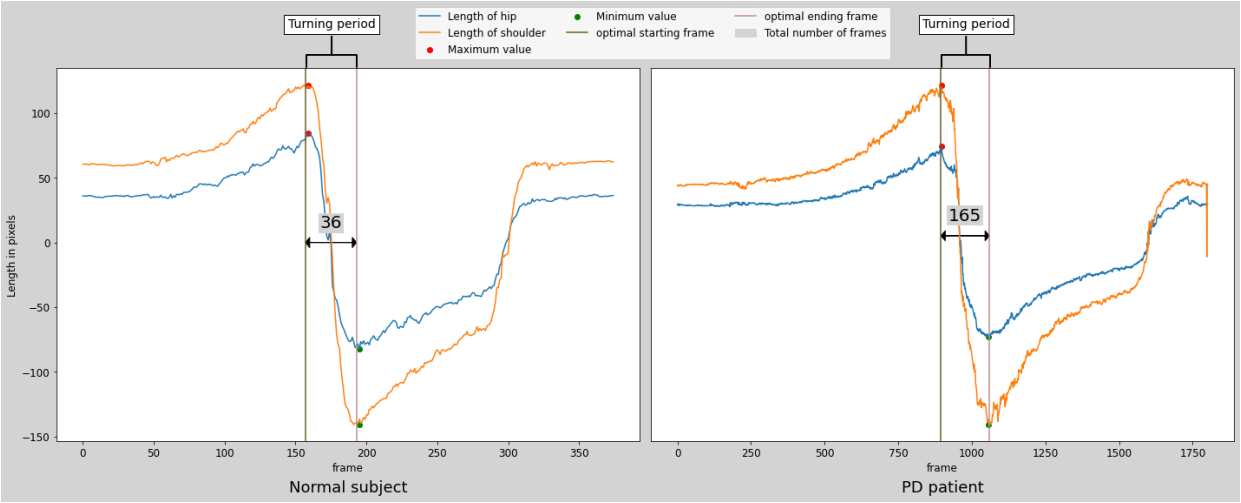



Figure 6. Results of turning detection algorithm for normal subject and PD patient.

Figure 6 illustrates the results of the turning detection algorithm that was applied to two different subjects, i.e., normal adult and PD patient. We observed a significant increase in the total number of frames during turning in PD patient, approximately five times greater than in normal subject. This significant difference highlighted the impact of PD on motor function, which contributed to the extended duration of the turning period, resulting in a higher number of frames. This led to our focus on extracting the subjects' turning features as well. The duration of the turning period in seconds was computed based on the frame numbers of the starting and ending frames. Specifically, referring to Eq (1):

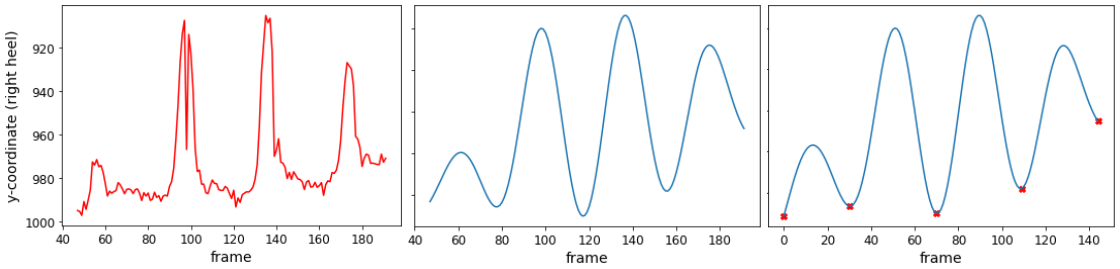
$$duration = \frac{t_{end} - t_{start}}{FPS} \tag{1}$$

By subtracting the starting frame number $t_{start}$ from the ending frame number $t_{end}$, we obtained the total number of frames encompassing the turning period. This value was then divided by the frame rate of the video (FPS) to convert it into the corresponding duration in seconds.

*iii. Stride segmentation*

The walking sequence of the subjects was divided into two sequences as described in the previous section. The stride segmentation process was performed on both feet for both feet in each of these sequences as their opposite directions could potentially influence the results. The final gait features were derived by calculating the average value of the two sequences, considering both feet. In the following section, our focus was solely on the stride segmentation of the right foot in the forward sequences.

