

Model-Based Feature Extraction and Classification for Parkinson's Disease Screening using Gait Analysis

Ming De Lim, Tee Connie, Michael Kah Ong Goh, Nor 'Izzati Saedon

Submitted to: JMIR Aging
on: August 21, 2024

Disclaimer: © The authors. All rights reserved. This is a privileged document currently under peer-review/community review. Authors have provided JMIR Publications with an exclusive license to publish this preprint on its website for review purposes only. While the final peer-reviewed paper may be licensed under a CC BY license on publication, at this stage authors and publisher expressly prohibit redistribution of this draft paper other than for review purposes.

Table of Contents

Original Manuscript.....	5
Supplementary Files.....	38
Figures	39
Figure 0.....	40
Figure 0.....	41
Figure 0.....	42
Figure 0.....	43
Figure 0.....	44
Figure 0.....	45
Figure 0.....	46
Figure 0.....	47
Figure 0.....	48
Figure 0.....	49
Figure 0.....	50
Figure 0.....	51
Figure 0.....	52
Figure 0.....	53
Figure 0.....	54
Figure 0.....	55
Figure 0.....	56
Figure 0.....	57
Figure 0.....	58
Figure 0.....	59
Figure 0.....	60
Figure 0.....	61
Figure 0.....	62
Figure 0.....	63
Figure 1.....	64
Figure 2.....	65
Figure 3.....	66
Figure 4.....	67
Figure 5.....	68
Figure 6.....	69
Figure 7.....	70
Figure 8.....	71
Figure 9.....	72
Figure 10.....	73
Figure 11.....	74
Figure 12.....	75
Figure 13.....	76
Figure 14.....	77
Figure 15.....	78
Figure 16.....	79
Figure 17.....	80
Figure 18.....	81
Figure 19.....	82
Figure 20.....	83
Figure 21.....	84
Figure 22.....	85
Figure 23.....	86
Figure 24.....	87
Figure 25.....	88
Figure 26.....	89

Model-Based Feature Extraction and Classification for Parkinson's Disease Screening using Gait Analysis

Ming De Lim¹; Tee Connie¹ PhD; Michael Kah Ong Goh¹ PhD; Nor 'Izzati Saedon²

¹Faculty of Information Science and Technology Multimedia University Melaka MY

²Department of Medicine, Faculty of Medicine Universiti Malaya Kuala Lumpur MY

Corresponding Author:

Tee Connie PhD

Faculty of Information Science and Technology

Multimedia University

Jalan Ayer Keroh Lama

Melaka

MY

Abstract

Background: Parkinson's Disease (PD) is a progressive neurodegenerative disorder that affects motor coordination, leading to gait abnormalities. Early detection of PD is crucial for effective management and treatment. Traditional diagnostic methods often require invasive procedures or are performed when the disease has significantly progressed. Therefore, there is a need for non-invasive techniques that can identify early motor symptoms, particularly those related to gait.

Objective: The study aims to develop a non-invasive approach for the early detection of Parkinson's Disease by analyzing model-based gait features. The primary focus is on identifying subtle gait abnormalities associated with PD using kinematic characteristics.

Methods: Data was collected through controlled video recordings of participants performing the Time-Up and Go (TUG) assessment, with particular emphasis on the turning phase. The kinematic features analyzed include shoulder distance, step length, stride length, knee and hip angles, leg and arm symmetry, and trunk angles. These features were processed using advanced filtering techniques and analyzed through machine learning methods to distinguish between normal and PD-affected gait patterns.

Results: The analysis of kinematic features during the turning phase of the TUG assessment revealed that individuals with PD exhibited subtle gait abnormalities, such as freezing of gait, reduced step length, and asymmetrical movements. The model-based features proved effective in differentiating between normal and PD-affected gait, demonstrating the potential of this approach in early detection.

Conclusions: This study presents a promising non-invasive method for the early detection of Parkinson's Disease by analyzing specific gait features during the turning phase of the TUG assessment. The findings suggest that this approach could serve as a sensitive and accurate tool for diagnosing and monitoring PD, potentially leading to earlier intervention and improved patient outcomes.

(JMIR Preprints 21/08/2024:65629)

DOI: <https://doi.org/10.2196/preprints.65629>

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

Please make my preprint PDF available to anyone at any time (recommended).

Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.

Only make the preprint title and abstract visible.

✓ **No, I do not wish to publish my submitted manuscript as a preprint.**

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain visible to all users.
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <http://www.jmir.org/preprint/65629>, the full manuscript will be available to all users.



Original Manuscript

Original Paper

Ming De Lim¹, Tee Connie^{1,*}, Michael Kah Ong Goh¹, Nor 'Izzati binti Saedon²

¹ Faculty of Information Science and Technology, Multimedia University, Jalan Ayer Keroh Lama, 75450, Melaka, Malaysia

² Department of Medicine, Faculty of Medicine, Universiti Malaya, 50603, Kuala Lumpur, Malaysia

Corresponding Author:

Tee Connie

Faculty of Information Science and Technology

Multimedia University

Jalan Ayer Keroh Lama

75450, Melaka, Malaysia

Phone: +606-2523592

Email: tee.connie@mmu.edu.my

Model-Based Feature Extraction and Classification for Parkinson's Disease Screening using Gait Analysis

Abstract

Background:

Parkinson's Disease (PD) is a progressive neurodegenerative disorder that affects motor coordination, leading to gait abnormalities. Early detection of PD is crucial for effective management and treatment. Traditional diagnostic methods often require invasive procedures or are performed when the disease has significantly progressed. Therefore, there is a need for non-invasive techniques that can identify early motor symptoms, particularly those related to gait.

Objective:

The study aims to develop a non-invasive approach for the early detection of Parkinson's Disease by analyzing model-based gait features. The primary focus is on identifying subtle gait abnormalities associated with PD using kinematic characteristics.

Methods:

Data was collected through controlled video recordings of participants performing the Time-Up and Go (TUG) assessment, with particular emphasis on the turning phase. The kinematic features analyzed include shoulder distance, step length, stride length, knee and hip angles, leg and arm symmetry, and trunk angles. These features were processed using advanced filtering techniques and analyzed through machine learning methods to distinguish between normal and PD-affected gait patterns.

Results:

The analysis of kinematic features during the turning phase of the TUG assessment revealed that individuals with PD exhibited subtle gait abnormalities, such as freezing of gait, reduced step length, and asymmetrical movements. The model-based features proved effective in differentiating between normal and PD-affected gait, demonstrating the potential of this approach in early detection.

Conclusions:

This study presents a promising non-invasive method for the early detection of Parkinson's Disease by analyzing specific gait features during the turning phase of the TUG assessment. The findings suggest that this approach could serve as a sensitive and accurate tool for diagnosing and monitoring PD, potentially leading to earlier intervention and improved patient outcomes.

Keywords: Model-based Features, Gait Analysis, Parkinson's Disease, Computer Vision, Support

Vector Machines.

Introduction

A common neurological disease that affects millions of individuals worldwide is Parkinson's disease (PD). This disorder gradually impairs a person's ability to move their body, resulting in a variety of crippling symptoms including tremors, stiff muscles, and sluggish motions. To effectively manage PD and provide timely measures, accurate and early diagnosis is essential.

Traditionally, clinical evaluations conducted by medical professionals, which can be arbitrary and inconsistent, have been used to diagnose PD. In distant or underdeveloped areas, access to specialised PD care may be restricted, delaying diagnosis and therapy. The prevalence of PD is predicted to rise as the world's population ages, placing additional demand on healthcare resources and highlighting the need for easily available and effective diagnostic methods.

In this regard, current technology has shown promise in addressing the difficulties associated with PD diagnosis, especially in the areas of machine learning and deep learning. The limits of conventional clinical procedures can be overcome with the promise for early and objective identification provided by machine learning techniques.

This research investigates the use of machine learning for PD identification with a focus on gait characteristics. The goal is to develop a simple and non-invasive method for screening PD symptoms early through analysing gait patterns.

In this study, extracting kinematic features during the turning phase of the Time-Up and Go (TUG) assessment is analysed. Gait variabilities occurred during body turning are more easily identified as turning involves complex motor coordination and balance adjustments, making it a challenging movement. This increased complexity can accentuate subtle abnormalities in gait patterns that might not be as apparent during straight walking. Turning is particularly difficult for individuals with PD, as it requires precise control and stability, often revealing difficulties such as freezing of gait, reduced step length, and asymmetrical movements. Therefore, analyzing the turning phase allows for a more sensitive and accurate detection of PD symptoms, providing a reliable indicator of whether a subject exhibits PD-related gait abnormalities. This focus enhances the effectiveness of the diagnostic tool, which contributes to the facilitation of early and accurate identification of Parkinson's Disease.

Conventional Methods

Challa et al. [1] proposed an advanced predictive model for PD by employing machine learning algorithms such as MLP, Boosted Logistic Regression, Random Forest, and BayesNet. Their investigation applied the PPMI database, an extensive dataset containing data from Parkinson's disease patients and healthy participants. The experimental results showcased remarkable performance improvements over existing methods, boasting accuracy rates of 96.09% for training and 95.45% for testing in the Multilayer Perceptron algorithm, 96.5854% for training and 96.02% for testing in the BayesNet algorithm, 95.45% for training and 94.87% for testing in the Random Forest algorithm, and the highest accuracy achieved by the Boosted Logistic Regression algorithm with 97.159% for training and 96.97% for testing. The area under curve of the ROC curve reached an impressive 98.9% for the Boosted Logistic Regression algorithm, emphasizing its robust predictive capabilities. This research represents a significant stride in healthcare, providing a reliable model for early Parkinson's disease prediction, crucial for timely diagnosis and intervention in addressing this global public health challenge.

In 2019, Polat. [2] investigated the recognition of freezing of gait (FoG) cases in individuals with PS, employing a logistic regression classifier trained and tested on a dataset comprising 16 samples. The study meticulously assessed the classifier's performance using a comprehensive set of 10 performance measures, such as accuracy, miss rate, false discovery rate, false positive rate, false

omission rate, sensitivity, specificity, precision, and negative predictive value. Impressively, the logistic regression classifier exhibited a noteworthy accuracy of 81.3% in accurately classifying FoG cases. The research further compared the performance of linear regression with four alternative models there are Linear Support Vector Machines (SVM), Quadratic SVM, Cubic SVM, and k-Nearest Neighbours (kNN) revealing that the proposed logistic regression model surpassed its counterparts with the highest accuracy of 81.3% in classifying FoG datasets for individuals with Parkinson's disease. This outcome underscores the superior performance of the logistic regression model in the context of FoG classification.

Vidya [3] conducted a comprehensive study on the application of multi-class SVM in using gait analysis to identify and grade the severity of PD. The researchers utilized a publicly accessible dataset containing VGRF and implemented kinematic analysis to extract spatiotemporal features crucial for the diagnostic process. Their suggested framework included a multi-regression strategy to normalise gait time series data and a correlation-based feature selection method. Four distinct SVM kernel functions such as linear, Gaussian, quadratic, and cubic were rigorously evaluated across three different walking tests to gauge their performance. Impressively, the quadratic SVM classifier emerged as the most effective, achieving an outstanding average accuracy of 98.65%. This result surpassed existing state-of-the-art methods, showcasing the robustness and efficacy of the proposed SVM-based approach for PD diagnosis and severity rating.

Besides, Fang. [4] performed a study focusing on predicting PD through the application of machine learning techniques. The research extensively compared the accuracy and recall of three distinct algorithms: KNN, Random Forest, and Naive Bayesian. Addressing the inherent limitations of equal weighting in traditional KNN, the study introduced an entropy weight method to enhance KNN's performance, specifically mitigating equal-weighting issues. Additionally, the research delved into a voice-based UPDRS prediction scheme, leveraging algorithms to effectively predict UPDRS scores from voice data of Parkinson's patients. Notably, the study utilized the UCI dataset, showcasing that the refined KNN algorithm surpassed its traditional counterpart, achieving a notable accuracy rate increase from 91.8% to 93.8%. This research contributes significantly to the advancement of Parkinson's disease prediction methodologies, particularly emphasizing the pivotal role of improved weighting mechanisms in enhancing algorithmic accuracy.

In addition, Gundala et al. [5] conducted a comprehensive study employing the Random Forest algorithm for the recognition of PD using the spiral handwritten dataset. The machine learning technique, widely recognized for its efficacy in processing handwritten designs, was applied by partitioning the dataset into subgroups based on features and constructing decision trees for each feature. The dataset, sourced from the Kaggle website, provided the foundation for training the Random Forest model. Notably, the algorithm's strength lies in its ability to combine outputs from multiple decision trees through majority voting, resulting in a remarkable accuracy rate of 91%. The study emphasizes the robustness of the Random Forest algorithm in accurately identifying Parkinson's disease based on features extracted from handwritten drawings. The results underscore the algorithm's effectiveness in leveraging the diversity of decision trees for enhanced predictive accuracy in the context of Parkinson's disease diagnosis.

Besides, Govindu and Palwa [6] analyzed MDVP audio data collected from both Parkinson's patients and healthy individuals. Among the machine learning models evaluated, the Random Forest classifier emerged as the most effective, achieving a detection accuracy of 91.83% and a sensitivity of 0.95. This model outperformed other techniques such as SVM, KNN and Logistic Regression, which were also assessed for their classification capabilities. The superior performance of the Random Forest classifier highlights its robustness and reliability for detecting PD in its early stages, demonstrating the promise of machine learning in enhancing diagnostic accuracy.

Deep Learning Methods

Pereira et al. [7] proposed an innovative method for early identification of PD using a CNN trained

on handwritten dynamics data obtained by a smart pen. The CNN effectively learned relevant features from the signals generated during individual exams, enabling discrimination between individuals with and without PD based on these learned features. Experimental results showcased the superiority of the CNN over raw data, with the ImageNet architecture, employing 128×128 images and a 75% training dataset split, yielding the best overall accuracy of 83.77%. Despite these promising results, the study acknowledged challenges in achieving consistent recognition rates over control individuals, emphasizing the need for further refinement in the proposed approach.

Later, Grover et al. [8] introduced a deep learning methodology utilizing a Deep Neural Network (DNN) constructed with TensorFlow and Keras, containing 3 hidden layers with 10, 20, and 10 neurons each and an input layer with 16 units, concluding with an output layer representing the classes "severe" and "non-severe." The DNN was trained on the Parkinsons Telemonitoring Voice Dataset from the UCI Machine Learning Repository, comprising biomedical of voice measurements from 42 participants. The experiments focused on predicting the severity of PD based on Total Unified Parkinson's Disease Rating Scale (UPDRS) and Motor UPDRS scores. The DNN exhibited significant accuracy improvements over previous research, achieving a classification accuracy of 94.4422% for Total UPDRS and 83.367% for Motor UPDRS on the training dataset. While test dataset results were comparatively lower, with 62.7335% accuracy for Total UPDRS and 81.6667% for Motor UPDRS, the proposed DNN classifier showcased enhanced performance compared to prior studies, emphasizing its potential for accurate severity prediction in Parkinson's Disease.

Aal et al. [9] conducted an optimized approach for early PD detection, employing speech features extracted from two datasets, DS1 and DS2, containing recordings from both healthy individuals and PD patients. Utilizing Mel-frequency cepstral coefficients (MFCCs) and delta MFCCs, the authors proposed a deep learning model combining a Recurrent Neural Network (RNN) with a Long Short Term Memory (LSTM) layer. Comparative evaluation of alternative ML techniques, including SVM, K-NN and RNN with stochastic gradient descent (SGD), revealed superior performance of the proposed RNN-LSTM model, optimized with the ADAM optimizer. The model exhibited remarkable testing accuracy rates of 95.8% on DS1 and 90.24% on DS2, accompanied by high recall, precision, and F-score on both datasets. Aal et al. further demonstrated the model's superiority over existing methods for PD detection, solidifying its potential as an effective tool for early diagnosis.

Apart from that, Ouhmida et al. [10] delved into the early recognition of voice-based PD through the application of Convolutional Neural Networks (CNN) and Artificial Neural Networks (ANN). Employing two distinct databases, the study showcased the superior accuracy of CNN over ANN in their experiments. "Database I" encompassed 195 voice recordings from 31 individuals, while "Database II" featured 240 recordings from 80 subjects, with a balanced distribution of 40 Parkinson's disease patients and 40 healthy individuals. The deep learning models, specifically CNN and ANN, underwent training and testing on both datasets, revealing CNN's accuracy rates of 93.10% and 88.89% for datasets I and II, respectively. In contrast, the ANN model achieved slightly lower accuracies, with 82.76% for dataset I and 72.22% for dataset II. Ouhmida et al. also highlighted the intricate layers comprising each model, with the ANN model featuring 2 hidden layers and the CNN model incorporating convolution, normalization, activation, softmax, and classification layers. The paper expressed an intent to extend the research by exploring additional deep learning methods and implementing a hybrid system integrating diverse techniques and datasets.

Biswas et al. [11] proposed an approach for the early recognition of PD by proposing two distinct deep learning models tailored for hand-drawn graphics. The first model, a 2D CNN, processed pre-processed images of spirals, circles, and meanders as input, achieving notable accuracies of 83.6% on circles, 61.5% on spirals, and 67.8% on meanders. Another model, an innovative LSTM model, operated on timeline-series signals and demonstrated an overall accuracy of 0.78. Leveraging the NewHandPD dataset for training and testing, the authors conducted four comprehensive experiments, presenting the results. The study posited that early PD detection through these advanced

models could potentially enhance treatment outcomes and elevate the overall quality of life for affected individuals.

Apart from that, Khaskhoussy and Ayed [12] evaluates the effectiveness of SVM CNN for classifying data obtained from speech tasks. Two types of input data were analyzed: the raw speech signal values and i-vector features of dimensions 100, 200, and 300. The classification performance was assessed using five evaluation metrics: accuracy, precision, recall/sensitivity, specificity, and F1-score. For a test dataset of 28 participants, the approach achieved outstanding results, including 100% accuracy, a precision of 0.99, a recall of 0.98, a specificity of 0.96, and an F1-score of 0.98. A table summarizing the state-of-the-art methods is presented in Table 1.

Table 1. A Summary of the State-of-the-Art Methods

Authors	Methods	Input Type	Dataset	Performance
Challa et al. [1]	MLP, BayesNet, RF, Boosted Logistic Regression	Non-motor symptoms	PPMI database	Best overall accuracy: 96.97%
Polat [2]	Logistic Regression	FoG	FoG Dataset (Parkinson's Disease)	81.3%
Vidya [3]	SVM	Gait features	Gait Analysis Dataset	98.65%
Fang [4]	KNN	Voice records	UCI Dataset	93.8%
Gundala et al. [5]	Random Forest	Handwritten drawings	Kaggle Handwritten Drawings Dataset	91%
Pereira et al. [7]	Convolutional Neural Network (CNN)	Handwritten dynamics	Public dataset of handwritten dynamics extracted by a smart pen	Best overall accuracy: 83.77%
Grover et al. [8]	DNN	Voice Data Set	Parkinsons Telemonitoring Voice Data Set	Total UPDRS: Train 94.44%, Test 62.73%; Motor UPDRS: Train 83.37%, Test 81.67%
Aal et al. [9]	RNN-LSTM	Speech features	DS1, DS2	DS1: 95.8%, DS2: 90.24%
Ouhmida et al. [10]	CNN ANN	Voice	Database I, Database II	Dataset I: 93.10%, Dataset II: 88.89%
Biswas et al. [11]	LSTM	Hand Drawings	NewHandPD	78.7%

Methods

The study entailed a methodical approach starting with data collection from young adults, the elderly, and Parkinson's disease patients, followed by obtaining consent, particularly from those diagnosed with the disease. The Time-Up and Go Assessment (TUG) was a key part of the data collection process. Video enhancements and preprocessing were performed to enhance the quality of the videos. After that, the keypoints on the human body are obtained using a human pose estimation technique. Next, features like shoulder distance, step and stride lengths, cadence, and speed were extracted to analyse the gait patterns. The Butterworth filter was applied to refine the data, and peaks were identified to calculate steps and turning durations. Finally, SVM (Polat, 2019) was utilized to distinguish between the different groups based on the extracted features, aiming to improve the prediction and analysis of PD symptoms

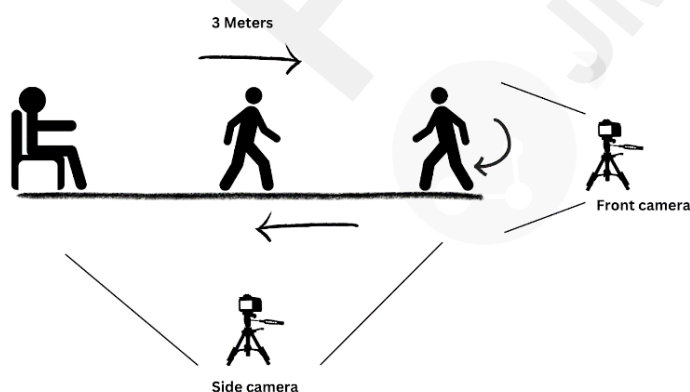
Dataset

The self-collected dataset comprises video recordings of 28 individuals performing the Time-Up and Go (TUG) assessment, a standard test used to evaluate mobility and balance. The dataset comprises three distinct cohorts: young adults age range between 21-33, elderly individuals above the age of 60, and elderly individuals diagnosed with PD. Consents were obtained from the participants prior to data collection. The videos capture various gait patterns, providing a comprehensive set of data for analyzing kinematic features such as step length, stride length, and joint angles. This dataset allows for controlled conditions and detailed annotations, ensuring high-quality data for feature extraction and analysis.

Time-Up and Go Assessment

Fig. 1 shows a TUG workflow, outlining the sequential steps involved in the activity. The process begins with the individual seated at the designated starting point. Upon initiating the recording, the participant stands up and proceeds to walk a distance of 3 meters in a typical manner. Upon reaching the 3 meters mark, the individual executes a turn, walking back to the starting point, where they conclude the TUG process by returning to a seated position. This comprehensive description encapsulates the entire TUG procedure, providing a clear understanding of the task's progression. To enhance data capture, two phones were used - one on the front and the other on the side.