The stride length was defined by the distance from heel contact of one foot to continuous heel contact of the same foot. Hence, we focus on the heel key point to segment the stride. Let $h = \{f_t, y_t\}$ denote the frame numbers and y-coordinates of the heel key point in frame *t*. Then, we defined the stride as the local minima of the $y_t$ with a minimal horizontal distance between them, specifically set to FPS of the video. This ensures that the detected strides are at least one second apart to avoid considering adjacent local minima that may represent noise. However, we found the output data of AlphaPose was very noisy, therefore we smoothed the data by using Low Pass Butterworth Filter before the segmentation. The Butterworth filter had a cutoff value of 1, the sampling frequency was set to the FPS of the video, and the order of the filter was set to 5. Due to the exclusion of frames in which subjects were at rest and during the turning period, there was a possibility that the starting and ending stride could not be accurately segmented. To address this issue, a solution was implemented to add one stride at the beginning and end if no stride was detected within the first and last 10% of the walking distance. This additional step ensured that the starting and ending strides were appropriately captured, even if they were initially missed during the segmentation process. The Figure 7 shows the y-coordinate of $h_{right}$ before and after smoothing as well as the detected right strides.




(a) Original data (b) After smoothing (c) Detected strides

Figure 7. Stride segmentation

We further derived the stride length, cadence, and gait speed from the detected strides. In terms of stride length, we first subtracted the difference of x-coordinate of every two successive detected strides which represents the distance in pixels between two strides. Considering that the distance between the subject and the camera changed in some videos, as described in section 3.1.3., instead of further converting the distance between the two strides into centimeters, we computed the ratio between the distance of the strides and the mean of subject's nose-to-hip height

over all frames. After that, we calculated the mean value as stride length of the subject. The computation process of the right stride length can be described as depicted in Eq (2) – Eq (6):

(i) First, let $Distance_{right} = \{d_1, d_2 \dots, d_{n-1}\}$ where n equals to the number of the detected strides, and then we calculate the distance as:

$$d_i = |x_i - x_{i-1}| \tag{2}$$

(ii) Calculate the of subject's nose-to-hip height, $nth$:

$$nth_t = y_t^{nose} - y_t^{hip} \tag{3}$$

where *t* denotes the frame numbers, $y_t^{nose}$ is equal to the nose key point, and $y_t^{hip}$ is the hip key point.

(iii) The mean of subject's nose-to-hip height, $\overline{nth}$ can be found as:

$$\overline{nth} = \frac{\sum_{i=1}^{t} nth_i}{t} \tag{4}$$

(iv) Calculate the ratio, $Ratio_{right} = \{r_1, r_2 \dots, r_{n-1}\}$:

$$r_i = \frac{d_i}{\overline{nth}} \tag{5}$$

(v) Calculate the mean value of $Ratio_{right}$, which is $\overline{Ratio_{right}}$ :

$$\overline{Ratio_{right}} = \frac{\sum_{i=1}^{n-1} r_i}{n-1} \tag{6}$$

The $\overline{Ratio_{right}}$represent the right stride length of the subject. We repeated the same calculations for the left stride to obtain the left stride length, $\overline{Ratio_{left}}$

Cadence was defined as the number of steps per minute. In order to compute the cadence, we first need to determine the total number of steps taken. We achieved this by counting the strides of both the right and left feet and subtracting 1 from each count due to the fact that two successive detected heel contacts form a single stride. Subsequently, the strides count for both feet were summed to obtain the total number of steps taken. For total time taken for walking, we considered the duration of the walking sequences. As the walking sequences were already segmented and

categorized, we had the specific frame numbers corresponding to the walking periods. By calculating the difference between the starting and ending frames of the walking sequences and then divided by FPS of the video, we obtained the total time taken in second for walking. We further converted the units to minutes. Then, we compute the cadence by taking the total number of steps divided by the total walking time, which can be written as below, referring to Eq (7) – Eq (9):

$$Cadence = \frac{Steps}{Time} \tag{7}$$

where $Steps$ is total number of steps,

$$Steps = \left(n_{right} - 1\right) + \left(n_{left} - 1\right) \tag{8}$$

and $Time$ refer to total time taken in minutes,

$$Time = \frac{t_{end} - t_{start}}{FPS * 60} \tag{9}$$

Gait speed was defined as the distance covered per second [25]. As stated earlier, the extracted stride length was represented as a ratio between the distance of the strides and the mean of the subject's nose-to-hip height. Therefore, we defined the gait speed as the average ratio covered per second. The computation process for gait speed was similar to that of cadence. First, we obtained the mean ratio of both feet by averaging the ratios calculated for each foot. Then, we divided this mean ratio by the total walking time. The calculation of gait speed can be described as depicted in Eq (10) – Eq (12):