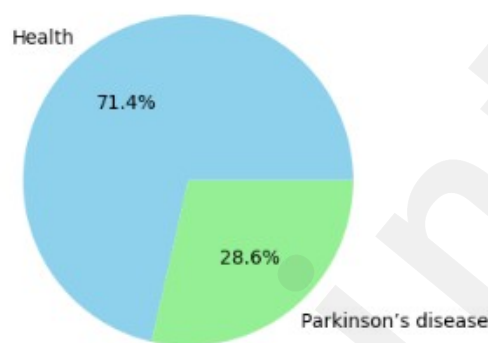
Fig. 1 TUG Shooting Site Configuration Plan



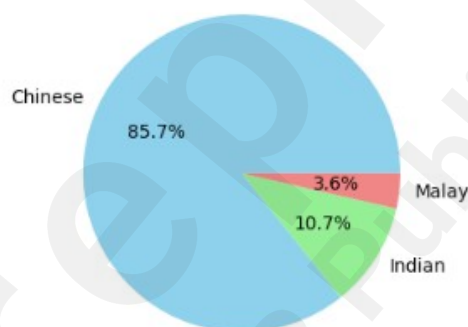
Subject Details

There are three groups of participants in the self-collected data set, namely patients with PD, the elderly and adolescents. The pie charts shown in Fig. 2 provide a comprehensive overview of the

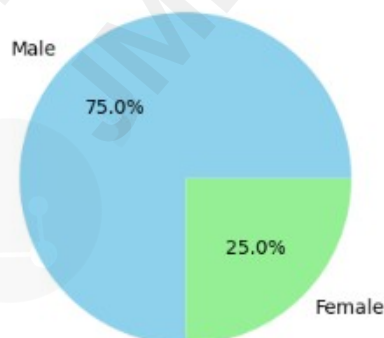
demographic and health-related characteristics of the study subjects. The Health Status Distribution chart shows that a significant portion of the subjects (29%) are diagnosed with PD, while the remaining subjects are either healthy or health. The Race Distribution chart indicates that the majority of the subjects are Chinese (79%), with a smaller representation of Indian (11%) and Malay (4%) individuals. The Gender Distribution chart reveals a higher proportion of male subjects (64%) compared to female subjects (36%). Lastly, the Age Range Distribution chart highlights that most subjects fall within the 60-69 age range (39%), followed by the 20-29 age range (32%), and smaller proportions in other age ranges. These charts collectively illustrate the diversity and focus of the study population.



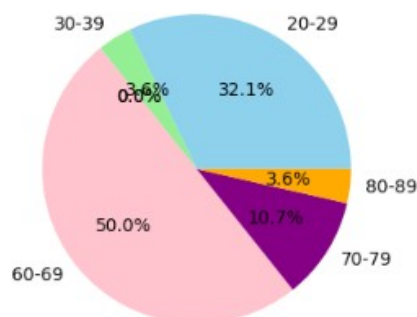
(a) Health Status Distribution



(b) Race Distribution



(c) Gender Distribution



(d) Age Range Distribution

Fig. 2 Demographic Information of the Subjects

Video Enhancement

The presence of background clutter in the original video introduces noise, which significantly affects the accuracy of letter pose extraction. Applying human pose estimation technique directly to unprocessed video may result in inaccurate keypoint coordinates, affecting the reliability of analysis results. To alleviate this problem, a preprocessing step is crucial. Specifically, the video was subjected to a customized processing method in which a black block was used to mask certain areas susceptible to noise. This strategic approach ensures that no noise interferes with the subsequent process, resulting in more precise and reliable coordination extraction. Fig. 3 shows an example for enhancement performed on the video images.

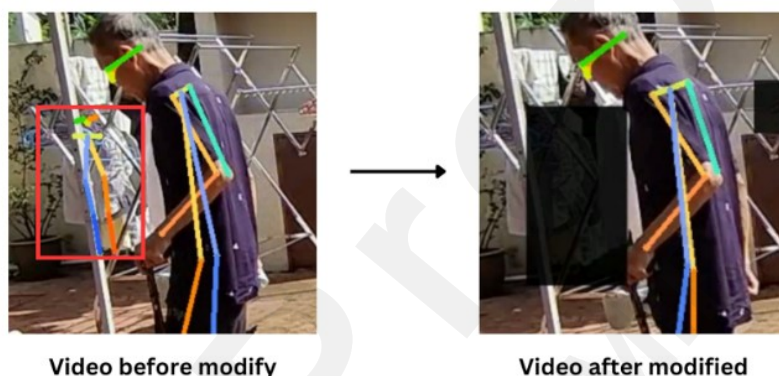


Fig. 3 Video Enhancement.

Human Pose Estimation

In this study, AlphaPose is used to identify 17 key points from the human body, each corresponding to a specific anatomical location. The process yields a JSON file containing coefficients for 17 key points. It is important to note that the JSON file, at this stage, has not undergone data processing. The raw file is populated with various elements, including image_id, category_id, key points, and scores within each set. The image_id represents the frame of the video, while the category_id serves to identify the object (set as 1 for a person). The key points section contains coordinates for body part locations and corresponding detection confidence, formatted as x1, y1, c1, x2, y2, c2, and so forth, with 'c' denoting the confidence score (refer Fig. 4).

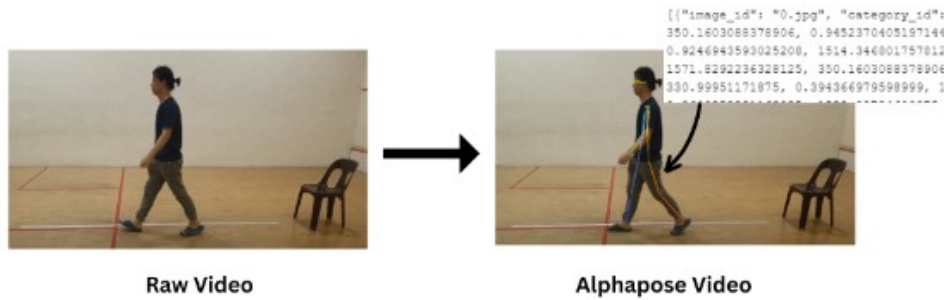


Fig. 4 Gait Analysis Using AlphaPose

Data Extraction

In the data preprocessing phase, the first step involves loading the JSON file and carefully filtering keypoints and image IDs. Additionally, confidence scores are eliminated. The next focus is to mark the 17 identified body key points, including nose, left eye (LEye), right eye (REye), left ear (LEar), right ear (REar), left shoulder (LShoulder), right shoulder (RShoulder), left elbow (LElbow), right elbow (RElbow), left wrist (LWrist), right wrist (RWrist), left hip (LHip), right hip (RHip), left knee (LKnee), right knee (RKnee), left ankle (LAnkle) and right ankle (RAnkle). Next, according to the specific video file name, key point coordinates for each image ID (frame) are marked, ranging from Nose_x, Nose_y to RAnkle_x, RAnkle_y.

The results from the data preprocessing stage boasts a comprehensive structure, encompassing a total of 36 columns that capture key information. Each column is curated to provide a detailed representation of the dataset. Starting with the file_name, which specifies the associated video file, and the image_id indicating the frame or image ID corresponding to each set of keypoints, the subsequent 34 columns focus on the x and y coordinates of 17 distinct body keypoints.

Frames Segmentation

Frames segmentation is a process to assign a time interval to each frame in the video which helps to determine the occurrence of a turning event. The total duration of the video is divided by the total number of frames (rows). This calculation yields the time interval per frame, assuming a constant frame rate throughout the video. With this interval calculated, a new column named 'time_duration' is added to the DataFrame. Each row in the 'time_duration' column is populated with the cumulative time, calculated as the frame's index in the DataFrame multiplied by the time interval. This method provides a time stamp for each frame, which is essential for synchronizing the data with the video and analyzing the timing of the detected turning body events. The pseudocode presented in Algorithm 1 describes the step-by-step procedure for frame segmentation.

Algorithm 1: Pseudocode for Frame Segmentation

Data:

Result: *turning_events*

df

1. **function** frame_segmentation(DataFrame df)
2. **for** each row in df:
3. $shoulder_distance = \sqrt{(shoulder_x - shoulder_x)^2 + (shoulder_y - shoulder_y)^2}$
4. **set** df[shoulder_distance] **to** shoulder_distance
5. **end for**
- 6.
7. **for** each row in df:
8. **if** df[shoulder_distance] > threshold **then**
9. **set** df[turning_event] **to** TRUE
10. **else**

```

11.  set df[turning_event] to FALSE
12.  end if
13.  end for
14.
15.  Return turning_events
16. End function
17.
18. # Main Process
19. DataFrame df = Load data into a DataFrame
20. Float distance_threshold = Set the distance threshold
21.
22. DataFrame turning_events = detect_turning_events(df)
23.
24. Integer total_rows = Length of DataFrame df # Total number of rows
25. Float total_duration_seconds = Set the total duration of the video in seconds
26.
27. Float time_interval = total_duration_seconds / total_rows
28.
29. For each row in DataFrame turning_events:
30.   Calculate time_duration = index_of_row * time_interval
31.   Add time_duration to DataFrame turning_events

```

Noise Filtering

The Butterworth filter is used as the noise filtering tool in this study. Butterworth filter a type of signal processing filter that plays a pivotal role in enhancing the clarity of movement coordinate data extracted from videos. By applying the Butterworth filter to the raw coordinates, high-frequency noise is effectively attenuated, resulting in a smoother trajectory on the graph. This smoothed representation provides a clearer visualization of the subject's movements, reducing interference from irrelevant fluctuations.

Combined with the Butterworth filter, a peak detection algorithm is used to identify important points on the graph. These peaks correspond to key events in motion, such as steps taken by the subject. The peak identification process helps in extracting basic features such as the number of steps throughout the video. This kind of feature extraction is particularly valuable for specific Parkinson's disease aspects of analysis and quantification. The Butterworth filter is expressed using the equation,

$$y(t) = \frac{1}{1 + (\frac{f_c}{f})^{2n}} \cdot x(t) \quad (1)$$

where:

- $y(t)$ is the filtered signal
- $x(t)$ is the input signal
- f_c is the cutoff frequency
- f is the frequency of the signal
- n is the order of the filter

In this study, an order of 4 was used for the Butterworth filters because it provides a smooth yet sufficiently steep roll-off in the transition band. While higher-order filters provide sharper roll-offs, they also require more computational resources. Therefore, a 4th-order filter appears to be a reasonable compromise, being computationally efficient while providing adequate filtering. On the other hand, the peak detection equation is given by,

$$Peaks = \{t_i | y(t_i) \geq height, \forall t_i \in T\} \quad (2)$$

where:

- $y(t)$ is the filtered signal
- T is the set of all time points in the signal $y(t)$

- t_i represents the time point where a peak occurs
- $height$ is the threshold height for peak detection

Only peaks that have an amplitude greater than or equal to the threshold, h , are detected and included in the output. In this study, the value for the height parameter was set to 100. An example of graph for noise filtering with peaks is illustrated in Fig. 5. This example shows the application of Butterworth filter makes the shape of the signal smoother and yields an accurate number of steps during body turning.

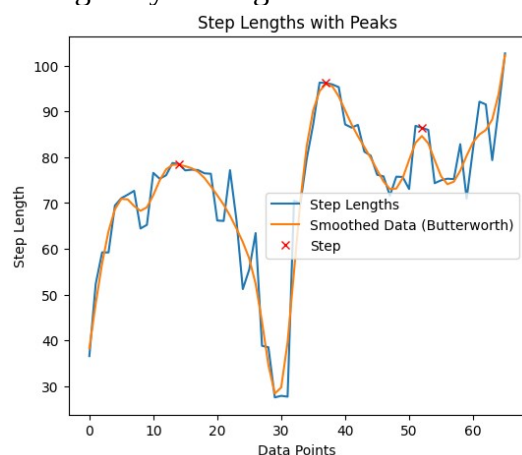


Fig. 5 Example of Graph by Butterworth with Peaks

PD Recognition using Model-based Features

Certain gait features can help identify abnormalities or changes in walking patterns, which are particularly useful for diagnosing and monitoring conditions like PD. By focusing on the body turning period, we aim to capture the most challenging part of the gait cycle, where gait variabilities for PD are more easily observed. Turning involves complex motor coordination and balance adjustments, making it a difficult movement. This complexity can accentuate subtle abnormalities in gait patterns that might not be as apparent during straight walking. Therefore, analyzing gait features during turns provides a more sensitive and accurate assessment of PD-related gait abnormalities, enhancing the effectiveness of early diagnosis and monitoring. The gait features utilized in this study are summarized in Table 2.

Table 2. Gait Analysis Feature Definitions and Formulas

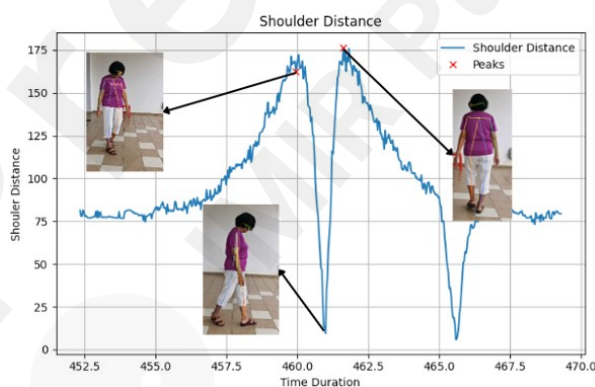
Features	Definition	Formula
Shoulder Distance	Average horizontal distance between the left and right shoulders	$mean(Shoulder_Distance(LShoulder_x, LShoulder_y, RShoulder_x, RShoulder_y))$
Step Length	Average distance between the left ankle and right ankle	$mean(Step_Length(LAnkle_x, LAnkle_y, RAnkle_x, RAnkle_y))$
Stride Length	Average distance covered in one full stride, which	$mean(Stride_Length(RAnkle_x, RAnkle_y))$

	consists of two steps (one by each foot)	
Angle of Both Knee	Degrees at the knee joint by considering the vectors formed by the knee to hip and knee to ankle points.	$\theta_{Knee}(Hip_x, Hip_y, Knee_x, Knee_y, Ankle_x, Ankle_y)$
Angle of Both Hip	Degrees at the hip joint by considering the vectors formed by the hip to knee and hip to shoulder midpoint points.	$\theta_{Hip}(Shoulder_{mid_x}, Shoulder_{mid_y}, Hip_x, Hip_y, Knee_x, Knee_y)$
Symmetric Leg	Degrees for each ankle by considering the vectors formed by the knee to hip midpoint and knee midpoint to hip midpoint points.	$\theta_{Leg}(Knee_{mid_x}, Knee_{mid_y}, Hip_{mid_x}, Hip_{mid_y}, Knee_x, Knee_y)$
Symmetric Arm	Degrees for each arm by considering the vectors formed by the arm to shoulder midpoint and arm midpoint to shoulder midpoint points.	$\theta_{Arm}(Shoulder_{mid_x}, Shoulder_{mid_y}, Elbow_{mid_x}, Elbow_{mid_y})$
Trunk Angle I (Vertical)	Degrees by considering the vertical reference vector and the vector from the hip midpoint to the nose.	$\theta_{Trunk1}(Hip_{mid_x}, Hip_{mid_y}, Nose_x, Nose_y)$

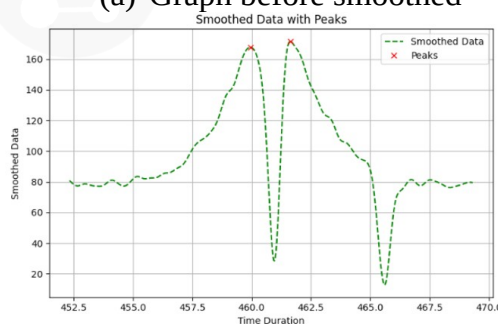
Trunk Angle II (Horizontal)	Degrees by considering the vector from the hip midpoint to the shoulder midpoint and a horizontal reference vector.	$\theta_{Trunk2}(Shoulder_{mid_x}, Shoulder_{mid_y}, Hip_{mid_x}, Hip_{mid_y})$
Shank Angle	Degrees at the hip joint by considering the vectors formed by the left ankle to knee midpoint and right ankle to knee midpoint points.	$\theta_{Shank}(Ankle_x, Ankle_y, Knee_{mid_x}, Knee_{mid_y})$

Turning Body Duration Calculation

Fig. 6 presents two graphs related to the analysis of shoulder distance during a body turn. The left graph displays the original, unprocessed data, illustrating the raw measurements of shoulder distance over time. The fluctuations observed may correspond to the natural movement variations during a turn. In contrast, the right graph exhibits data that has been smoothed using a Butterworth filter, a technique that mitigates short-term fluctuations and reveals the underlying pattern of movement more clearly.



(a) Graph before smoothed



(b) Graph after smoothed

Fig. 6 Shoulder Distance Analysis with Peak Detection

The smoothed data enables effective determination of the peaks of the signal, which identifies the significant turning points. These peaks are highlighted by the red 'X' marks in the figure and represent moments where the shoulder distance reaches its maximum, indicating a complete turn or a change in direction. By focusing on the two highest peaks, the graph underscores the most substantial turning events, thereby minimizing the potential for misinterpreting minor variations as significant movements.

In the context of shoulder distance, $LShoulder$ and $RShoulder$ would correspond to the left and right shoulders, with their respective x and y coordinates. The shoulder distance, $Shoulder_Distance$, which is a measure of how far apart the shoulders are, is calculated as follows,

$$Shoulder_Distance = \sqrt{(RShoulder_x - LShoulder_x)^2 + (RShoulder_y - LShoulder_y)^2} \quad (3)$$

The indexes between these two peaks are then taken to determine the exact time the turn occurred. The difference between these peak times is calculated to find the duration between consecutive turning points. Let $T_{peak,i}$ be the time of the i^{th} peak and n is the total number of peaks detected. Equation (4) calculates the total time duration between the first and the last peak by adding up the durations between each pair of consecutive peaks,

$$Turning_Time = \sum_{i=1}^{n-1} (T_{peak,i+1} - T_{peak,i}) \quad (4)$$

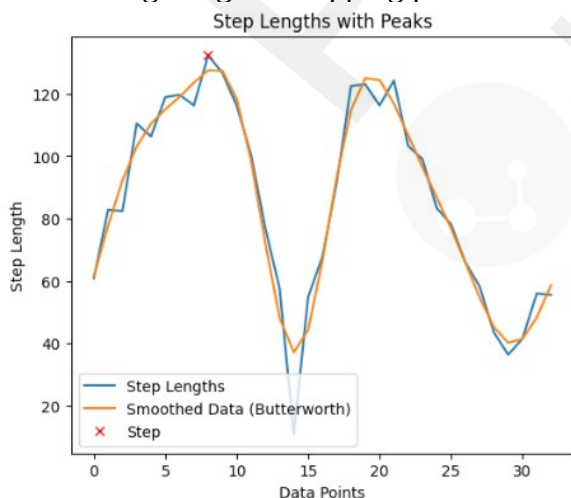
Total Step During Turning Body

Fig. 7 presents three-line graphs which corresponds to the step lengths for different demographic group: Young, Elder, and PD. The step lengths are measured during a turning sequence.

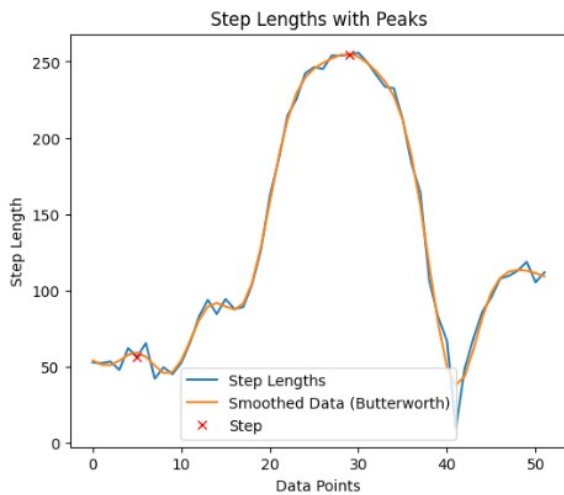
The Young graph shows a more variable stepping pattern with pronounced peaks, suggesting agility and a higher range of motion during turns. The steps are uneven in length, but the peaks return to baseline quickly, indicating swift changes in direction and a potentially more dynamic gait.

In contrast, the Elder graph has fewer, more rounded peaks, indicative of a more cautious or steady turning strategy. The step lengths are generally more uniform, with smoother transitions between steps. This could reflect a more deliberate and potentially less stable gait, as is often seen with aging.

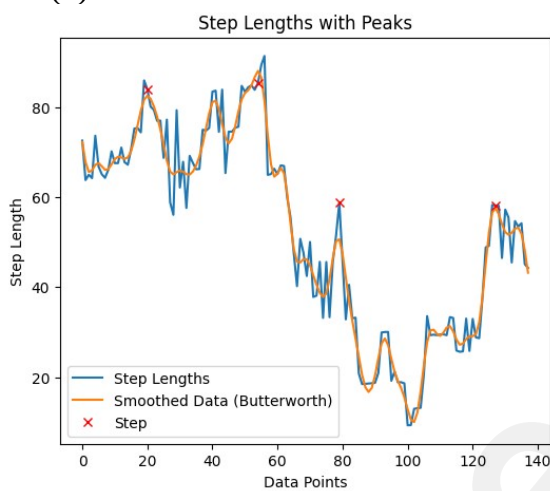
The PD graph shows a significantly different pattern, with smaller step lengths and a higher frequency of steps, which may point to the short, shuffling steps often associated with PD. The peaks are less pronounced and more erratic, highlighting the challenges individuals with PD may face in maintaining a regular stepping pattern.



(a) Young



(b) Elder



(c) PD

Fig. 7 Comparative Step Analysis by Step Length between Groups

Step Length

Step length refers to the distance between the left and right ankle for each step. It is determined by the coordinates of the left ankle ($LAnkle_x, LAnkle_y$) and the right ankle ($RAnkle_x, RAnkle_y$) at a particular instance in time. The formula to calculate step length is provided in Equation (5). This formula gives the straight-line distance between two points in a 2-dimensional space, such as a video frame or a motion capture system's coordinate plane.

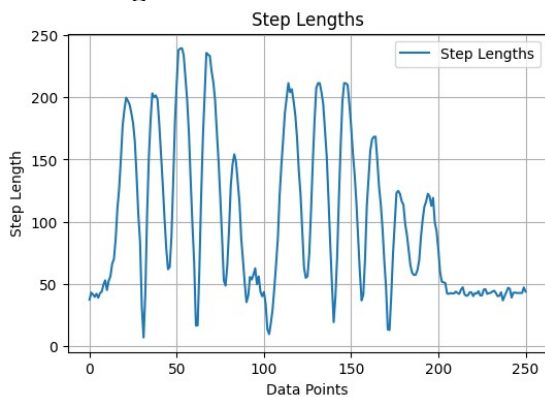
$$Step_Length = \sqrt{(RAnkle_x - LAnkle_x)^2 + (RAnkle_y - LAnkle_y)^2} \quad (5)$$

In Fig. 8, the graphs illustrate the differences in gait patterns across three demographic groups during a TUG test. Each graph plots the variation in step lengths, which are a reflection of the distance between the left and right ankle during individual steps.