$$Speed = \frac{Ratio}{Time} \tag{10}$$

where $Ratio$ can be found as,

$$Ratio = \frac{\overline{Ratio_{right}} + \overline{Ratio_{left}}}{2} \tag{11}$$

and $Time$ indicate total time taken in seconds,

$$Time = \frac{t_{end} - t_{start}}{FPS} \tag{12}$$

In terms of the steps taken during the turn, we applied the same stride segmentation process to the turning period of the subject in order to segment the strides of both feet. After that, we extracted the number of steps using a similar approach as in the extraction process for cadence. However, since we did not include an additional stride at the beginning and end of the turning period, we did not subtract 1 from the total number of strides for each foot. In other words, the number of steps in the turn is equal to the sum of the left and right strides detected during the turn. In summary, steps taken during the turn can be found as shown in Eq (13):

$$Steps = n_{right} + n_{left} \tag{13}$$

*iii. Online Dataset*

The gait feature extraction process for the online dataset was adjusted due to the variation in recording conditions present in the online video. These variations included differences in recording angles, camera movement, distance between the camera and the subject, and availability of turning periods. Therefore, we were unable to develop a general algorithm that could accurately detect the subject's turning periods. To address this issue, we manually labeled frame numbers indicating when the subject started and finished turning. For the videos that lacked any turning period, we assigned a null value to indicate the absence of such periods.

Additionally, we implemented a different algorithm to calculate the total time taken for walking. Let $h_i = \{f_t, y_t\}$ denote the detected right stride. To determine the total walking time, we first identified the minimum frame numbers between the first detected strides of both feet. Then, we determined the maximum frame numbers between the last detected strides of both feet. Afterwards, we computed the difference between these values and divided it by the FPS of the video to obtain the total walking time in seconds. The computation process can be expressed as in Eq (14):

$$Time = \frac{argmax\left(t_{n-1}^{right}, t_{n-1}^{left}\right) - argmin\left(t_0^{right}, t_0^{left}\right)}{FPS} \tag{14}$$

Admittedly, certain videos in our collected online dataset, featuring a front-view perspective, had led to inaccuracies in the computed stride length. Especially for normal people, their stride length was estimated to be extremely small. To address this problem, we considered an alternative approach to compute their stride length. We modified the original method of subtracting x-coordinates to instead subtracting y-coordinates. This modification was based on our observation that, in certain cases, the y-coordinate movement provided more reliable and informative data for estimating stride length. However, this alternative approach did not completely resolve the issue of inaccurate stride length. Therefore, we replaced the stride length with the results of the new approach only when the computed value exceeded the original stride length.

## IV. EXPERIMENTAL RESULTS

### *A. Experiment Setup*

Our dataset comprised 6 gait features extracted from a total of 330 subjects, as detailed in the previous section. In the following section, we will use the following terms to denote the extracted gait features, which are ratioR (right stride length), ratioL (left stride length), cadence (cadence), mean_speed (gait speed), turning_duration (duration of turning period) and turning_steps (number of steps taken during).

We handled the missing values in our dataset by filling in the median values. Then, we divided our dataset into training and testing subsets using a ratio of 80% for training and 20% for testing. To maintain consistency and reproducibility, we set a fixed random state value of 42 during the dataset split as well as the subsequent models training in the following section. After splitting, the training subset included 264 subjects, while the test subset consisted of 66 subjects. To address the varying range of values exhibited in our dataset, we performed feature scaling by using the standard scaler method to normalize our data. Then, we trained all the models using this dataset.

### *B. Overall performance of classification models*

In this study, we implemented seven machine learning algorithms to classify PD patients and non-PD subjects. The algorithms used were Logistic Regression (LR), Random Forest (RF), K-Nearest Neighbors (KNN), Decision Tree (DT), eXtreme Gradient Boosting (XGB), Support Vector Machine (SVM), and Multilayer Perceptron (MLP). All algorithms were trained on the same dataset, and the random state was set to 42. We optimized the algorithms by performing hyperparameter tuning using the grid search method with 4-fold cross-validation. The training accuracy of all algorithms is summarized in the Table 3 below.