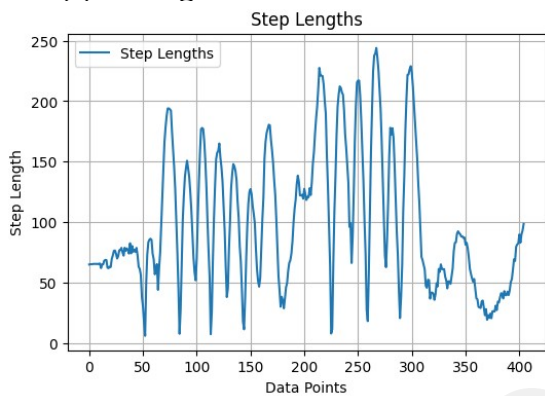
The Young graph is characterized by high peaks and deep valleys, indicating significant variability in step lengths. This could suggest a dynamic and robust gait, with the ability to take both long and short steps, perhaps adjusting speed or direction more frequently.

In contrast, the Elder graph shows a moderate level of variability with somewhat rounded peaks. The reduced height of the peaks compared to the young group suggests shorter steps on average, which may be a sign of a more cautious approach to movement, possibly due to decreased mobility or balance concerns that come with age.

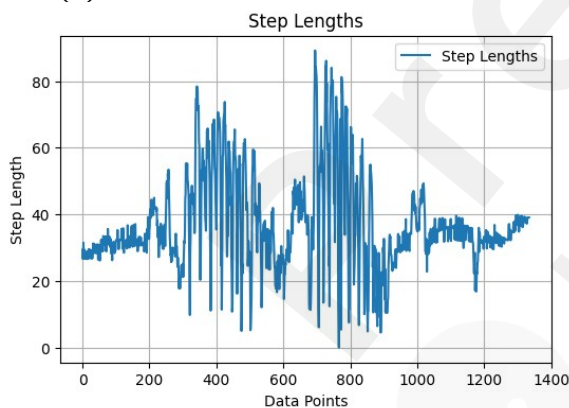
The PD graph differs markedly from the other two. It has a much higher frequency of smaller fluctuations, and the overall range of step lengths is noticeably lower. This pattern is indicative of the short, shuffling steps that are often observed in individuals with PD, reflecting the challenges they face with gait initiation and continuation.



(a) Young



(b) Elder



(c) PD

Fig. 8 Comparative Step Length Analysis between Groups

Stride Length

Stride length is typically defined as the distance covered in one full stride, consisting of two steps (one by each foot). The equation to calculate the stride length for each frame is presented in Equation (6). It calculates the stride length as the Euclidean distance from a point (possibly the origin) to the coordinates of the right ankle for each frame.

$$\text{Stride_Length} = \sqrt{(\text{RAnkle}_x)^2 + (\text{RAnkle}_y)^2} \quad (6)$$

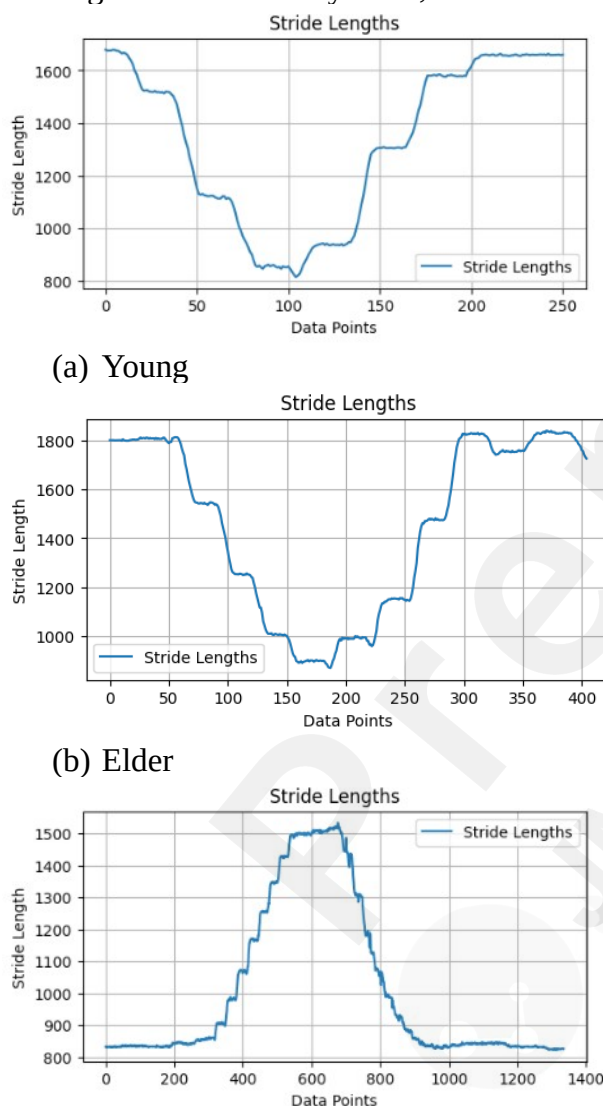
Fig. 9 shows the stride length of the right ankle throughout a single TUG video. The graph exhibits the dynamic nature of stride lengths, with the vertical axis representing the stride length and the

horizontal axis corresponding to data points that likely represent sequential frames of the video.

Noteworthy in this graph is the clear distinction between two types of motion: the rising and falling edges indicate the leg in motion, not in contact with the ground, and the flat regions signify the stance phase where the leg is stationary on the floor, awaiting the next step. The consistency in the flat regions indicates moments when the leg is at rest between strides.

It is evident that the young and elder groups display similar frequencies in stride patterns, whereas the PD group's graph significantly diverges. The PD graph illustrates a strikingly different frequency in stride lengths, standing out from the more consistent rhythmic patterns observed in the young and elder groups.

Moreover, from these charts, it is obvious that when young and old people complete a stride, their feet leave the ground to a greater extent, because a normal stride will take a relatively large range. On the contrary, when patients with PD complete a stride, their feet leave the ground. The amplitude of the ground is relatively short, which can be considered a symptom of PD.



(c) PD
Fig. 9 Comparative Stride Length Analysis between Groups

Angle of Both Knees

By analysing the positions of the hip, knee, and ankle for each leg to determine the angles at both knees can assist in recognizing PD (refer Fig. 10).

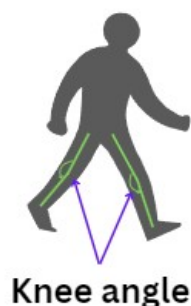


Fig. 10 Illustration of Knee Angle Position

The knee angles provide crucial insights into the degree of bending or extension at the knee joints, which are essential for understanding gait patterns. PD often affects movement and posture, leading to distinctive gait abnormalities. By quantifying these knee angles, the calculation helps in detecting and assessing these gait characteristics, aiding in the identification and monitoring of PD symptoms. Given the knee, hip and ankle joint positions, the knee angle, θ_{Knee} , can be computed as,

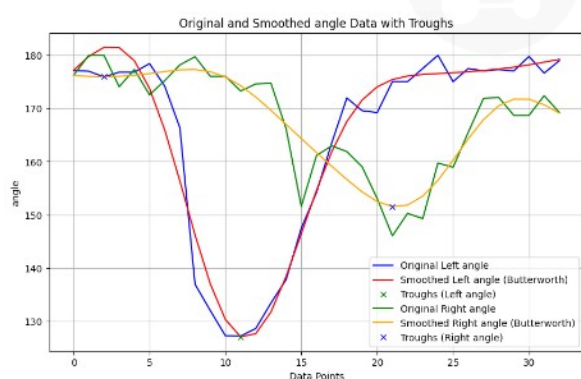
$$\theta_{Knee} = \arccos \left(\frac{(Hip_x - Knee_x) \cdot (Ankle_x - Knee_x) + (Hip_y - Knee_y) \cdot (Ankle_y - Knee_y)}{\sqrt{(Hip_x - Knee_x)^2 + (Ankle_x - Knee_x)^2} \times \sqrt{(Hip_y - Knee_y)^2 + (Ankle_y - Knee_y)^2}} \right) \quad (7)$$

Fig. 11 presents three-line graphs comparing the knee joint angles of the three groups: Young, Elder, and PD. The graph for the young group illustrates smooth and regular oscillations in knee joint angles for both the left and right legs. As they turn their body, the angles show significant flexion and extension, reflecting the youthful ability to execute the turn in a single, fluid step. The peaks and troughs are well-defined, indicating robust and agile movements typical of healthy, young individuals. This pattern highlights their efficient and coordinated gait.

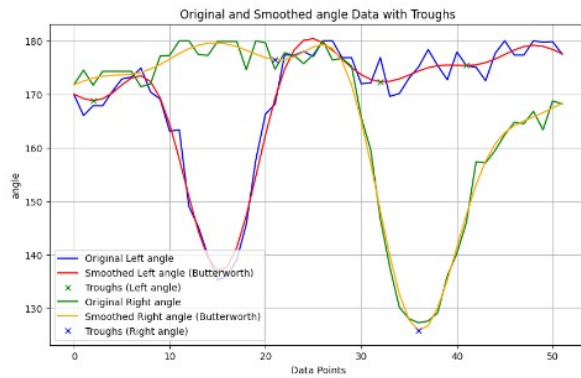
In the elder group's graph, the knee joint angles are more subdued, with two distinct curves for each turn. This indicates that elderly individuals take two steps to complete the body turn, reflecting a more cautious and segmented approach. The movement patterns are slower and less dynamic compared to the young group, with reduced range of motion.

The graph for the PD group shows highly irregular and erratic knee joint angles. The curves exhibit frequent sharp peaks and drops, indicating inconsistent and disrupted movement as individuals with Parkinson's Disease attempt to turn their bodies. These fluctuations reflect the challenges faced by PD patients, including tremors, rigidity, and difficulty maintaining smooth and coordinated movements. The graph captures their struggle to control knee flexion and extension, leading to a more disordered and interrupted gait pattern during the turning motion.

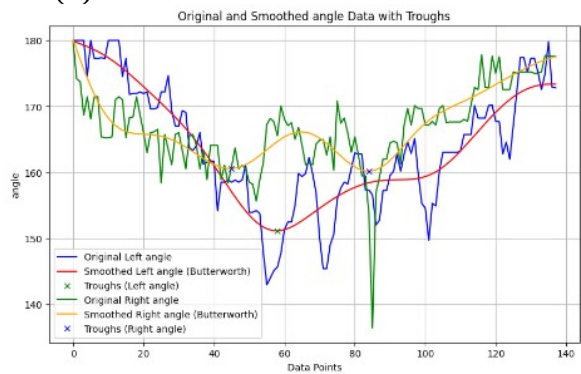
These graphs highlight clear differences in turning movements: young individuals exhibit fluid, single step turns; elders use a careful two-step approach; and PD patients show irregular, disrupted turning patterns.



(a) Young



(b) Elder



(c) PD

Fig. 11 Comparative Knee Angle between Group

Angle of Both Hip

The range of motion (ROM) angle while turning the body is used to determine the angle of both hip (Fig. 12). The positions of the hip, knee, and shoulder are used to find the angle formed at the hip joint. This measurement is crucial for assessing how the hip moves during turns, providing insights into the flexibility and coordination of the lower body. Understanding hip ROM is especially valuable for identifying movement patterns in different populations, such as detecting mobility issues or assessing gait in individuals with PD.



Fig. 12 Illustration of Hip Angle Position

The hip angle is calculated as follows,

$$\theta_{Hip} = \frac{(Knee_x - Hip_x) \cdot (Shoulder_{mid_x} - Hip_x) + (Knee_y - Hip_y) \cdot (Shoulder_{mid_y} - Hip_y)}{\sqrt{(Knee_x - Hip_x)^2 + (Knee_y - Hip_y)^2} \times \sqrt{(Shoulder_{mid_x} - Hip_x)^2 + (Shoulder_{mid_y} - Hip_y)^2}} \quad (8)$$

The graphs showing the cosine of the hip angle are shown in Fig. 12. The figure shows three-line

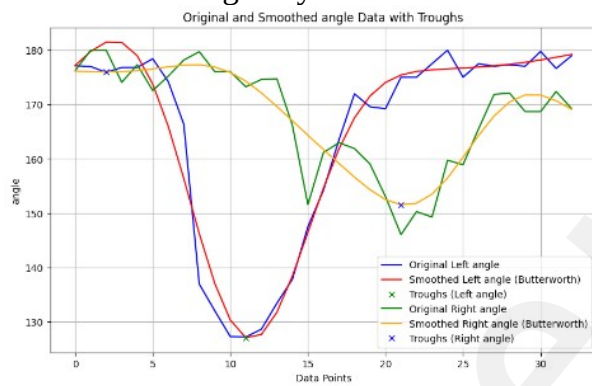
graphs comparing hip angles during body turns for the three groups. Each graph represents the hip angles for both the left and right sides, with lines depicting both the original and smoothed data.

The young group indicates relatively stable and moderate fluctuations in hip angles during body turns. Both hips show consistent movement with shallow peaks and troughs, reflecting balanced and controlled motion. Young individuals maintain a steady range of hip angles, suggesting efficient and coordinated hip movement when turning.

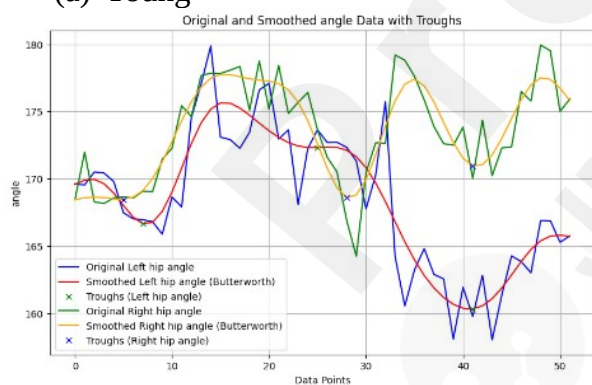
In the elder group's graph, the hip angles display more pronounced and variable fluctuations compared to the young group. The curves reveal that elder individuals experience larger and more variable hip movements during turns. This increased variability suggests that elders adjust their hip movements significantly to maintain balance and stability, resulting in less smooth and more fluctuating hip angles.

The PD group shows significant irregularity and instability in hip angles. The curves exhibit erratic and sharp changes, highlighting the difficulty individuals with PD face in maintaining consistent hip movement during turns. The frequent and abrupt fluctuations in hip angles are characteristic of the disease, where symptoms like rigidity and tremors disrupt smooth and coordinated movements.

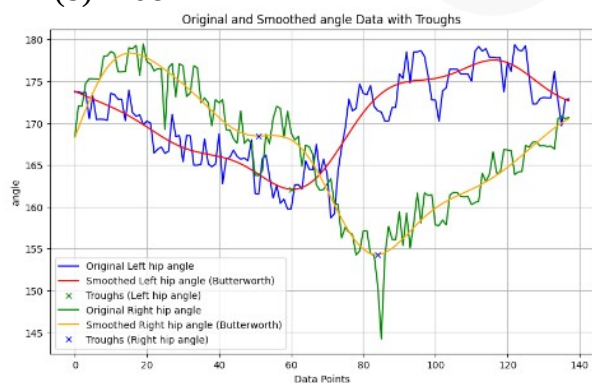
Fig. 13 illustrate clear differences in hip angle dynamics during turns among the three groups. Young individuals have controlled and consistent hip movements, elders show more variability and larger fluctuations in hip angles, and those with PD experience erratic and unstable hip movements. These patterns reflect how age and neurological conditions affect the coordination and efficiency of hip movement during body turns.



(a) Young



(b) Elder



(c) PD

Fig. 13 Comparative Hip Angle between Groups

Leg Symmetry

The symmetry of leg movements is calculated using the angles at the knee joints (Fig. 14). The goal is to measure and evaluate the balance and symmetry in the leg movements which is crucial for understanding gait and stability, especially during turning motions. This information is particularly useful for analysing the movement patterns in different populations, such as detecting gait irregularities in individuals with conditions like PD.

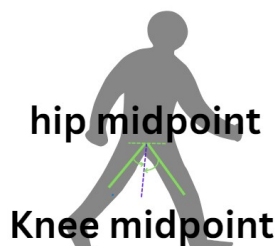


Fig. 14 Illustration of Symmetric Leg Position

The leg angle is computed as,

$$\theta_{Leg} = \frac{(Knee_x - Hip_{mid_x}) \cdot (Knee_{midpoint_x} - Hip_{mid_x}) + (Knee_y - Hip_{mid_y}) \cdot (Knee_{midpoint_y} - Hip_{midpoint_y})}{\sqrt{(Knee_x - Hip_{mid_x})^2 + (Knee_y - Hip_{mid_y})^2} \times \sqrt{(Knee_{midpoint_x} - Hip_{mid_x})^2 + (Knee_{midpoint_y} - Hip_{midpoint_y})^2}} \quad (9)$$

Next, the angles of the left and right legs are compared,

$$difference = |\theta_{Left_Leg} - \theta_{Right_Leg}| \quad (10)$$

The leg symmetry is then assessed as,

$$symmetric = \begin{cases} 1 & \text{if } difference \leq 1 \\ 0.5 & \text{if } 1 < difference \leq 2 \\ 0 & \text{if } difference > 2 \end{cases} \quad (11)$$

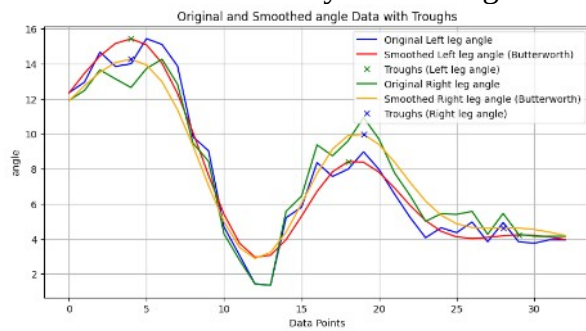
Fig. 15 displays three-line graphs comparing the leg symmetry during body turns for the three participant groups. Each graph plots the symmetrical leg movements, showing the original and smoothed angles of the left and right legs as they turn.

For the young group shows closely aligned curves for the left and right leg angles, indicating high symmetry in their leg movements during turns. Both legs exhibit similar patterns with minimal deviation between them. This close alignment reflects the balanced and coordinated movements typical of young, healthy individuals, suggesting that they maintain a consistent and symmetrical gait.

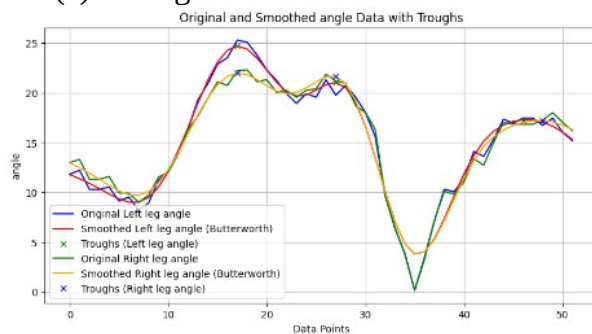
The elder group's graph reveals more noticeable fluctuations between the left and right leg angles. While the overall patterns still follow a similar trajectory.

On the other hand, the PD group's graph shows significant irregularities and less alignment between the left and right leg angles. The curves are erratic and display frequent sharp deviations, reflecting low symmetry in their leg movements. This lack of alignment is characteristic of PD, where symptoms like tremors and rigidity cause disrupted and uncoordinated movements.

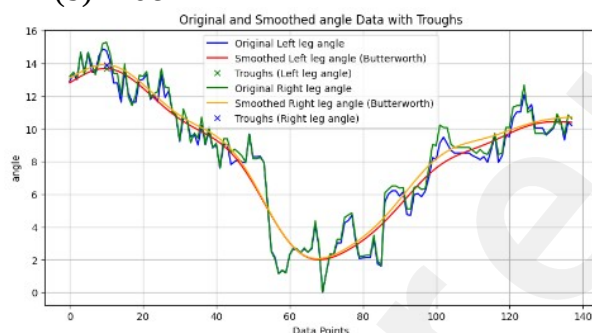
This comparison highlights how age, neurological conditions and health status impact the ability to maintain balanced and symmetrical gait during turning motions.



(a) Young



(b) Elder



(c) PD

Fig. 15 Comparative Symmetric of Leg between Groups

Arm Symmetry

The symmetry of arm movements are analysed by using the angles at the elbows in relation to the shoulders (Fig. 16). The purpose is to measure how balanced and coordinated the arm movements are, especially during activities like turning the body. This assessment is important for understanding the range of motion and the symmetry in arm movements, which can provide insights into overall body coordination and detect possible imbalances or movement disorders.

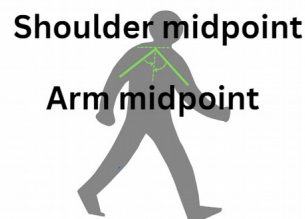


Fig. 16 Illustration of Symmetric Arm Position

The calculation for arm symmetry is similar to that of finding the symmetry in legs,

$$\theta_{Arm} = \frac{(Elbow_x - Shoulder_{midpoint_x}) \cdot (Elbow_{midpoint_x} - Shoulder_{midpoint_x}) + (Elbow_y - Shoulder_{midpoint_y}) \cdot (Elbow_{midpoint_y} - Shoulder_{midpoint_y})}{\sqrt{\left((Elbow_x - Shoulder_{midpoint_x})^2 + (Elbow_y - Shoulder_{midpoint_y})^2 \right) \times \left((Elbow_{midpoint_x} - Shoulder_{midpoint_x})^2 + (Elbow_{midpoint_y} - Shoulder_{midpoint_y})^2 \right)}} \quad (12)$$

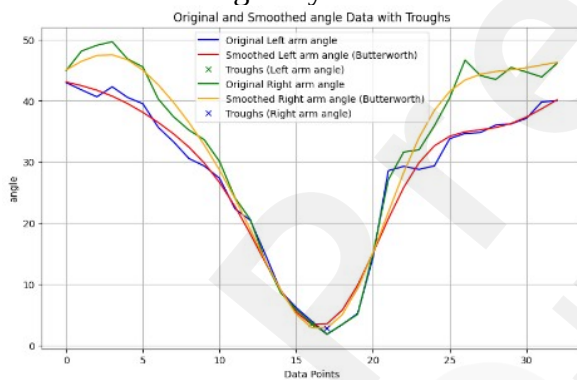
Fig. 17 presents three-line graphs comparing the arm symmetry during body movements for three groups. Each graph plots the angles of the left and right arms over time, showing both the original and smoothed data.

For the young group displays closely aligned curves for the left and right arm angles. Both arms move in a highly synchronized manner, with minimal differences between the left and right angles throughout the motion.

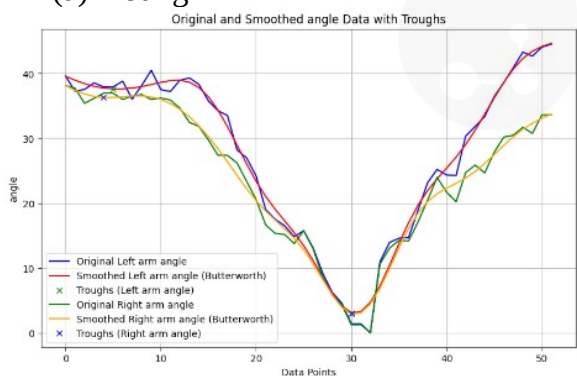
Elder group's graph shows a slightly more varied pattern compared to the young group. While the overall movements of the left and right arms remain relatively synchronized, there are noticeable differences in the magnitude and timing of the angles.

For the PD group exhibits significant irregularities and discrepancies between the left and right arm angles. The curves are erratic and often diverge sharply, indicating substantial asymmetry in arm movements.

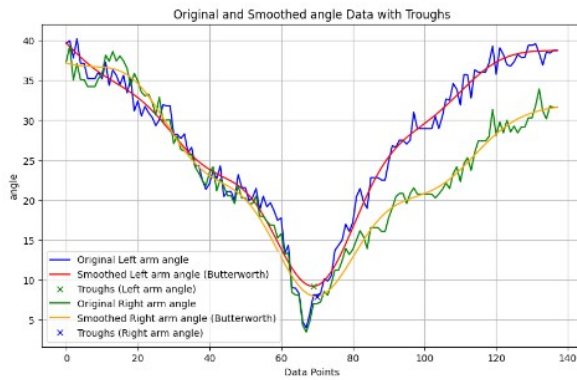
We observe that young individuals demonstrate highly synchronized and symmetrical arm movements, elders show moderate symmetry with some variability, and those with Parkinson's Disease experience significant asymmetry and irregularity in their arm movements. These patterns underscore how age and neurological conditions impact the coordination and balance of arm movements during body turns.



(a) Young



(b) Elder



(c) PD

Fig. 17 Illustration of Trunk Angle I Vertical Position

Trunk Angle I (Vertical)

The relative positions of the hips and nose are used to calculate the trunk angle (Fig. 18). The trunk angle helps to understand how much the upper body tilts or bends relative to the lower body, especially during activities like turning. This angle helps in assessing posture and balance, providing insights into the coordination and alignment of the trunk in relation to the hips during movement. It is particularly useful in evaluating movement patterns and detecting postural deviations in various populations, including those with movement disorders such as PD.

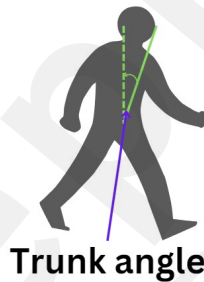


Fig. 18 Illustration of Trunk Angle I Vertical Position

The calculation of the trunk angle I is given as,

$$\theta_{Trunk1} = \frac{(Hip_{midpoint_x} - Hip_{midpoint_x}) \cdot (Nose_x - Hip_{midpoint_x}) + (Hip_{midpoint_y} - 10_y) \cdot (Nose_y - Hip_{midpoint_y})}{\sqrt{(Hip_{midpoint_x} - Hip_{midpoint_x})^2 + (Hip_{midpoint_y} - 10_y)^2} \times \sqrt{(Nose_x - Hip_{midpoint_x})^2 + (Nose_y - Hip_{midpoint_y})^2}} \quad (13)$$

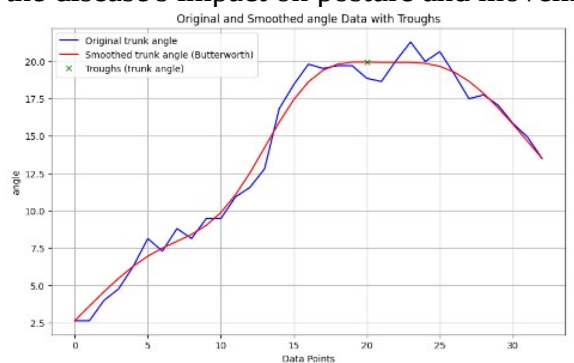
Fig. 19 displays 3 graphs comparing trunk angles relative to the vertical axis for three groups. Each graph shows the trunk angle over a series of data points during body movements, with both original and smoothed data represented. The trunk angle is used to observe the severity of hunchback, providing insights into the degree of forward bending or curvature of the upper body.

For the young group exhibits a consistent and smooth progression of trunk angles. The trunk angle gradually increases and decreases within a narrow range, reflecting a well-coordinated and balanced posture with minimal forward bending. Young individuals maintain a steady alignment with minimal deviations, indicating a lack of significant hunchback severity.

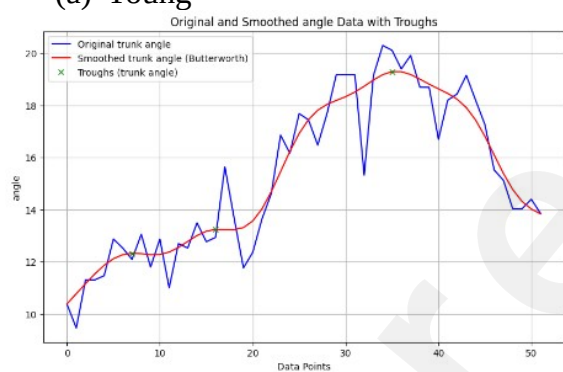
Elder group's graph shows more variability in trunk angles compared to the young group. The curves

display noticeable fluctuations and less consistency, with several abrupt changes in angle. These variations suggest that elder individuals experience more difficulty in maintaining a stable trunk posture, leading to less smooth and more erratic movements. The increased trunk angles and variability may indicate a greater tendency towards forward bending, suggesting a moderate severity of hunchback as they struggle to maintain an upright posture.

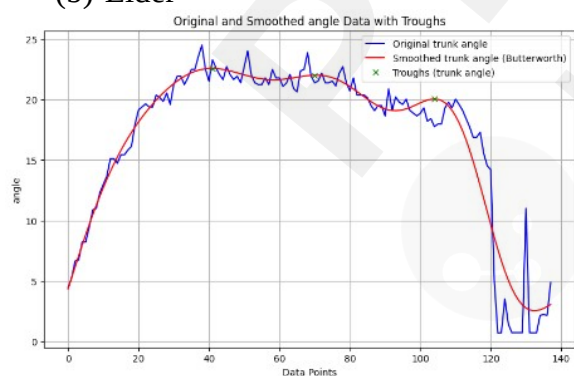
PD group's graph reveals significant irregularities and instability in trunk angles. The curves are highly erratic, with frequent sharp changes and a wide range of deviations from the vertical alignment. These abrupt shifts indicate substantial difficulties in maintaining consistent trunk posture. Individuals with Parkinson's Disease struggle with controlling their trunk movements, leading to frequent tilting and misalignment relative to the vertical axis. The pronounced forward bending and high variability in trunk angles reflect a severe severity of hunchback, characteristic of the disease's impact on posture and movement control.



(a) Young



(b) Elder



(c) PD

Fig. 19 Comparative Trunk Angle I between Groups

Trunk Angle II (Horizontal)

Trunk angle II analyses the positions of the shoulders and hips (Fig. 20). This angle is crucial for understanding the alignment of the trunk relative to the lower body, particularly during movements like turning. By evaluating the trunk angle, we can assess the severity of hunchback or forward bending, which provides valuable insights into posture and stability.

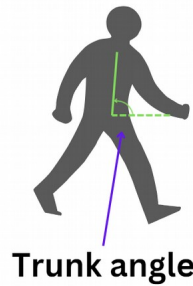


Fig. 20 Illustration of Trunk Angle II Horizontal Position

The calculation for trunk angle II is similar to that for trunk angle I. The only difference is that as trunk angle II is horizontal, the default value must be set at the x coordinate instead of the keypoint. The formula to calculate trunk angle II is given as,

$$\theta_{Trunk2} = \frac{(Shoulder_{midpoint_x} - Hip_{midpoint_x})(10_x - Hip_{midpoint_x}) + (Shoulder_{midpoint_y} - Hip_{midpoint_y})(Hip_{midpoint_y} - Hip_{midpoint_y})}{\sqrt{(Shoulder_{midpoint_x} - Hip_{midpoint_x})^2 + (Shoulder_{midpoint_y} - Hip_{midpoint_y})^2} \times \sqrt{(10_x - Hip_{midpoint_x})^2 + (Hip_{midpoint_y} - Hip_{midpoint_y})^2}} \quad (14)$$

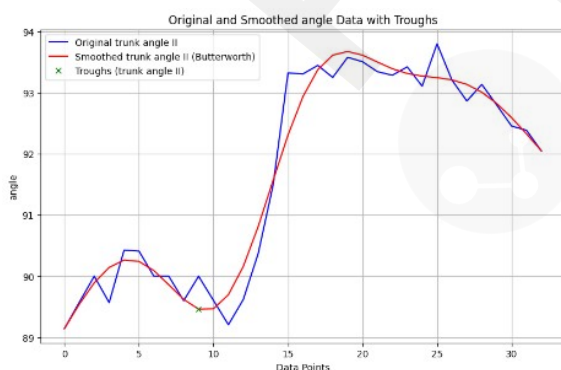
Fig. 21 shows graphs comparing trunk angles relative to the horizontal axis for three groups. These angles help assess the severity of forward bending or misalignment.

The young group exhibits stable and controlled trunk angles with minimal deviations, indicating well-balanced posture and low severity of hunchback.

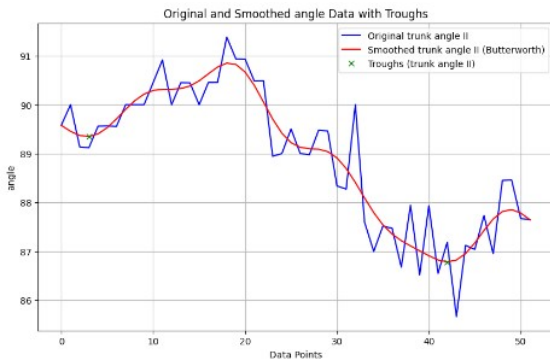
The older group showed more variability and sudden changes in trunk angle, reflecting moderate forward lean and greater difficulty maintaining stable horizontal alignment, indicating moderate kyphosis severity.

The PD group shows significant instability and frequent sharp changes in trunk angles, indicating severe difficulties in maintaining consistent trunk posture and severe hunchback severity due to the disease's impact on movement control.

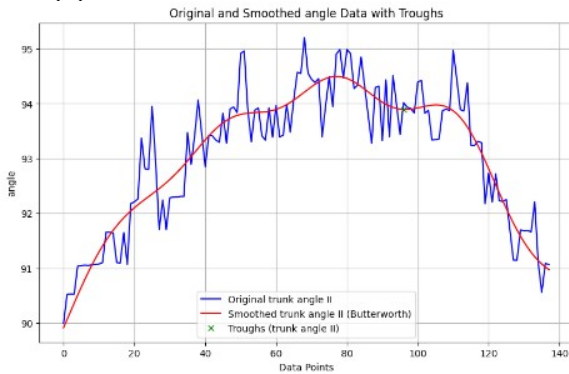
Overall, Young individuals maintain a low severity of hunchback, elders exhibit moderate severity, and those with Parkinson's Disease show severe hunchback, reflecting their challenges in maintaining horizontal alignment.



(a) Young



(b) Elder

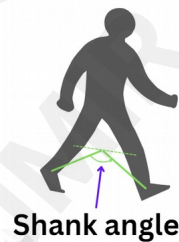


(c) PD

Fig. 21 Comparative Trunk Angle II Horizontal between Groups

Shank Angle

The shank angle computes the positions of the knees and ankles (Fig. 22). This angle is useful to understand the alignment and movement of the lower leg, especially during activities like turning. By evaluating the shank angle, we can assess the coordination and stability of the lower leg movements, providing valuable insights into gait patterns and detecting any abnormalities in movement, which is particularly useful for analysing conditions such as PD.



Shank angle

Fig. 22 Illustration of Shank Angle Position

The formula to calculate shank angle is provided by,

$$\theta_{Shank} = \frac{(Ankle_x - Knee_{midpoint_x}) \cdot (Ankle_x - Knee_{midpoint_x}) + (Ankle_y - Knee_{midpoint_y}) \cdot (Ankle_y - Knee_{midpoint_y})}{\sqrt{(Ankle_x - Knee_{midpoint_x})^2 + (Ankle_y - Knee_{midpoint_y})^2} \times \sqrt{(Ankle_x - Knee_{midpoint_x})^2 + (Ankle_y - Knee_{midpoint_y})^2}} \quad (15)$$

Fig. 23 presents the graphs comparing shank angles during body movements for three groups. Each graph shows the shank angle over a series of data points, with both original and smoothed data represented.

For the young group demonstrates relatively stable and consistent shank angles. Both the original and smoothed curves align closely, indicating smooth and coordinated movements. The angles show

moderate fluctuations within a narrow range, reflecting balanced and controlled lower leg movements typical of healthy young individuals.

The elder group's graph shows shank angles that are stable and consistent, like the young group. Both the original and smoothed curves align closely, indicating smooth and coordinated lower leg movements. The angles show moderate fluctuations within a narrow range, reflecting balanced and controlled movements. This suggests that elder individuals perform nicely, maintaining stable shank angles comparable to the young group.

The PD group's graph reveals significant irregularities and instability in shank angles. The curves are highly erratic, with frequent sharp changes and a wide range of deviations. These abrupt shifts indicate substantial difficulties in maintaining consistent lower leg movements. Individuals with PD struggle with controlling their shank angles, leading to frequent misalignment and instability. This pattern is characteristic of the disease, where symptoms such as rigidity and tremors disrupt smooth and coordinated movements.

Fig. 23 highlights the differences in shank angle behavior among the three groups. Young and elder individuals exhibit stable and coordinated shank movements with minimal variability, indicating balanced and controlled lower leg movements. In contrast, those with PD display significant instability and erratic shank movements, underscoring the impact of the condition on lower leg control and stability.

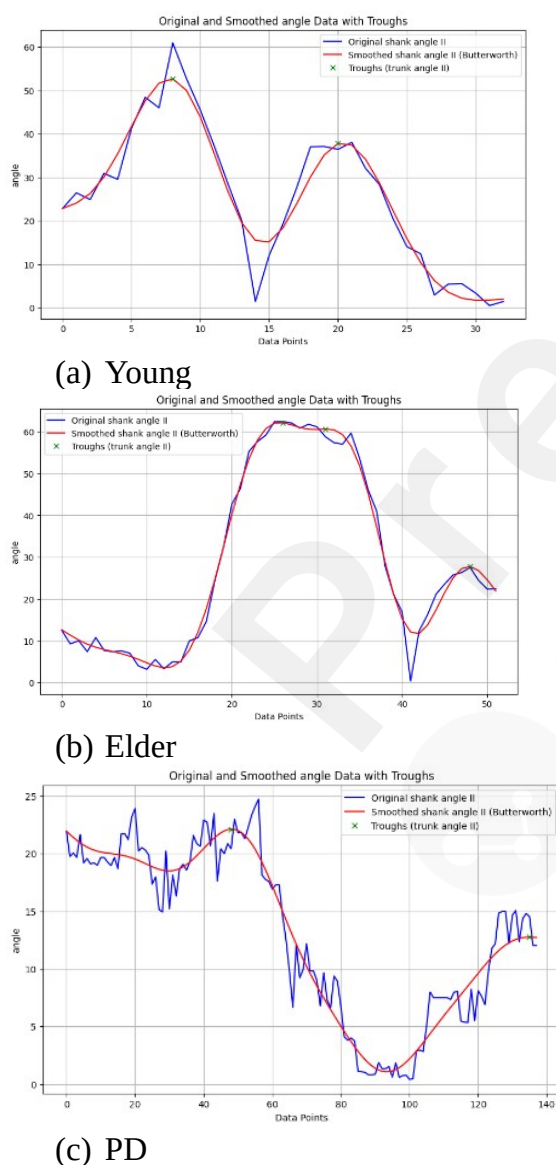


Fig. 23 Comparative Shank Angle between Groups

Results

This section presents the experiment results and discussion for the gait features extracted from the Self-Collected Dataset, focusing on the effectiveness of these features in distinguishing between normal and PD-affected gait patterns using a Support Vector Machine (SVM) classifier.

For the experiments, PCA was applied to reduce the dimension of the dataset. The number of components for PCA was set to 8, corresponding to the number of subjects with PD. This step helps in capturing the most significant features while reducing computational complexity.

The dataset was additionally divided in a 70:30 ratio. This indicates that the model was trained using 70% of the data, with the remaining 30% set aside for testing and performance evaluation. By splitting the data this way, a significant piece of the data is used to train the model while maintaining enough for a thorough assessment. Ten trials of the experiments had been performed and the average results were recorded.

Classification Report

A comprehensive evaluation of the SVM model is provided by the classification report for the test set, as shown in Table 3. The report includes the overall accuracy of the model, as well as each class's recall, precision, and F1-score. Although the training accuracy is 0.95, the test accuracy is 0.89, indicating that the model fits the training data well with minimal overfitting. The classification report shows that class 0 (normal class) has high precision and recall, while class 1 (PD class) has lower recall. This disparity in recall may be due to the smaller sample size of class 1.

Table 3. Classification Report for Distinguishing Normal and PD Classes using Gait Features

Test Accuracy: 0.89				
Train Accuracy: 0.95				
Classes/Metrics	Precision	Recall	F1-Score	Support
0 (Normal)	0.88	1.00	0.93	7.00
1 (PD)	1.00	0.50	0.67	2.00
Accuracy			0.89	9.00
Macro Avg	0.94	0.75	0.80	9.00
Weighted Avg	0.90	0.89	0.87	9.00

Area under Curve (AUC) and Loss Curve

Fig. 24 provides specifics regarding the model's capacity to distinguish between the two classes. Moderate discrimination capability is indicated by the test set's AUC of 0.6429. An indicator of how well the model performs across various training data subsets is provided by the cross-validation scores. With a mean cross-validation score of 0.9000, the model's cross-validation scores are [0.75, 1.00, 1.00, 0.75, 1.00]. This shows that the model continues to work well when the data is folded in different ways.

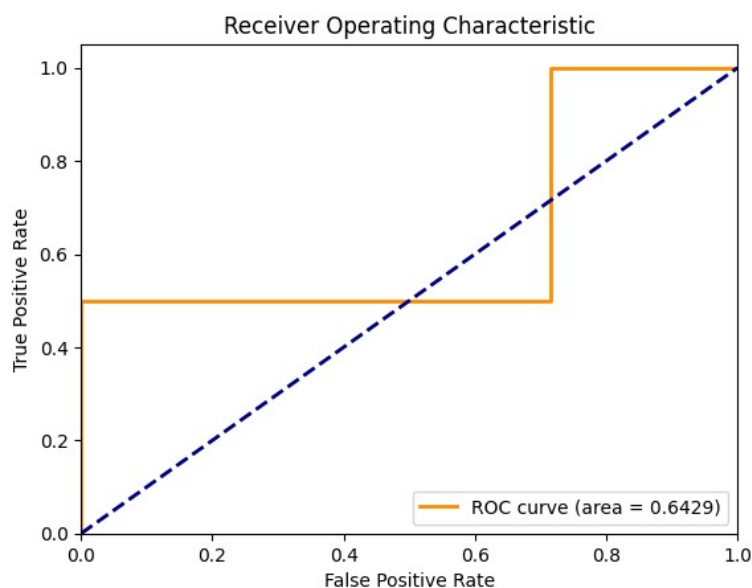


Fig. 24 Receiver Operating Characteristic (ROC) Curve.

Fig. 25 shows the train/validation loss curve, illustrating the log loss for both training and validation sets across various folds of cross-validation. The blue curve indicates the training loss, which stays relatively stable with minor fluctuations. In contrast, the red curve, representing the validation loss, exhibits significant fluctuations, peaking notably at fold 3. This implies that while the model fits the training data well, it experiences considerable variability in performance on unseen data, suggesting possible overfitting or sensitivity to specific data subsets. The difference between the training and validation loss also underscores the model's difficulty in generalizing from the training set to the validation set, highlighting the need for potential adjustments in model complexity or training strategy to achieve more consistent performance.

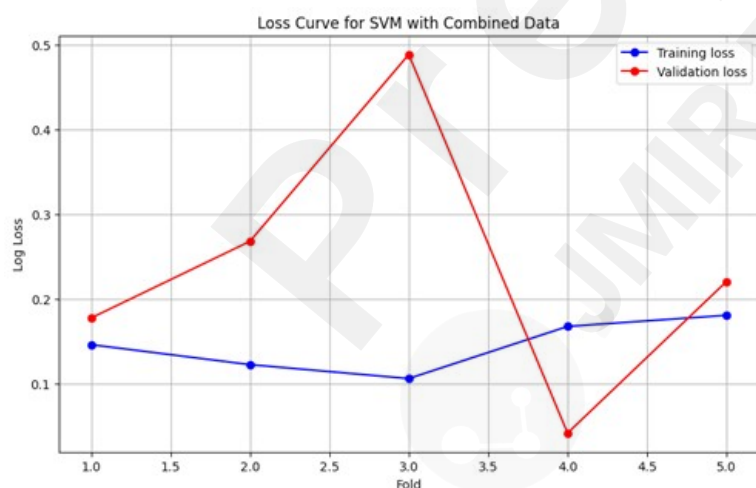


Fig. 25 Train/Validation Loss Curve.

Confusion Matrix

The confusion matrix depicted in Fig. 26 provides a thorough analysis of the predictions of the model on the test set. It reveals that while the model correctly predicted 7 out of 7 occurrences of class 0 (NP), it incorrectly classified 1 out of 2 instances of class 1 (PD). This suggests that although the model is highly accurate in predicting class 0, it faces challenges with class 1, likely due to class imbalance in the dataset.