Table 3. Training accuracy of every algorithm

| Algorithms | Hyperparameters used | Accuracy |
|---|---|---|
| LR | {'C': 0.01, 'fit_intercept': True, 'max_iter': 1000, 'penalty': 'none', 'solver': 'sag'} | 0.834 |
| RF | {'max_depth': 10, 'min_samples_leaf': 5, 'n_estimators': 100} | 0.859 |
| KNN | {'leaf_size': 1, 'n_neighbors': 5} | 0.826 |
| DT | {'max_depth': 5, 'min_samples_leaf': 10, 'splitter': 'best'} | 0.856 |
| XGB | {'booster': 'gbtree', 'gamma': 0.1, 'learning_rate': 0.05, 'max_depth': 7, 'n_estimators': 500, 'reg_alpha': 1, 'reg_lambda': 0.01} | **0.875** |
| SVM | {'C': 10, 'gamma': 0.1, 'kernel': 'rbf'} | 0.841 |
| MLP | {'activation': 'relu', 'hidden_layer_sizes': (200, 1000), 'max_iter': 200, 'solver': 'adam'} | 0.844 |

We observed that the gradient boosting algorithm XGB was outperforming other algorithms in terms of training accuracy. Therefore, we used XGB algorithm to perform classification on the testing dataset. The performance of the selected algorithm on testing dataset is shown in the following classification report (Table 4). Class 0 indicates the non-PD subject while class 1 indicates PD patients.

Table 4. Classification report of XGB on testing dataset

| Algorithm | Accuracy | Class | Precision | Recall | F1-Score |
|---|---|---|---|---|---|
| XGB | 0.879 | 0 | 0.903 | 0.848 | 0.875 |
| | | 1 | 0.857 | 0.909 | 0.882 |
| | | Average: | 0.880 | 0.879 | 0.879 |

The XGB algorithm achieved an accuracy of 0.879 when classifying the testing dataset. When considering the class of non-PD subjects, it exhibited a precision of 0.903, a recall of 0.848, and an F1-score of 0.875. Regarding PD patients, the algorithm attained a precision of 0.857, a recall of 0.909, and an F1-score of 0.882. Taking both classes into account, the average precision, recall, and F1-score were 0.880, 0.879, and 0.879, respectively.

*C. Features investigating*

In order to identify the most informative and relevant gait features that contribute to the predictive accuracy of the PD classification model, we conducted a features investigation to analyze the characteristics and properties of the gait features in our dataset.