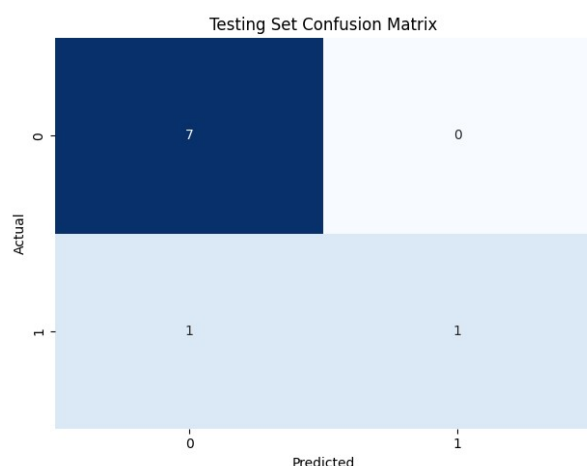


Fig. 26 Testing Set of Confusion Matrix for Model-based Features

Discussion

Conclusions

This study presents a non-invasive approach for early detection of PD through the analysis of model-based gait features. By utilizing kinematic characteristics such as shoulder distance, step length, stride length, knee and hip angles, leg and arm symmetry, and trunk angles, we aimed to identify subtle gait abnormalities associated with PD. Data was collected through controlled video recordings of the TUG assessment, and the extracted features were processed using advanced filtering techniques and analyzed using SVM classifier.

The results demonstrated that the model-based features are highly effective in distinguishing between normal and PD-affected gait patterns, achieving high accuracy, precision, recall, and F1-score. These findings support the potential of model-based gait analysis as a non-invasive, accessible tool for early diagnosis and monitoring of PD. The study also highlighted the importance of addressing class imbalance and refining the feature extraction process to further enhance the performance of the classifier.

Future research should focus on validating these findings with larger datasets, exploring other machine learning models, and integrating additional features to improve the robustness and accuracy of PD detection systems. This approach could significantly benefit clinical settings, providing a reliable, non-invasive method for early diagnosis and improved patient outcomes.

Acknowledgements

This research is supported by the Fundamental Research Grant Scheme (FRGS/1/2020/ICT02/MMU/02/5) and Malaysian-Jordan Matching Fund.

Conflicts of Interest

none declared

References

1. Challa, K. N., Pagolu, V. S., Panda, G., & Majhi, B. An improved approach for prediction of parkinson's disease using Machine Learning Techniques. 2016 International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES). 2016; doi: 10.1109/scopes.2016.7955679

2. Polat, K. 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). 2019; doi: 10.1109/ebbt.2019.8742042
3. Vidya, B., & P, S. Gait based parkinson's disease diagnosis and severity rating using multi-class support vector machine. *Applied Soft Computing*. 2021; 113:107939. doi: 10.1016/j.asoc.2021.107939
4. Fang, Z. Improved KNN algorithm with information entropy for the diagnosis of parkinson's disease. 2022 International Conference on Machine Learning and Knowledge Engineering (MLKE). 2022; doi: 10.1109/mlke55170.2022.00024
5. Gundala, S., Harichandana, C., Jaasmitha, A., Brundha, P. L. Parkinson's disease identification by voice and handwritten drawings using Xgboost and random forest algorithms. 2023 3rd International Conference on Computing and Information Technology (ICCIT). 2023; doi: 10.1109/iccit58132.2023.10273905
6. Govindu, A., Palwe, S. Early detection of Parkinson's disease using machine learning. *Procedia Computer Science*. 2023; 218: 249-261. doi: 10.1016/j.procs.2023.01.007.
7. Pereira, C. R., Weber, S. A., Hook, C., Rosa, G. H., & Papa, J. P. Deep learning-aided parkinson's disease diagnosis from handwritten dynamics. 2016 29th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI). 2016; doi: 10.1109/sibgrapi.2016.054
8. Grover, S., Bhartia, S., Akshama, Yadav, A., & K.R., S. Predicting severity of parkinson's disease using Deep Learning. *Procedia Computer Science*. 2018; 132:1788–1794. doi: 10.1016/j.procs.2018.05.154
9. Aal, H. A., Taie, S. A., & El-Bendary, N. An optimized RNN-LSTM approach for parkinson's disease early detection using speech features. *Bulletin of Electrical Engineering and Informatics*. 2021; 10(5):2503–2512. doi: 10.11591/eei.v10i5.3128
10. Ouhmida, A., Terrada, O., Raihani, A., Cherradi, B., & Hamida, S. Voice-based deep learning medical diagnosis system for parkinson's disease prediction. *International Congress of Advanced Technology and Engineering (ICOTEN)*. 2021; doi: 10.1109/icoten52080.2021.9493456
11. Biswas, S., Kaur, N., & K.R, S. Early detection of parkinson's disease from hand drawings using CNN and LSTM. 2022 4th International Conference on Artificial Intelligence and Speech Technology (AIST). 2022a; doi: 10.1109/aist55798.2022.10065159
12. Khaskhoussy, R., Ayed, Y.B. Improving Parkinson's disease recognition through voice analysis using deep learning. *Pattern Recognition Letters*. 2023; 168: 64-70. doi: 10.1016/j.patrec.2023.03.011.

Supplementary Files

Figures

Fig. 13 Comparative Hip Angle between Groups. (c) PD.

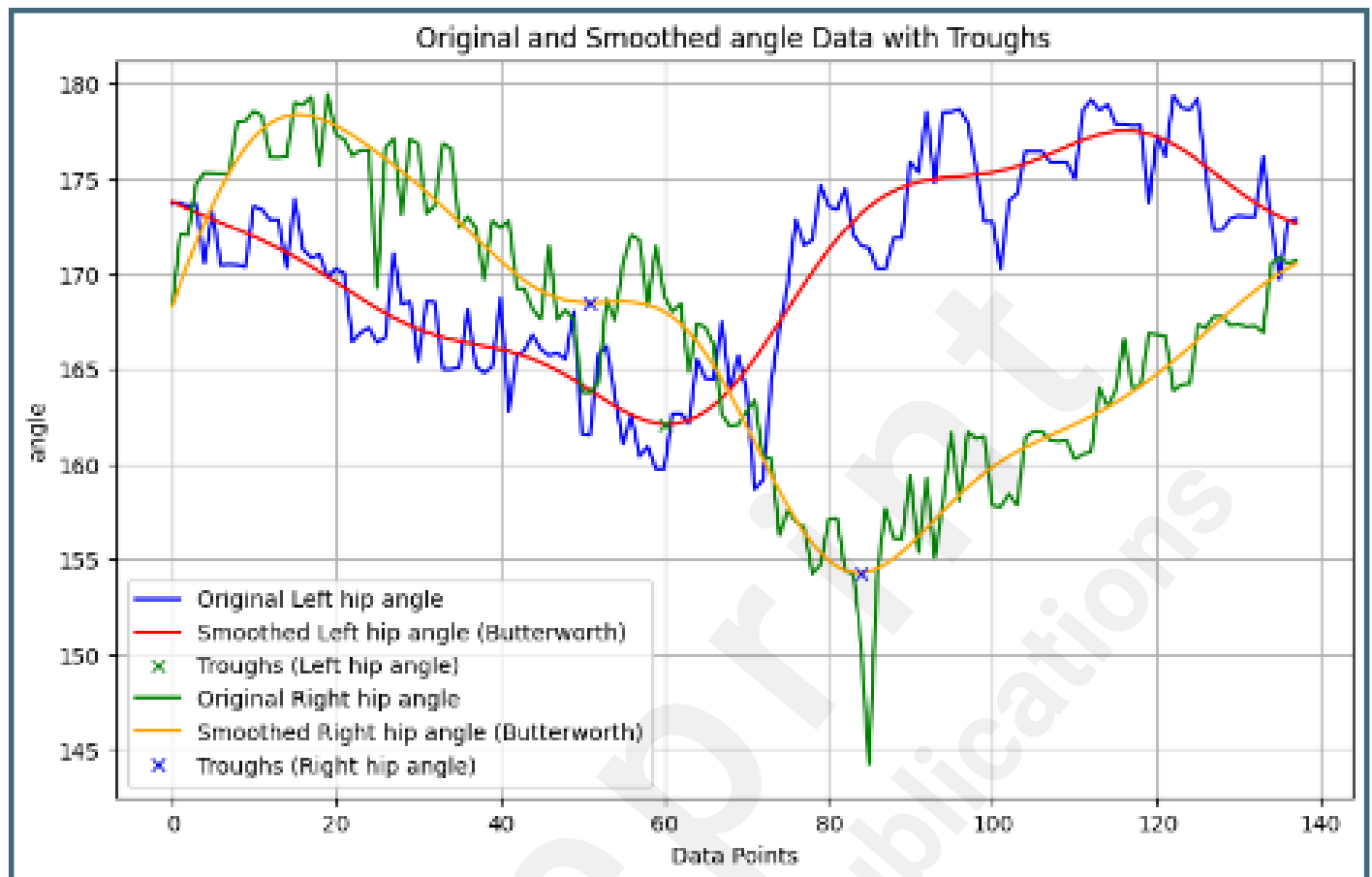


Fig. 19 Comparative Trunk Angle I between Groups. (b) Elder.

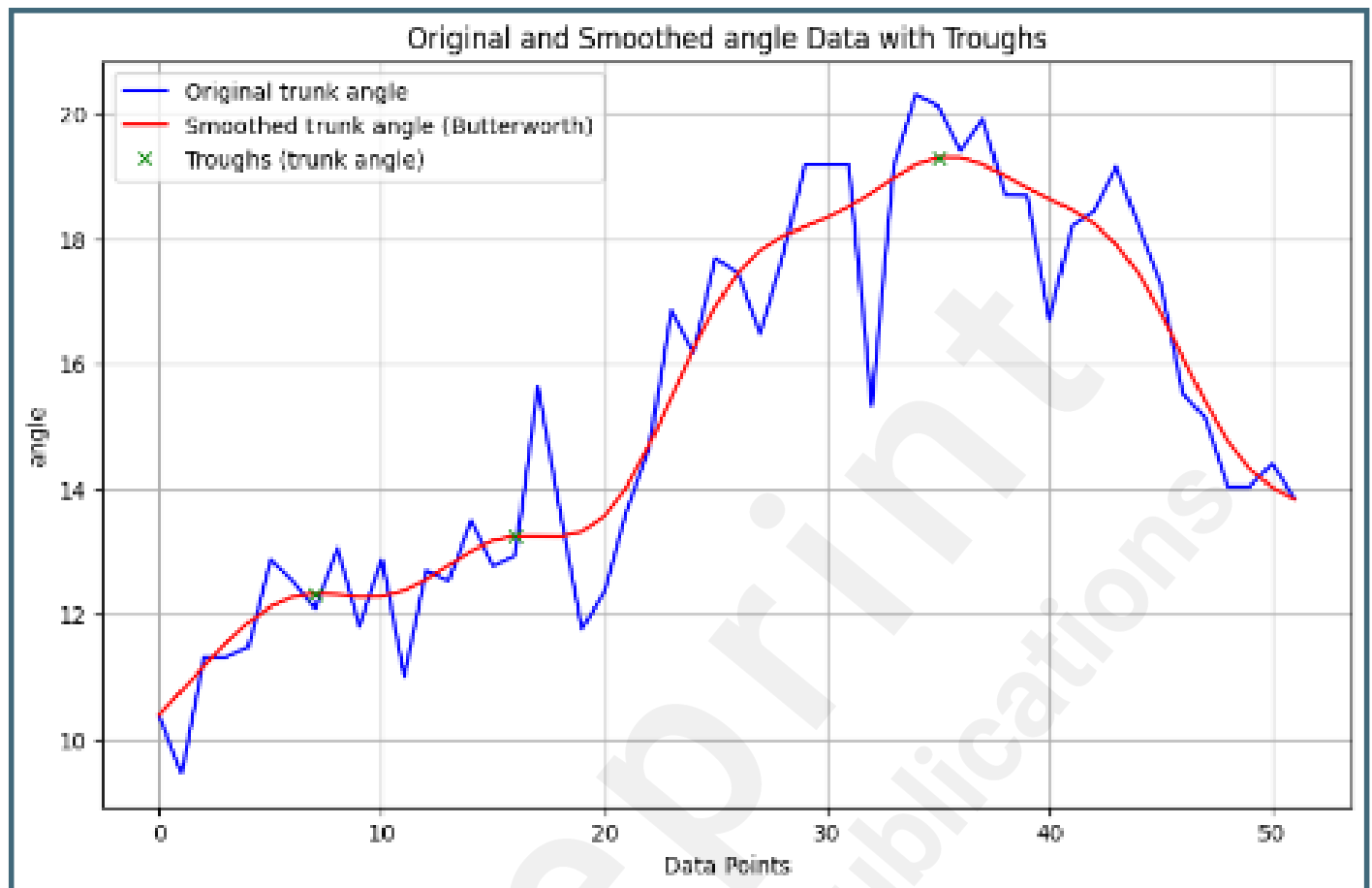


Fig. 11 Comparative Knee Angle between Group. (c) PD.

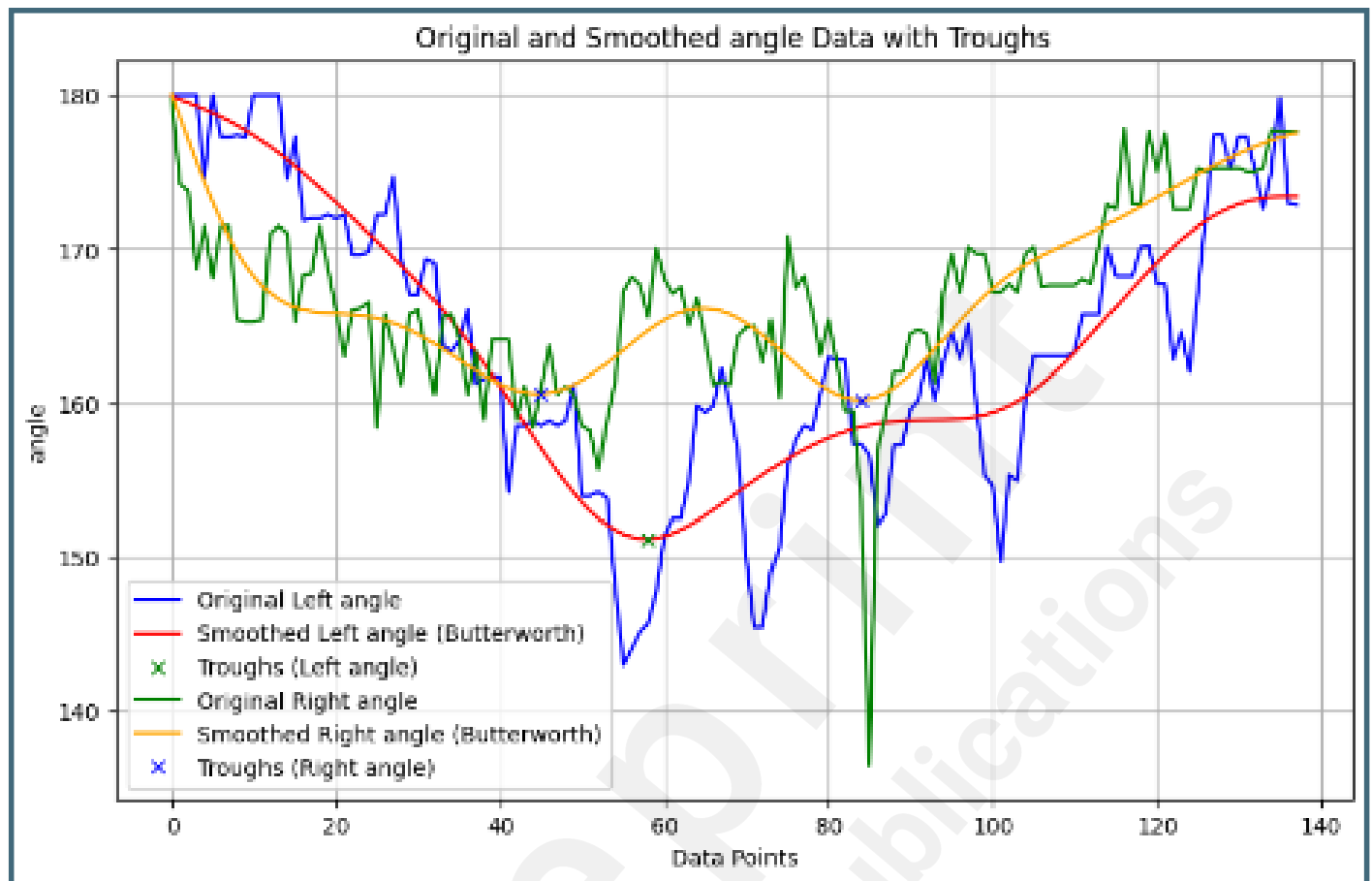


Fig. 11 Comparative Knee Angle between Group. (b) Elder.

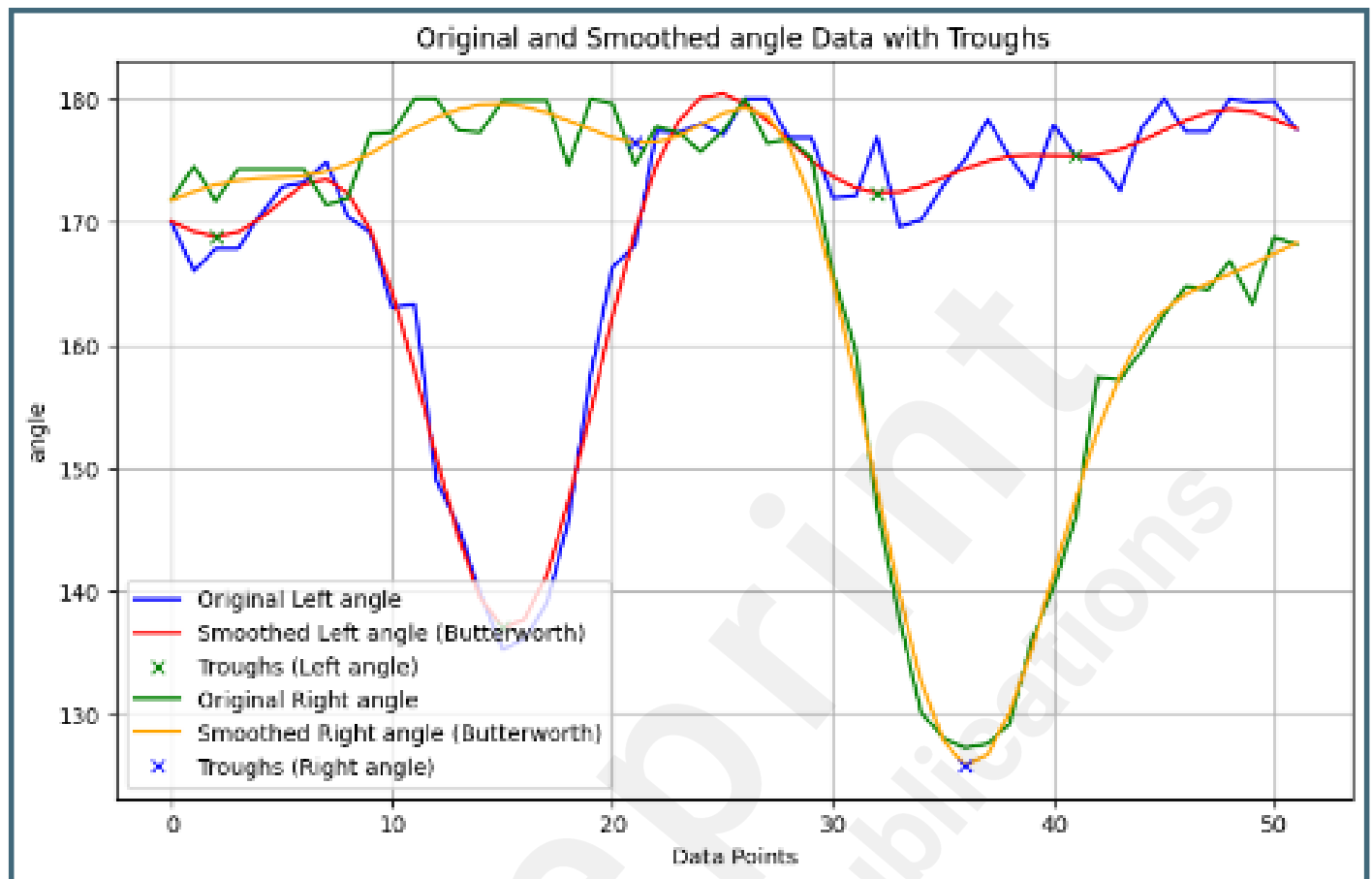


Fig. 15 Comparative Symmetric of Leg between Groups. (c) PD.

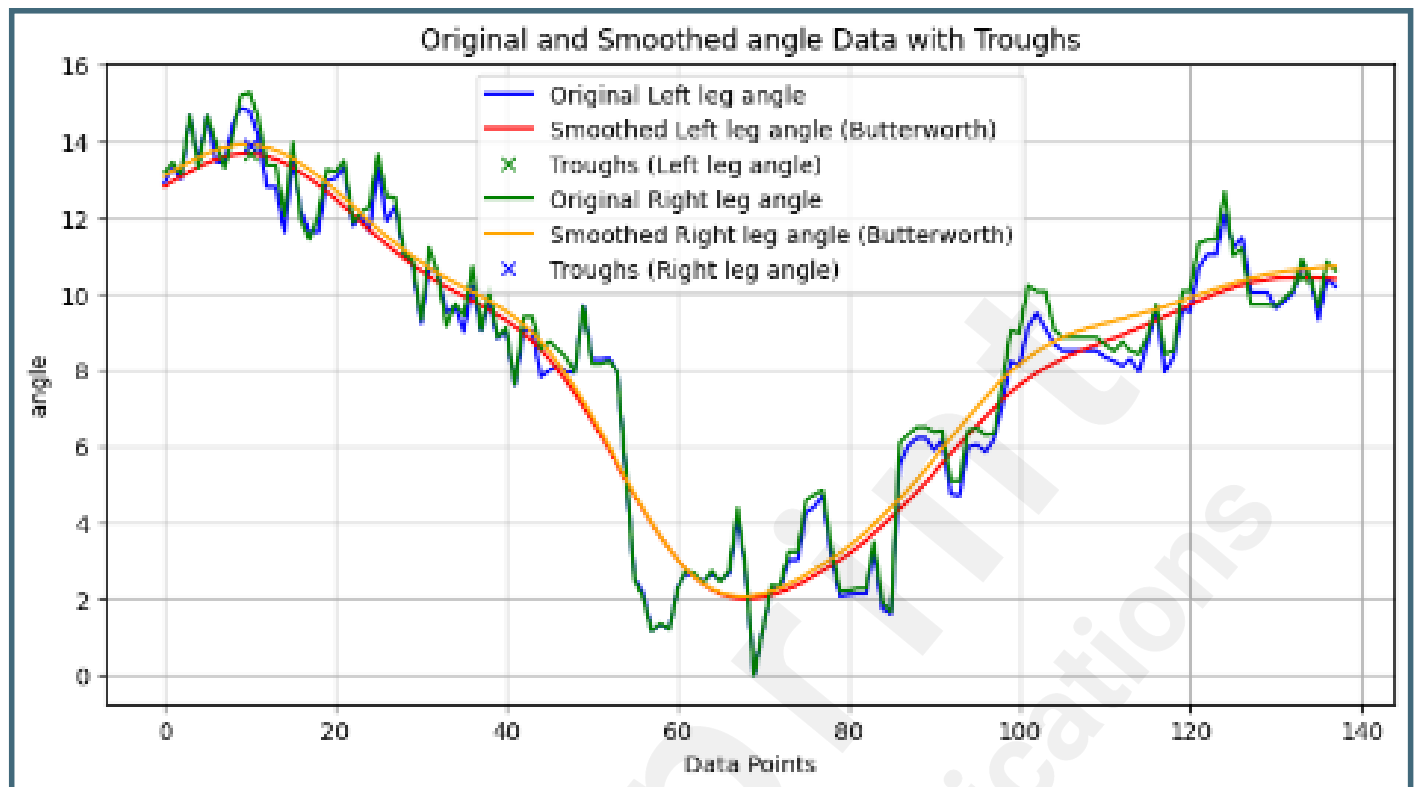


Fig. 17 Illustration of Trunk Angle I Vertical Position. (b) Elder.

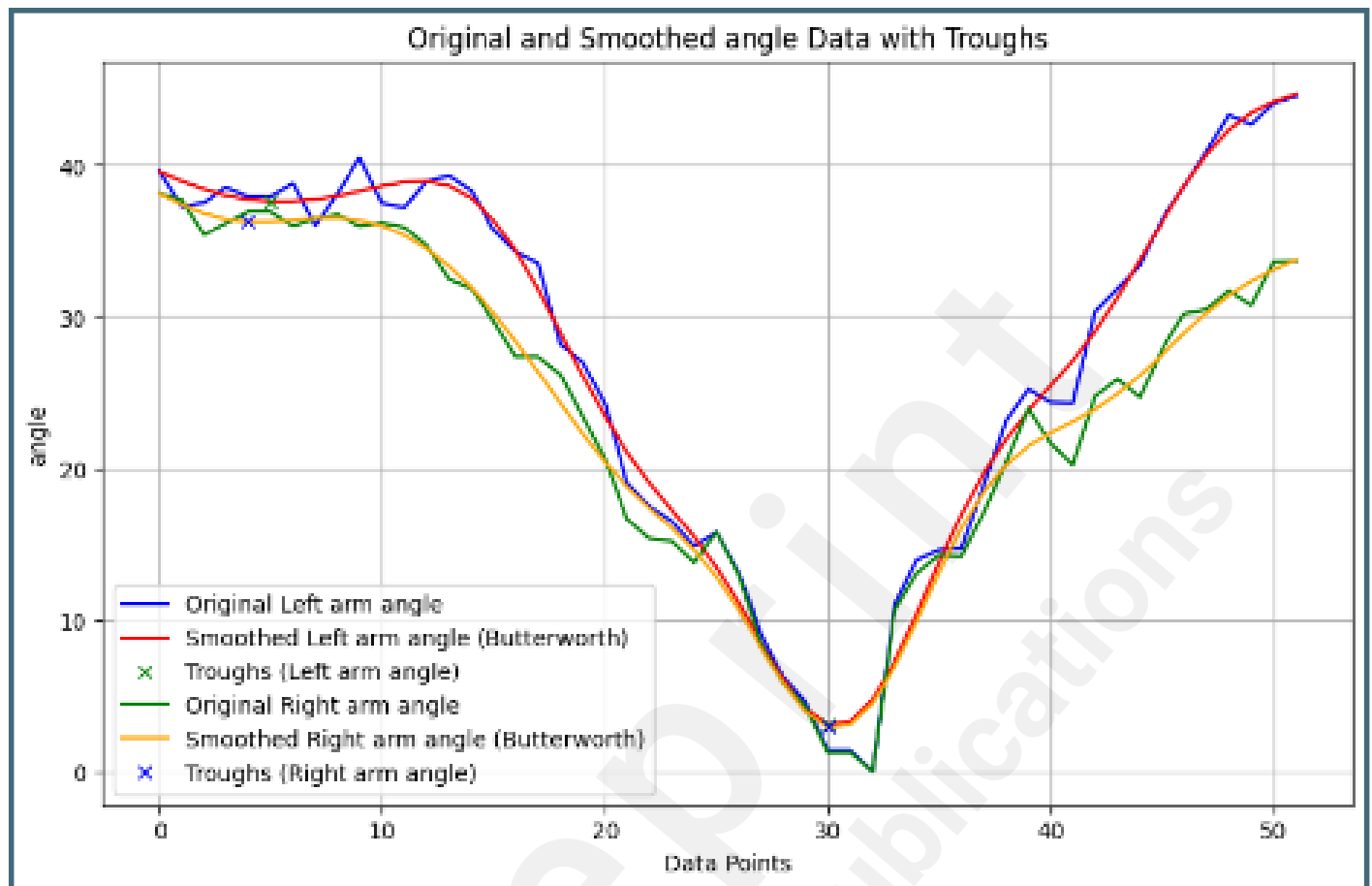


Fig. 9 Comparative Stride Length Analysis between Groups. (c) PD.

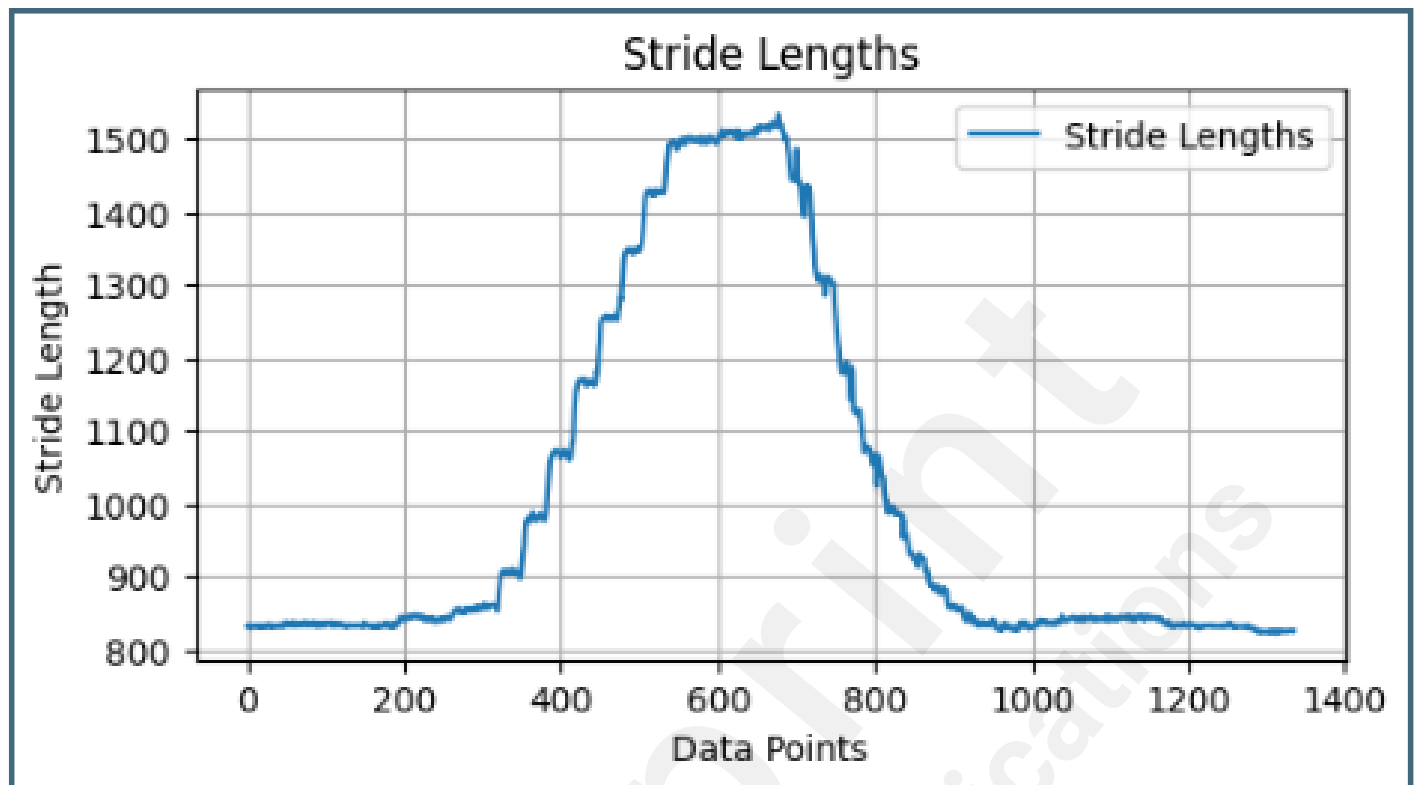


Fig. 9 Comparative Stride Length Analysis between Groups. (b) Elder.

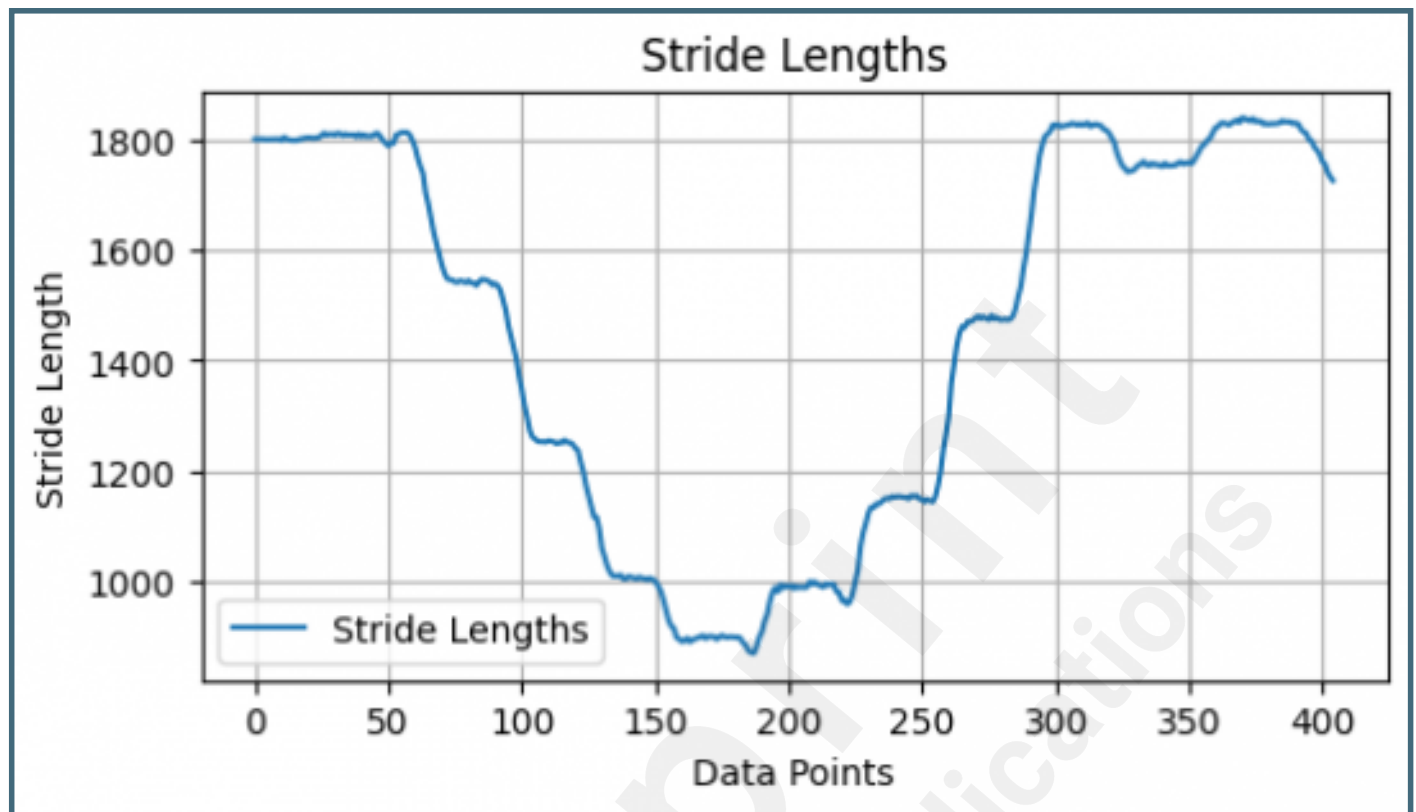


Fig. 17 Illustration of Trunk Angle I Vertical Position. (c) PD.

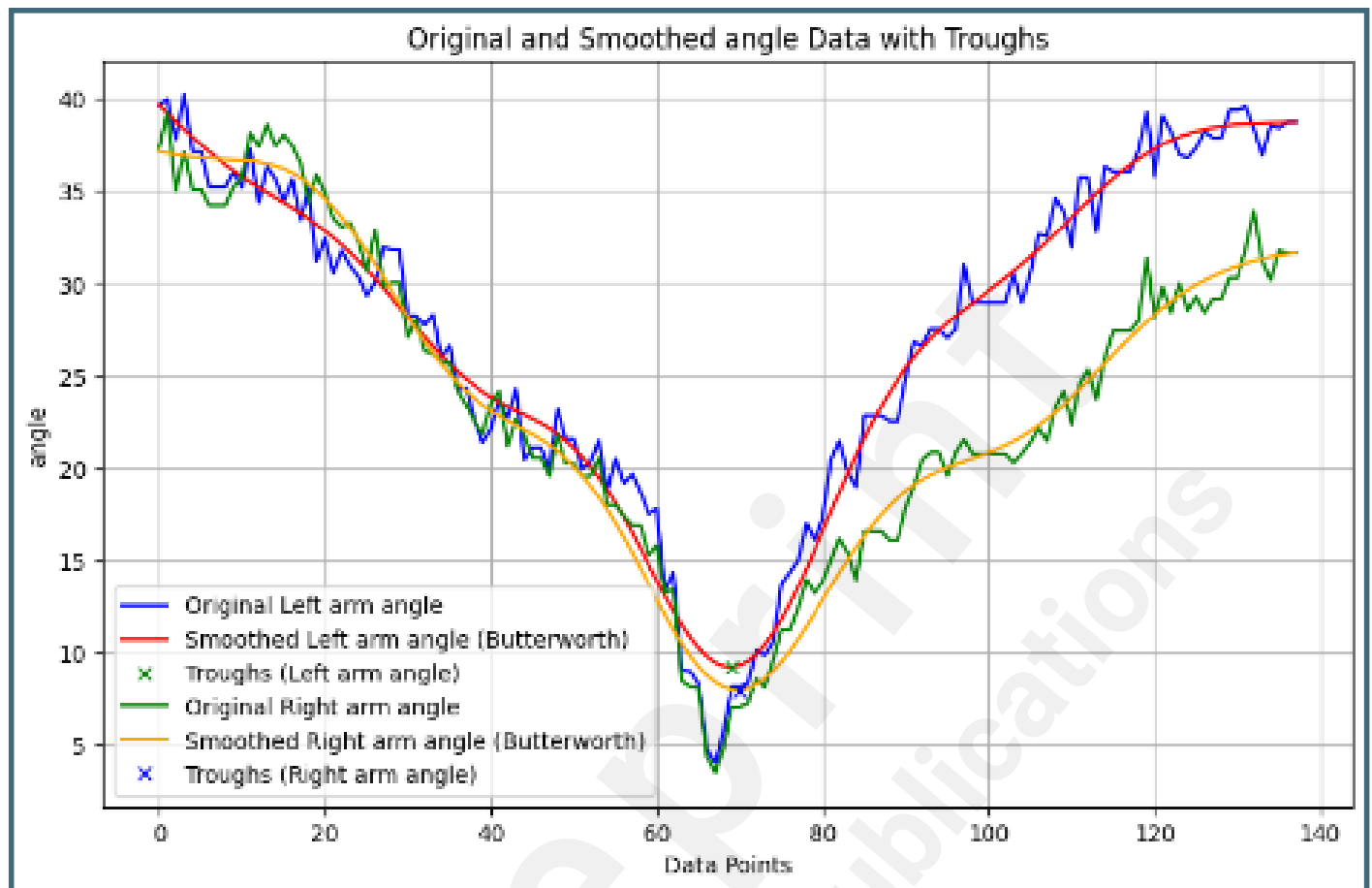


Fig. 8 Comparative Step Length Analysis between Groups. (c) PD.

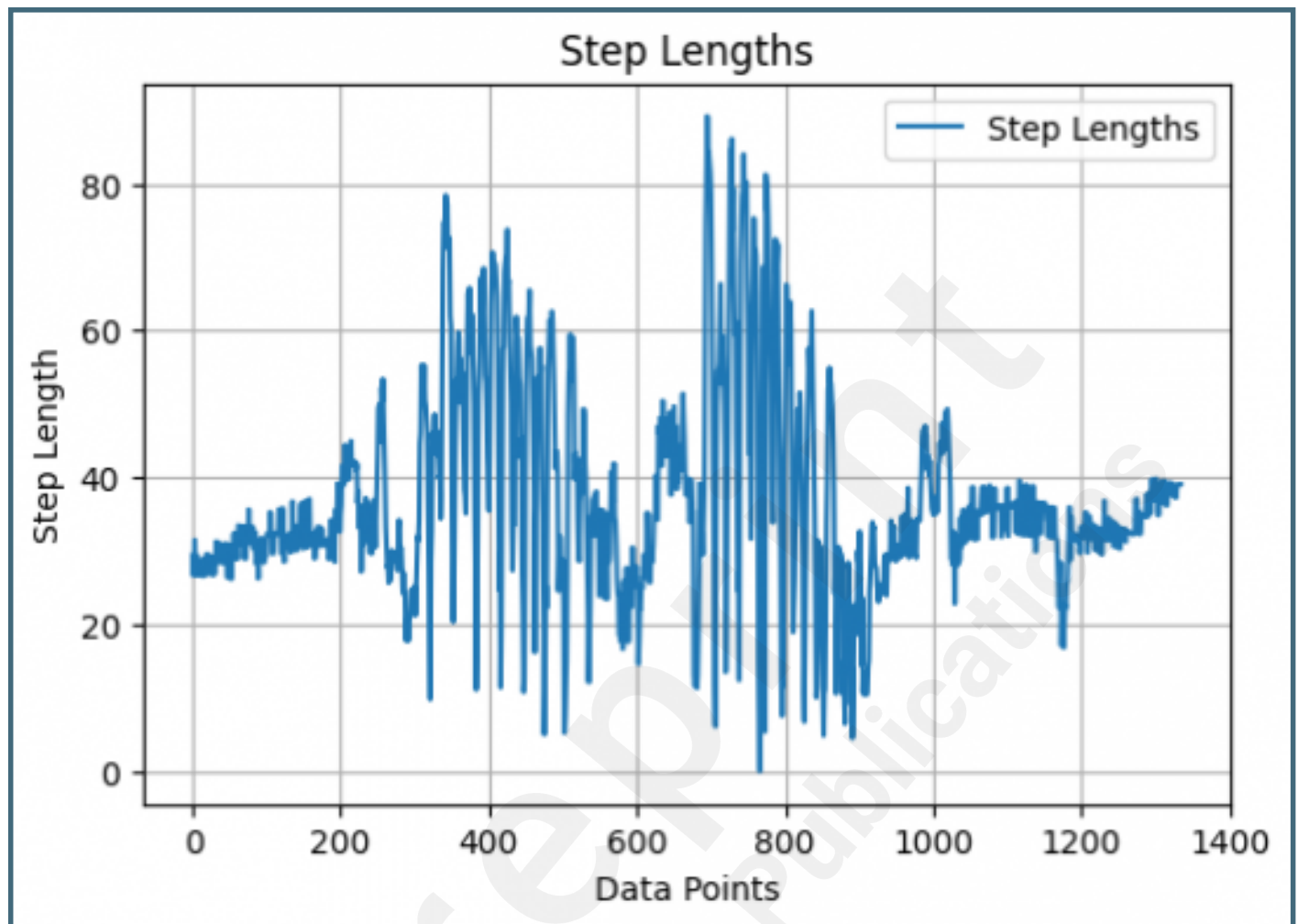


Fig. 8 Comparative Step Length Analysis between Groups. (b) Elder.