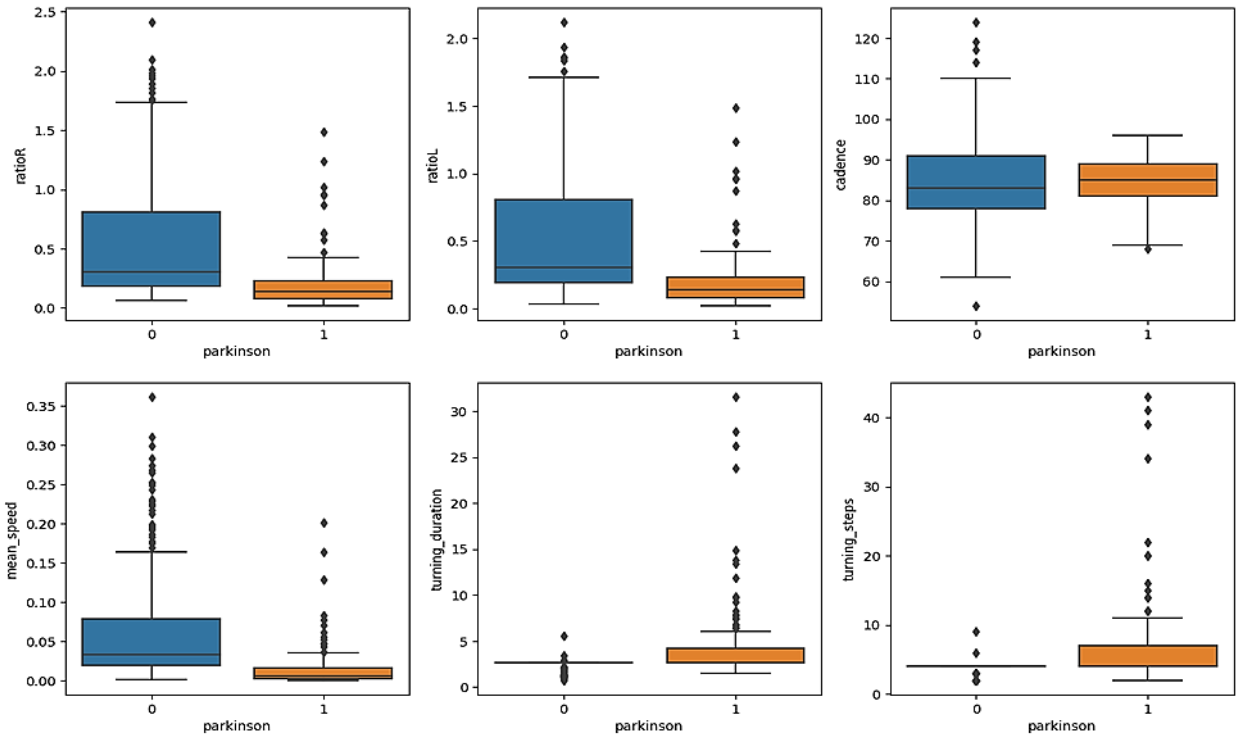



Figure 8. Box plot of the extracted features

In Figure 8, we visually represented the distribution and central tendency spread of each extracted gait feature from our dataset. It was observed that there was significant overlap in the distribution of data between the two classes. Specifically, by observing the median value, for features such as ratioR, ratioL, cadence, and mean_speed, approximately 50% of the non-PD subjects' data fell within the same distribution as that of the PD patients' class. We believed that this overlap occurred due to limitations in our gait features extraction algorithms, which was affected by the uncontrollable recording conditions of the online videos. Factors such as horizontal camera movement or shaking resulted in increased distances in the x-coordinates of detected strides, camera shaking in an upward direction led to incorrect stride detection. Moreover, many videos in the online dataset only provided front-view perspectives, lacking crucial depth information for accurate gait analysis. Although there was a significant overlap in our dataset regarding these issues, the non-PD class generally exhibited larger median values, maximum values, and values for the third quartile (Q3) compared to the PD class, except cadence. This observation is consistent with the findings of other studies in which patients with PD tend to have slower gait speed and shorter stride length.

In terms of turning_duration and turning_steps, due to a shortage of the videos that showcasing normal subjects turning in the online dataset, the majority of data for non-PD subjects was represented by the median value of the dataset. Therefore, in our further analysis of turning features, we focused only on the data that does not contain any null values.

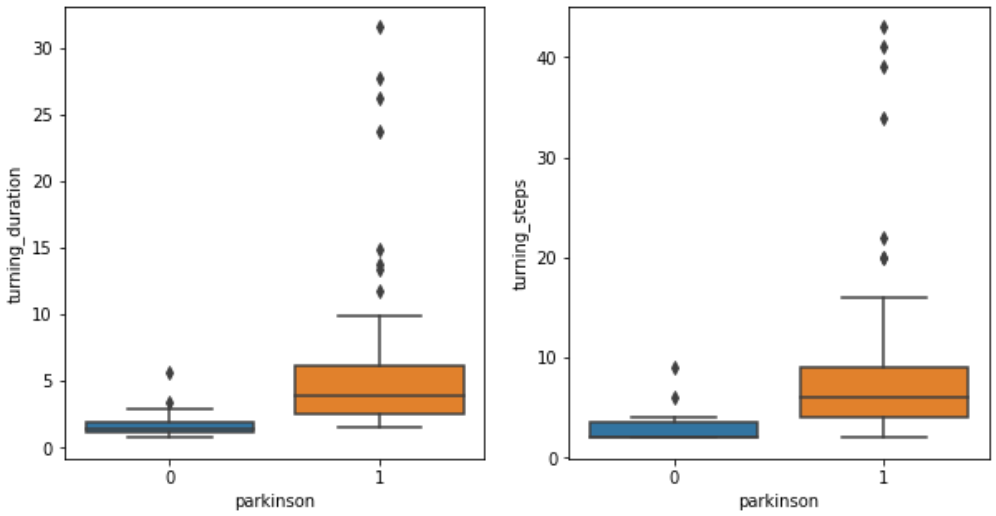



Figure 9. Box plot of turning features after removing data containing null values.

After removing the data containing null values, a notable observation is the distinct separation in data distribution for the two turning features across the two classes, as shown in Figure 9. For example, the minimum value of turning_duration for the PD class exceeded the median value of the non-PD class, while the maximum value of turning steps for the non-PD class fell below the first quartile (Q1) of the PD class. We believe that this significant difference in the data distribution of turning features could greatly improve the classification performance in PD patients.

*D. Effect of different features on classification performance*

In this section, the impact of the extracted gait features on classification performance will be analysed and compared. We retrained the XGB classifier using each single gait feature and recorded its performance on the testing dataset. The results are presented in the following Table 5.

Table 5. Performance of stand-alone gait features.

| Gait features | Accuracy | Class | Precision | Recall | F1-score |
|---|---|---|---|---|---|
| ratioR | 0.773 | 0 | 0.725 | 0.879 | 0.795 |
| | | 1 | 0.846 | 0.667 | 0.746 |
| | | Average: | 0.786 | 0.773 | 0.770 |
| ratioL | 0.788 | 0 | 0.757 | 0.848 | 0.800 |
| | | 1 | 0.828 | 0.727 | 0.774 |
| | | Average: | 0.792 | 0.788 | 0.787 |
| cadence | 0.591 | 0 | 0.583 | 0.636 | 0.609 |
| | | 1 | 0.600 | 0.545 | 0.571 |
| | | Average: | 0.592 | 0.591 | 0.590 |
| mean_speed | **0.818** | 0 | 0.818 | 0.818 | 0.818 |
| | | 1 | 0.818 | 0.818 | 0.818 |
| | | Average: | 0.818 | 0.818 | 0.818 |
| turning_duration | 0.652 | 0 | 0.604 | 0.879 | 0.716 |
| | | 1 | 0.778 | 0.424 | 0.549 |
| | | Average: | 0.691 | 0.652 | 0.633 |
| turning_steps | 0.636 | 0 | 0.582 | 0.970 | 0.727 |
| | | 1 | 0.909 | 0.303 | 0.455 |
| | | Average: | 0.745 | 0.636 | 0.591 |

In the above Table 5, we found that the classification model trained with mean_speed obtained the highest overall performance, while the overall performance was the lowest for the model trained with cadence. This aligns with what

was observed in Figure 8, where the data distribution of mean_speed was the most pronounced and the data distribution of the cadence was most overlapped. Furthermore, the precision of turning_steps for the PD patients' class was exceptionally high, surpassing even the precision of the original classification model trained on all gait features. This result highlighted the significant impact of the turning_steps feature in accurately identifying true PD patients.

We further experimented with different combinations of gait features to investigate their impact on classification performance. The resulting classification accuracy on the testing dataset was recorded and summarized in the following Table 6.

Table 6. Performance of different combinations of gait features

| Gait features | | | | | | Accuracy |
|---|---|---|---|---|---|---|
| ratioR | ratioL | cadence | mean_speed | turning_duration | turning_steps | |
| ✓ | ✕ | ✕ | ✕ | ✕ | ✕ | 0.773 |
| ✓ | ✓ | ✕ | ✕ | ✕ | ✕ | 0.788 |
| ✓ | ✓ | ✓ | ✕ | ✕ | ✕ | 0.697 |
| ✓ | ✓ | ✓ | ✓ | ✕ | ✕ | 0.803 |
| ✓ | ✓ | ✓ | ✓ | ✓ | ✕ | 0.864 |
| ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 0.879 |

As we progressively included more gait features during the training process, we observed a general trend of improving accuracy. However, the inclusion of the cadence feature resulted in a drop in accuracy by 0.091. This decrease can be attributed to the confusion caused by cadence due to its overlapped data distribution between two classes, as described previously. Hence, we exclude the cadence from the model to determine if the accuracy of the classification model could be improved without the inclusion of cadence.

Table 7. Classification performance on testing dataset after excluding cadence.

| Algorithm | Accuracy | Class | Precision | Recall | F1-Score |
|---|---|---|---|---|---|
| XGB | 0.894 | 0 | 0.882 | 0.909 | 0.896 |
| | | 1 | 0.906 | 0.879 | 0.892 |
| | | Average: | 0.894 | 0.894 | 0.894 |

As shown in Table 7, the accuracy of the classification model improved slightly after removing the cadence in the training process. The precision of class 0 and the recall of class 1 were slightly decreased. This means that the number of actual PD patients incorrectly classified as non-PD subjects has slightly increased. On the other hand, the precision of class 1 and recall of class 0 were improved due to a slight decrease in the number of actual non-PD subjects being misclassified as PD patients. As the overall performance improved, we concluded that this modified model would be considered as our final classification model.