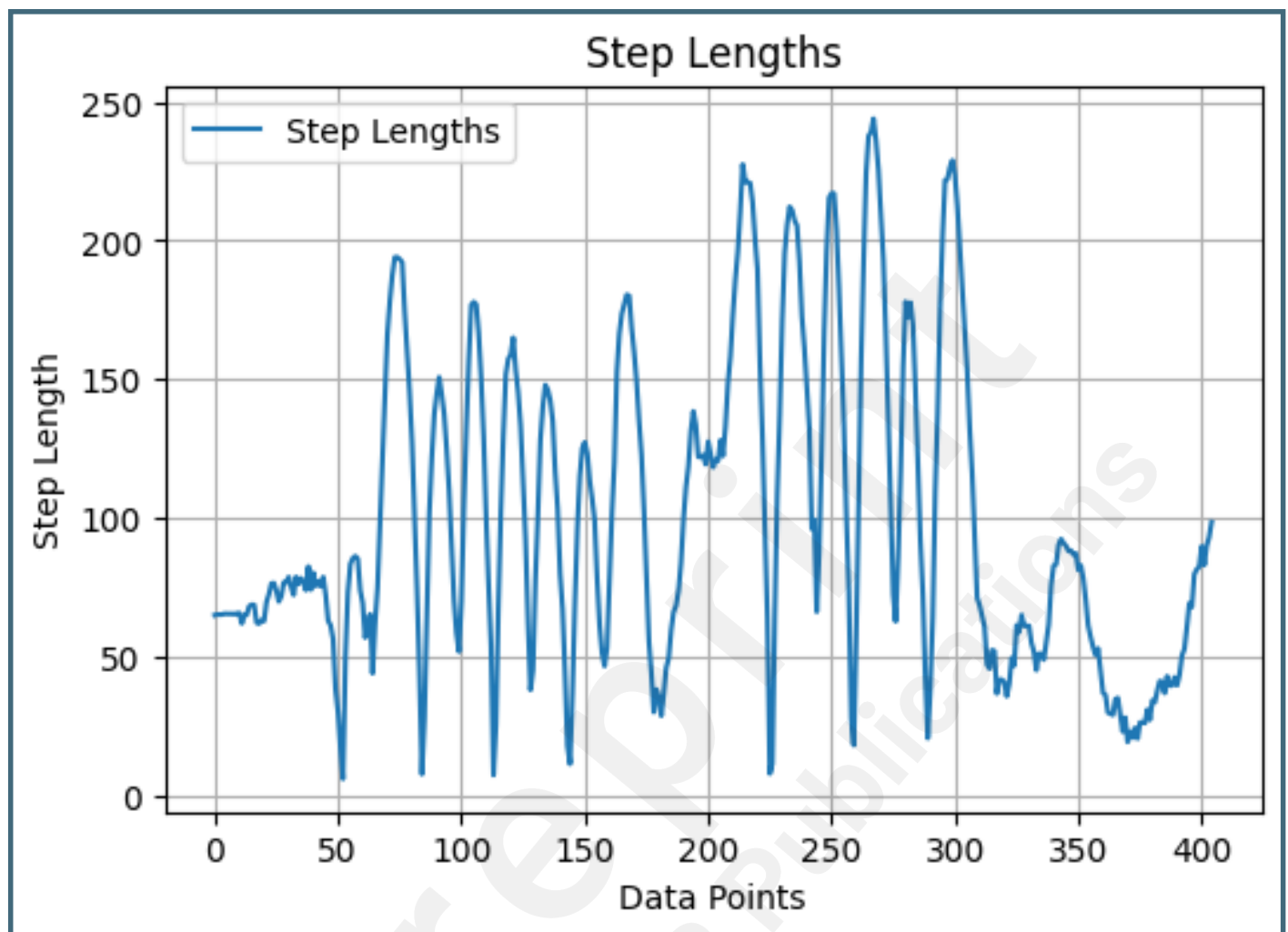


Fig. 7 Comparative Step Analysis by Step Length between Groups. (c) PD.

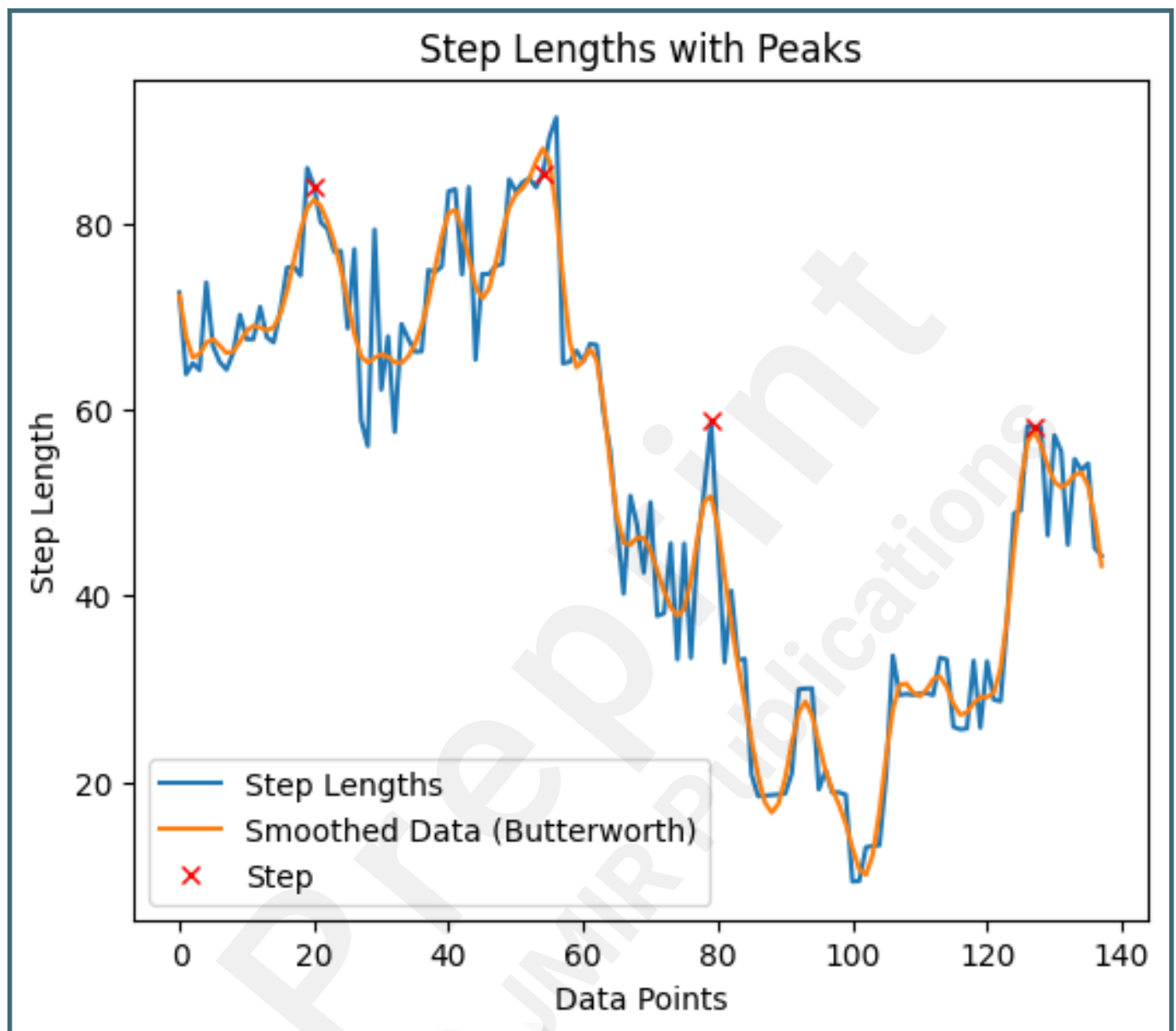


Fig. 13 Comparative Hip Angle between Groups. (b) Elder.

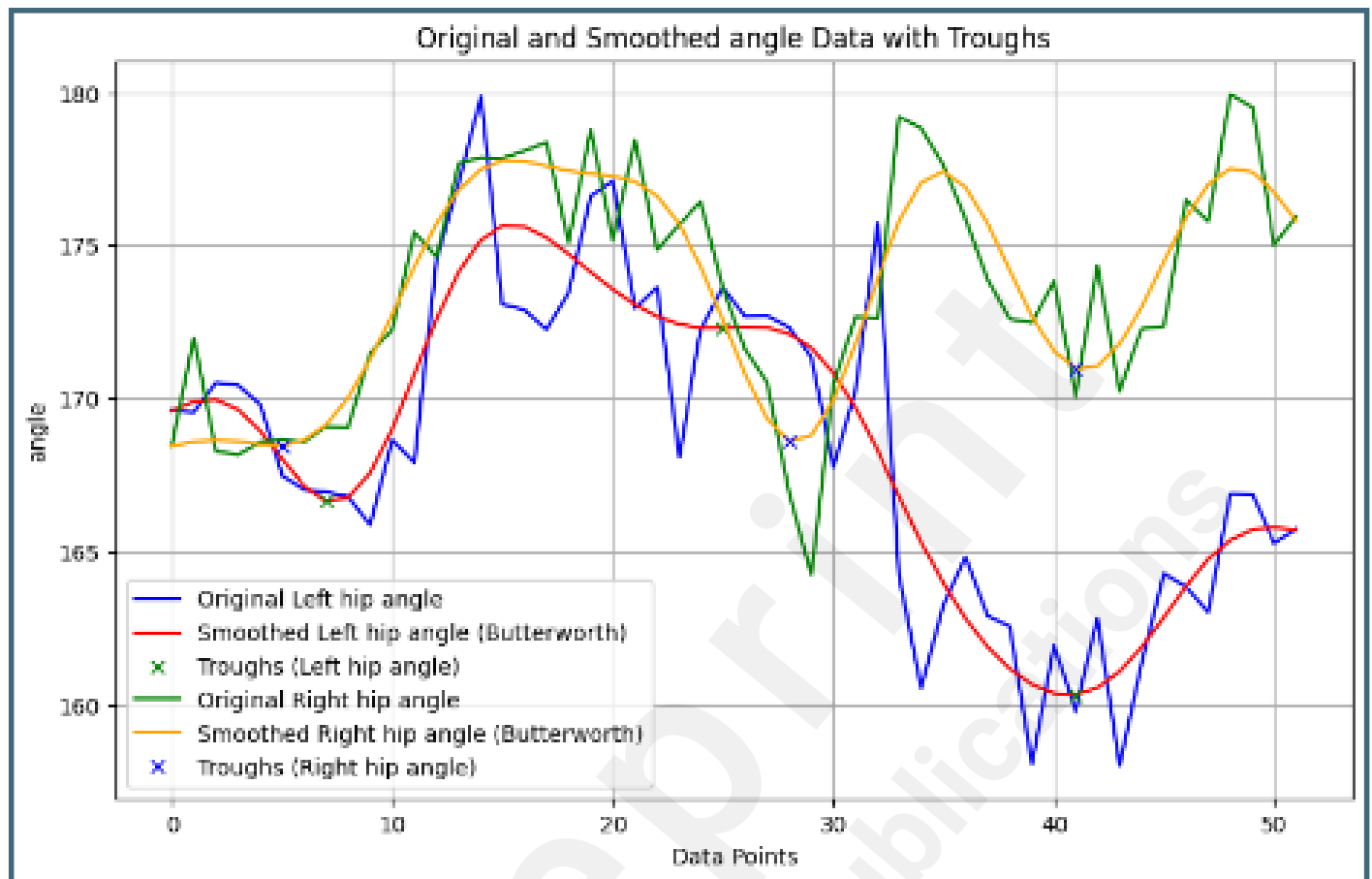


Fig. 7 Comparative Step Analysis by Step Length between Groups. (b) Elder.

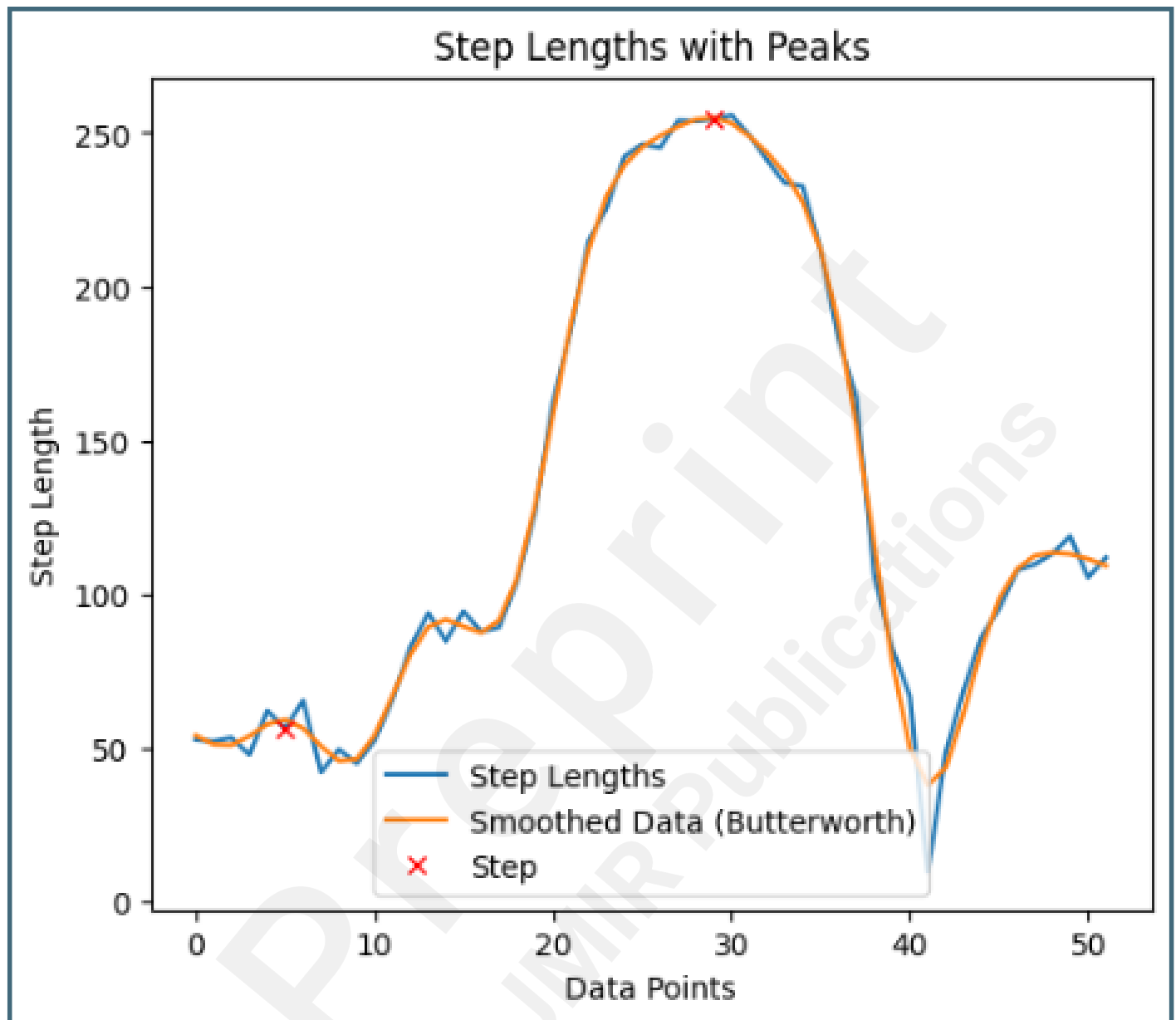


Fig. 19 Comparative Trunk Angle I between Groups. (c) PD.

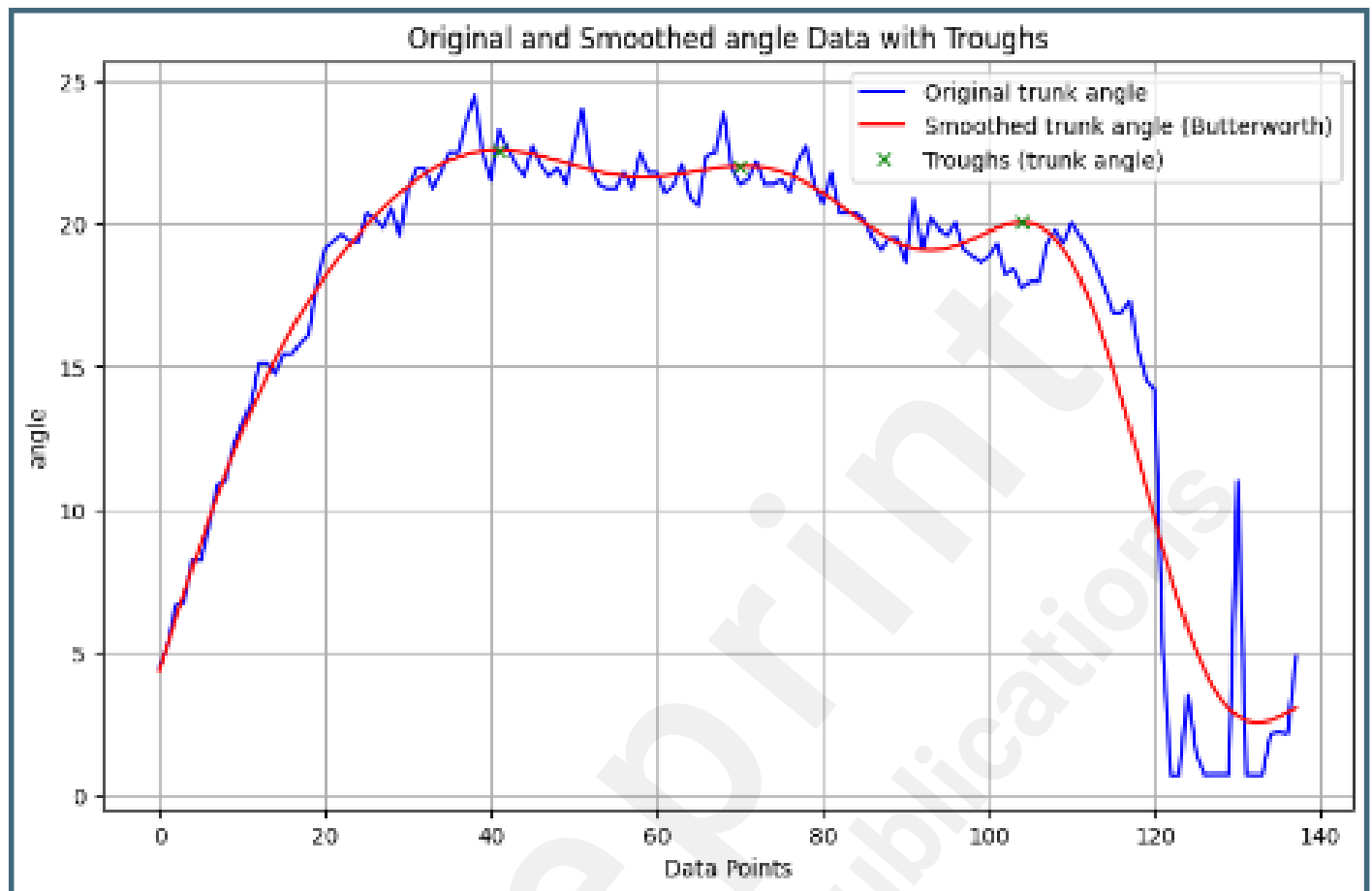


Fig. 6 Shoulder Distance Analysis with Peak Detection. (b) Graph after smoothed.

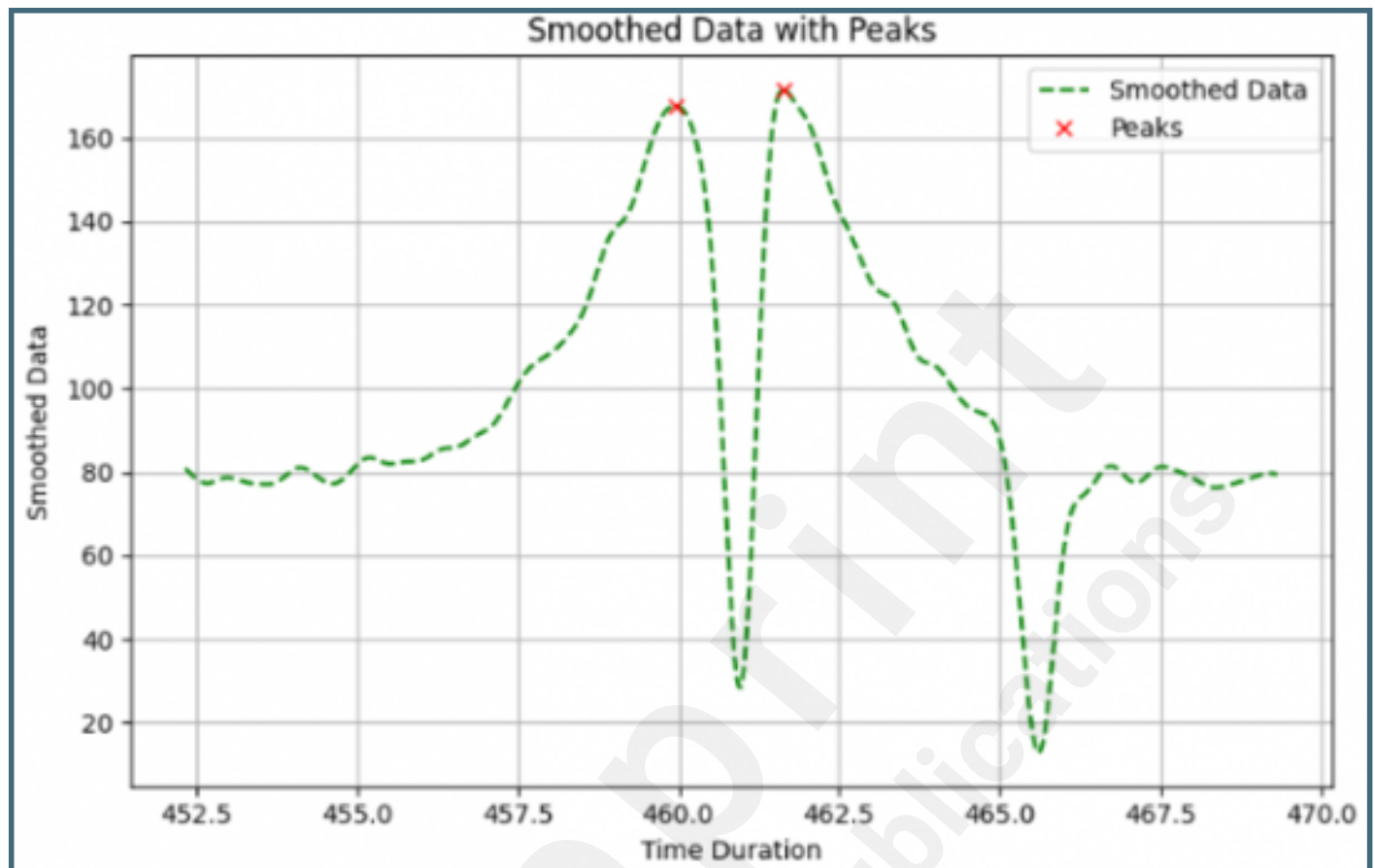


Fig. 21 Comparative Trunk Angle II Horizontal between Groups. (b) Elder.