## V. CONCLUSION

This paper presents a novel approach to detect neurodegenerative disorders through vision-based gait analysis. To collect data for our study, we asked our participants to perform TUG test and their walking patterns were recorded by using two smartphones. Then, we used a pose estimation tool called AlphaPose to estimate the human pose of the subjects from these videos. Gait features such as stride length, cadence, gait speed, turning duration, and the number of steps taken during turning were extracted from the estimated human pose. In our experiment, we trained a variety of conventional machine learning classifiers, including K-Neighbors, Random Forest, XGB, SVM, Decision Tree, Logistic Regression, and MLP, using the extracted gait features. These classifiers were then evaluated based on their training dataset accuracy. Among them, the XGB classifier outperformed others with a training accuracy of 87.5% and was selected for classification on the testing dataset, where it achieved an accuracy of 87.87%. To further improve the performance, we investigated the extracted features. We observed a significant difference in the data distribution of the turning features. However, due to the lack of videos capturing the subjects' turning movements in our dataset,

these turning features did not significantly improve the classification performance of our classifier. Furthermore, our findings revealed that gait speed was the most significant feature in classification and including cadence as a feature had caused confusion in the classification process of our classifier. By removing cadence from the classification model, the XGB classifier achieved the highest accuracy of 89.39%. In conclusion, this study showed the positive potential of a markerless vision-based gait analysis system for PD detection.

## ACKNOWLEDGEMENT

This project is funded by a Multimedia University and Universitas Telkom Joint research grant (MMUE/210063) and Fundamental Research Grant Scheme (FRGS/1/2020/ICT02/MMU/02/5).

## REFERENCES

[1] S. Chaudhary & N. Tyagi (2022). A Review on Parkinson's Disease: Overview and Management. International Journal of Pharmaceutical Sciences Review and Research, 18–24. https://doi.org/10.47583/ijpsrr.2022.v76i01.004

[2] B. Yvette (2023, April 4). Parkinson's disease: Early signs, causes, and risk factors. https://www.medicalnewstoday.com/articles/323396

[3] J. Parkinson (1817). An Essay on the Shaking Palsy. In J Neuropsychiatry Clin Neurosci (Vol. 14, Issue 2).

[4] S. Soltaninejad, A. Rosales-Castellanos, M. A. Ibarra-Manzano & L. Cheng (2018). Body Movement Monitoring for Parkinson's Disease Patients Using A Smart Sensor Based Non-Invasive Technique. 2018 IEEE 20th International Conference on E-Health Networking, Applications and Services (Healthcom), 1–6. https://doi.org/10.1109/HealthCom.2018.8531197

[5] K. Polat (2019). Freezing of Gait (FoG) Detection Using Logistic Regression in Parkinson's Disease from Acceleration Signals. 2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT), 1–4. https://doi.org/10.1109/EBBT.2019.8742042

[6] K. Devi Das, A. J. Saji & C. S. Kumar (2017). Frequency analysis of gait signals for detection of neurodegenerative diseases. 2017 International Conference on Circuit ,Power and Computing Technologies (ICCPCT), 1–6. https://doi.org/10.1109/ICCPCT.2017.8074273

[7] C. W. Cho, W.H. Chao, S.H. Lin & Y.Y. Chen (2009). A vision-based analysis system for gait recognition in patients with Parkinson's disease. Expert Systems with Applications, 36(3), 7033–7039. https://doi.org/10.1016/j.eswa.2008.08.076

[8] S. Shetty & Y. S. Rao (2016). SVM based machine learning approach to identify Parkinson's disease using gait analysis. 2016 International Conference on Inventive Computation Technologies (ICICT), 1–5. https://doi.org/10.1109/INVENTIVE.2016.7824836

[9] M. N. Alam, A. Garg, T. T. K. Munia, R. Fazel-Rezai & K. Tavakolian (2017). Vertical ground reaction force marker for Parkinson's disease. PLOS ONE, 12(5), e0175951. https://doi.org/10.1371/journal.pone.0175951

[10] D. Buongiorno, I. Bortone, G. D. Cascarano, G. F. Trotta, A. Brunetti & V. Bevilacqua (2019). A low-cost vision system based on the analysis of motor features for recognition and severity rating of Parkinson's Disease. BMC Medical Informatics and Decision Making, 19(S9), 243. https://doi.org/10.1186/s12911-019-0987-5

[11] D. Trabassi, M. Serrao, T. Varrecchia, A. Ranavolo, G. Coppola, R. De Icco, C. Tassorelli & S. F. Castiglia (2022). Machine Learning Approach to Support the Detection of Parkinson's Disease in IMU-Based Gait Analysis. Sensors, 22(10), 3700. https://doi.org/10.3390/s22103700

[12] M. Shaban (2021). Automated Screening of Parkinson's Disease Using Deep Learning Based Electroencephalography. 2021 10th International IEEE/EMBS Conference on Neural Engineering (NER), 158–161. https://doi.org/10.1109/NER49283.2021.9441065

[13] M. Sivakumar, A. H. Christinal & S. Jebasingh (2021). Parkinson's disease Diagnosis using a Combined Deep Learning Approach. 2021 3rd International Conference on Signal Processing and Communication (ICPSC), 81–84. https://doi.org/10.1109/ICSPC51351.2021.9451719

[14] S. S. Upadhya & A. N. Cheeran (2018). Discriminating Parkinson and Healthy People Using Phonation and Cepstral Features of Speech. Procedia Computer Science, 143, 197–202. https://doi.org/10.1016/j.procs.2018.10.376

[15] A. H. Butt, F. Cavallo, C. Maremmani, & E. Rovini (2020). Biomechanical parameters assessment for the classification of Parkinson Disease using Bidirectional Long Short-Term Memory. 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), 5761–5764. https://doi.org/10.1109/EMBC44109.2020.9176051

[16] R. Kaur, R. W. Motl, R. Sowers & M. E. Hernandez (2022). A Vision-Based Framework for Predicting Multiple Sclerosis and Parkinson's Disease Gait Dysfunctions—A Deep Learning Approach. IEEE Journal of Biomedical and Health Informatics, 27(1), 190–201. https://doi.org/10.1109/JBHI.2022.3208077

[17] T. Aşuroğlu & H. Oğul (2022). A deep learning approach for parkinson's disease severity assessment. Health and Technology, 12(5), 943–953. https://doi.org/10.1007/s12553-022-00698-z

[18] R. Atri, K. Urban, B. Marebwa, T. Simuni, C. Tanner, A. Siderowf, M. Frasier, M. Haas & L. Lancashire (2022). Deep Learning for Daily Monitoring of Parkinson's Disease Outside the Clinic Using Wearable Sensors. Sensors, 22(18), 6831. https://doi.org/10.3390/s22186831

[19] D. Podsiadlo & S. Richardson (1991). The Timed “Up & Go”: A Test of Basic Functional Mobility for Frail Elderly Persons. Journal of the American Geriatrics Society, 39(2), 142–148. https://doi.org/10.1111/j.1532-5415.1991.tb01616.x

[20] R. Baker (2006). Gait analysis methods in rehabilitation. Journal of NeuroEngineering and Rehabilitation, 3. https://doi.org/10.1186/1743-0003-3-4

[21] S. L. Colyer, M. Evans, D. P. Cosker & A. I. T. Salo (2018). A Review of the Evolution of Vision-Based Motion Analysis and the Integration of Advanced Computer Vision Methods Towards Developing a Markerless System. In Sports Medicine - Open (Vol. 4, Issue 1). Springer. https://doi.org/10.1186/s40798-018-0139-y

[22] P. Y. Chou & S. C. Lee (2013). Turning deficits in people with Parkinson’s disease. In Tzu Chi Medical Journal (Vol. 25, Issue 4, pp. 200–202). https://doi.org/10.1016/j.tcmj.2013.06.003

[23] J. Das, R. Vitorio, A. Butterfield, R. Morris, L. Graham, G. Barry, C. McDonald, R. Walker, M. Mancini & S. Stuart (2022). Visual Cues for Turning in Parkinson’s Disease. Sensors, 22(18). https://doi.org/10.3390/s22186746

[24] M. Mancini, A. Weiss, T. Herman & J. M. Hausdorff (2018). Turn around freezing: Community-living turning behavior in people with Parkinson’s disease. Frontiers in Neurology, 9(JAN). https://doi.org/10.3389/fneur.2018.00018

[25] Y. F. Ti, T. Connie, M.K.O. Goh. (2023). GenReGait: Gender Recognition using Gait Features. Journal of Informatics and Web Engineering, 2(2). https://doi.org/10.33093/jiwe.2023.2.2.10