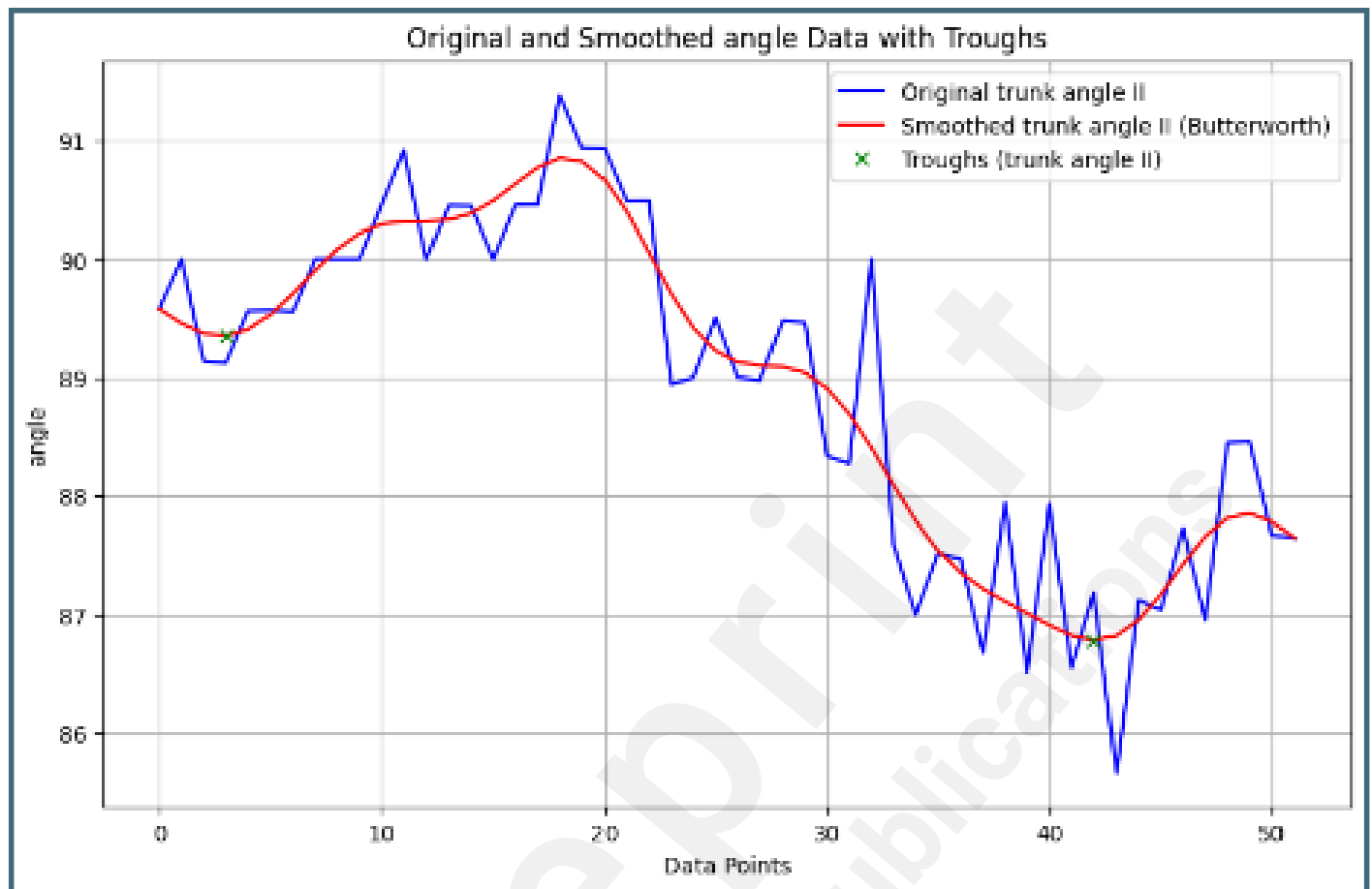


Fig. 21 Comparative Trunk Angle II Horizontal between Groups. (c) PD.

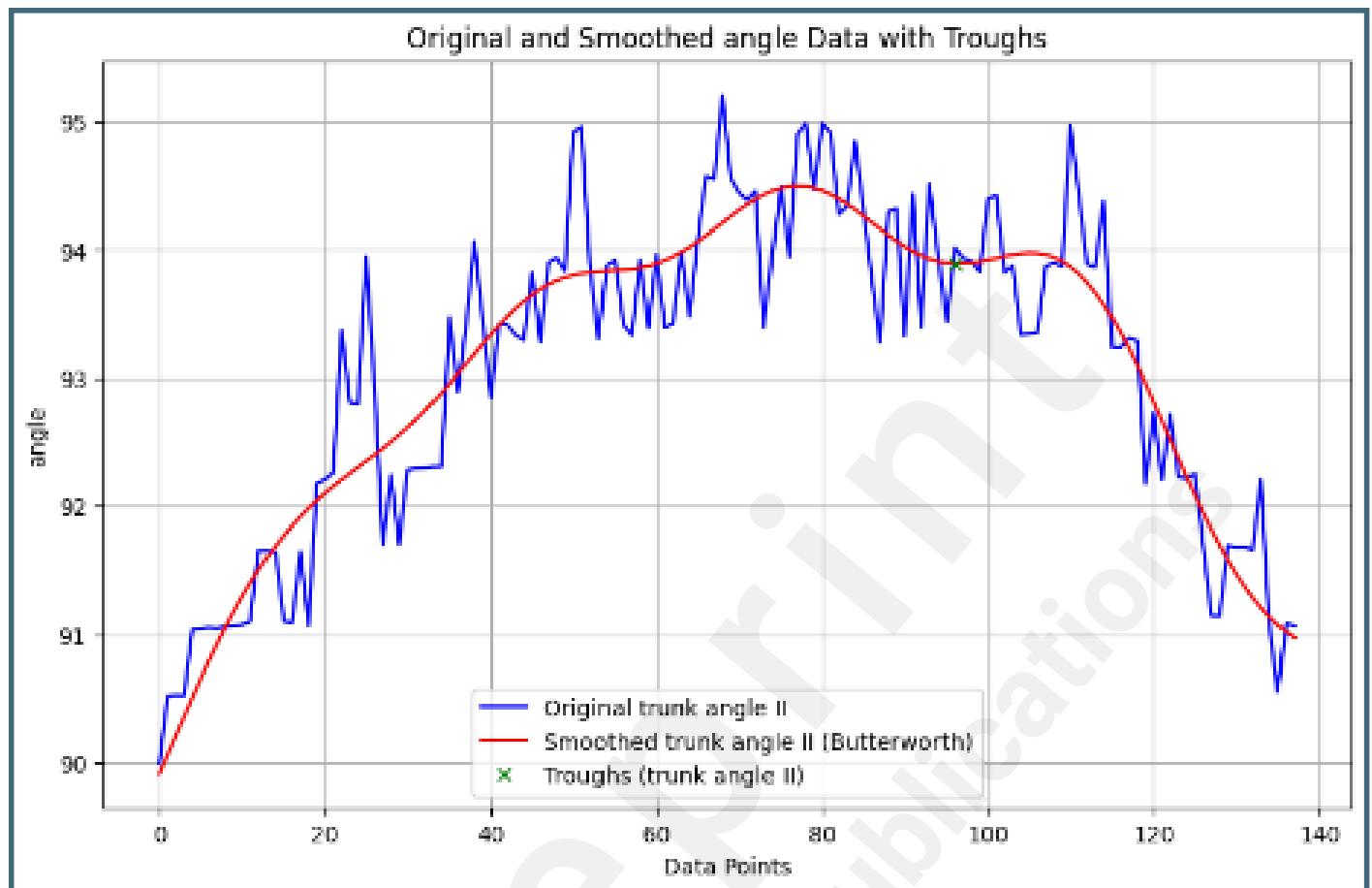


Fig. 23 Comparative Shank Angle between Groups. (b) Elder.

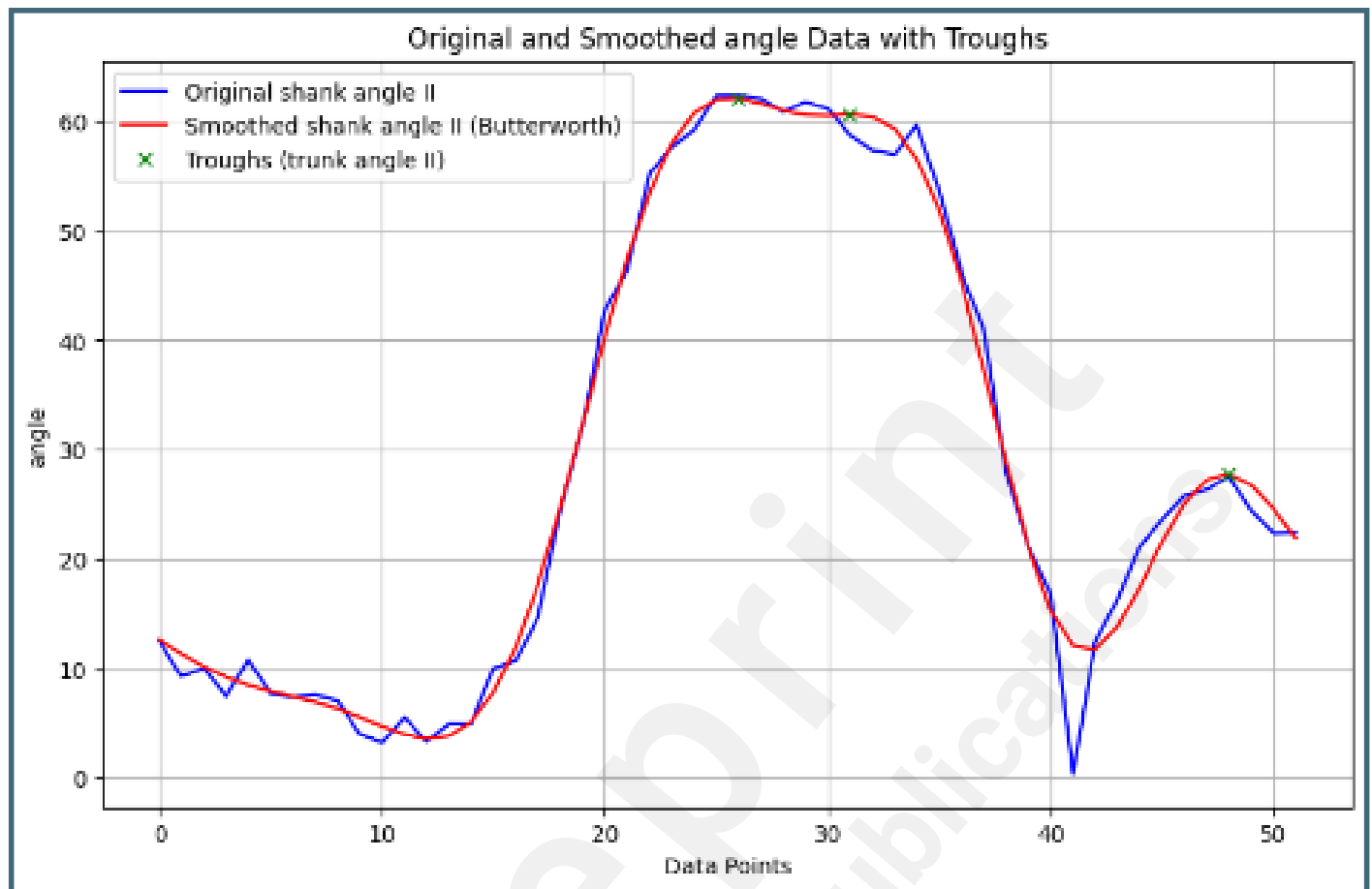


Fig. 23 Comparative Shank Angle between Groups. (c) PD.

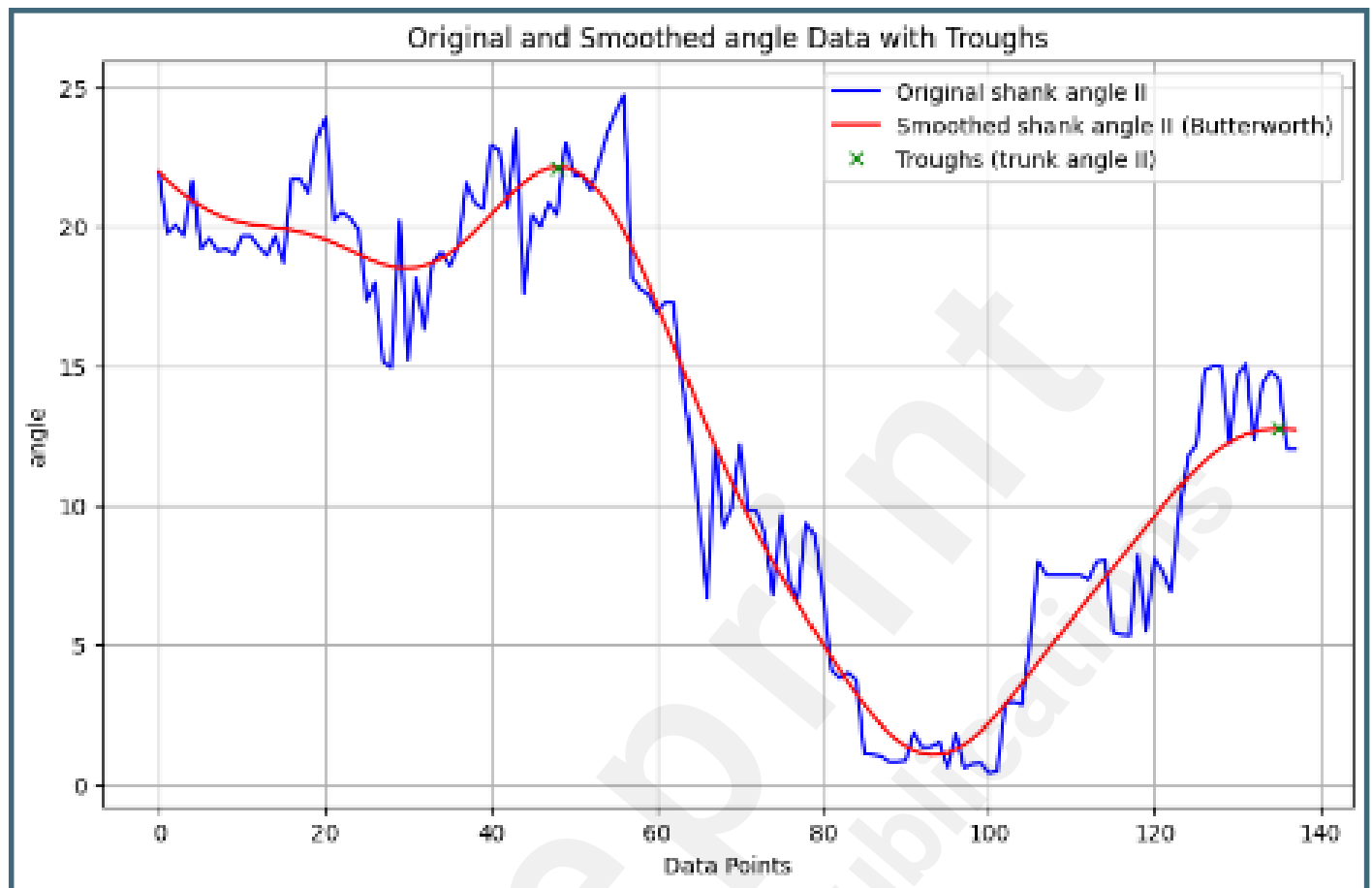


Fig. 2 Demographic Information of the Subjects. (d) Age Range Distribution.

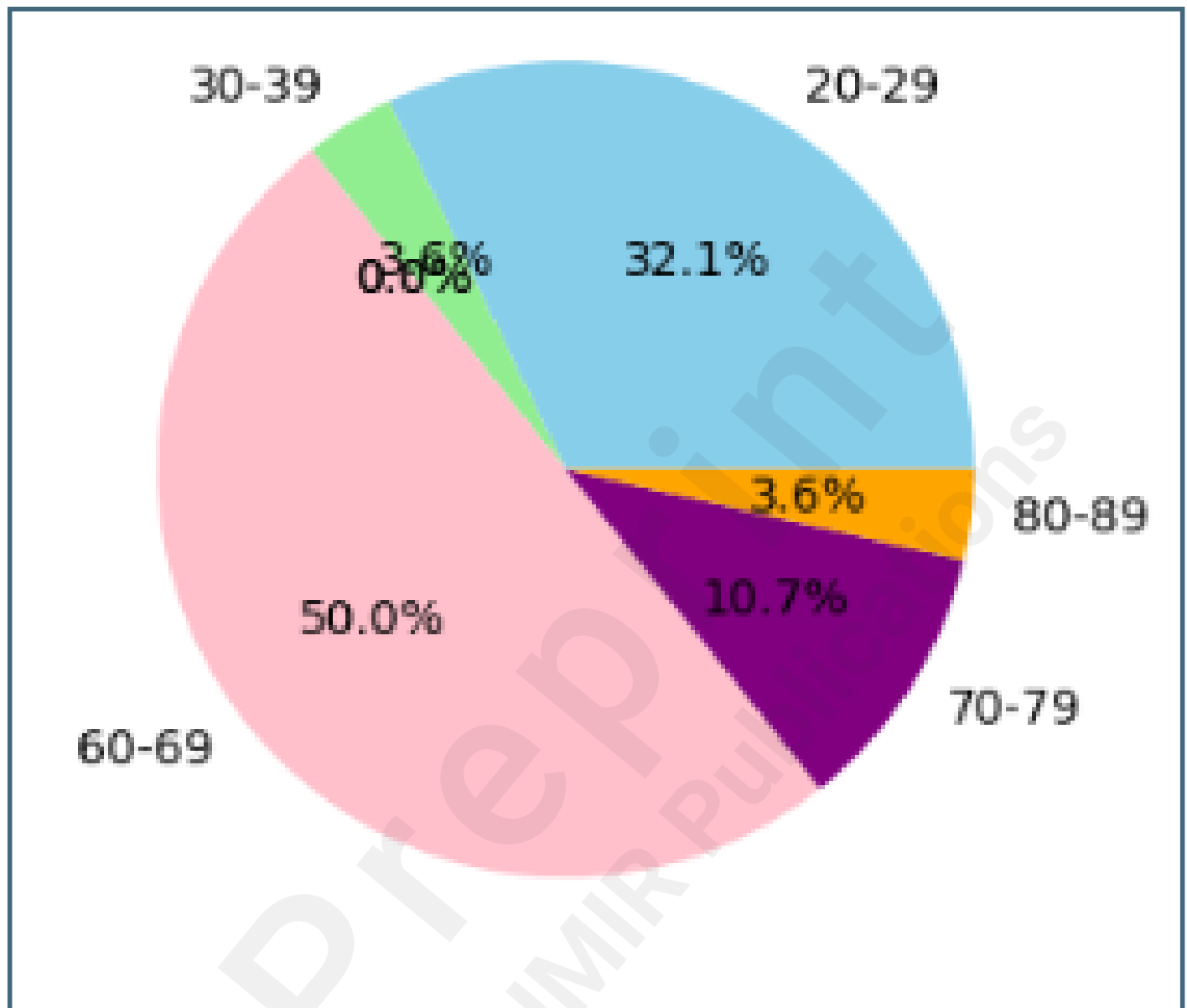


Fig. 2 Demographic Information of the Subjects. (c) Gender Distribution.

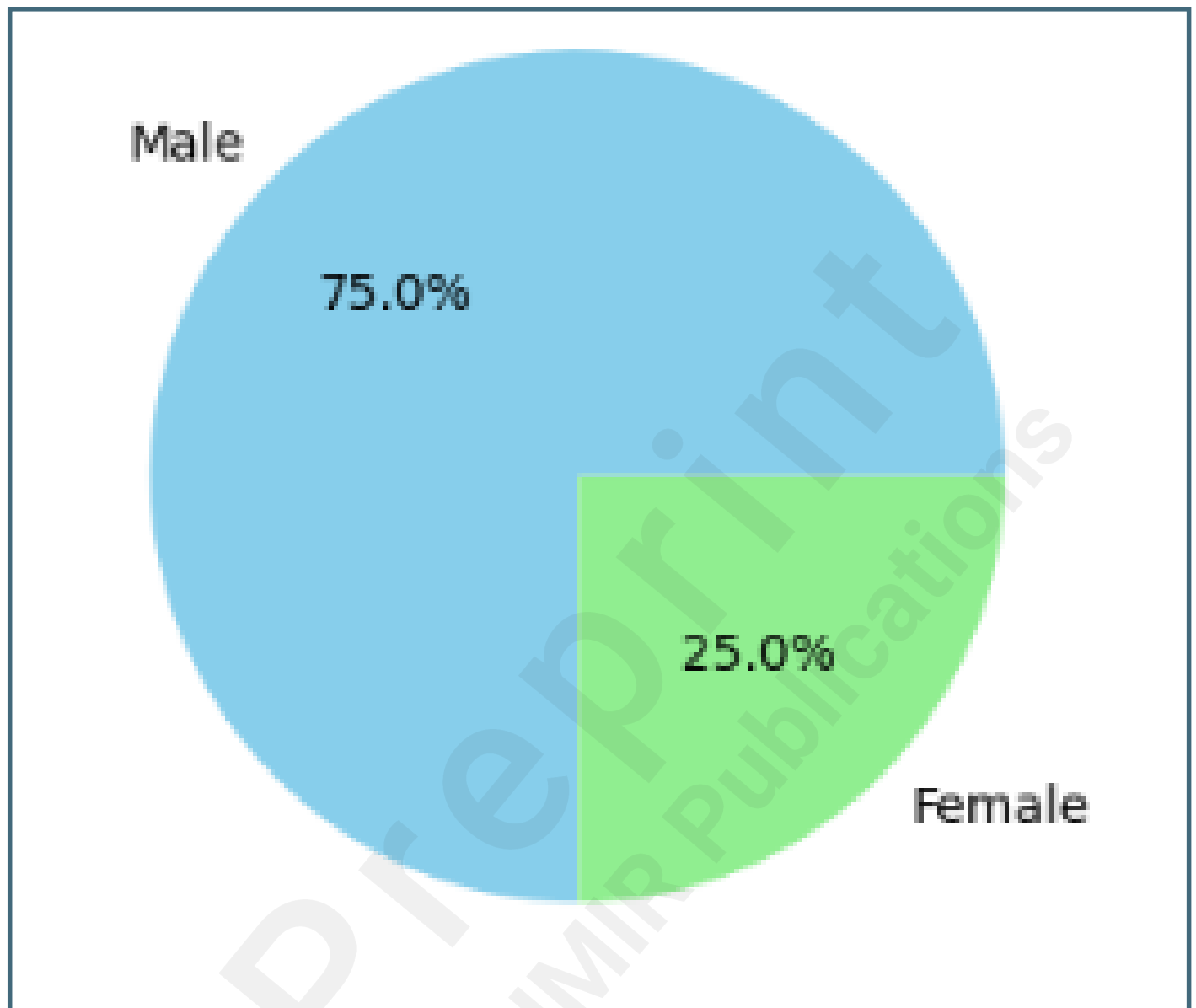


Fig. 2 Demographic Information of the Subjects. (b) Race Distribution.

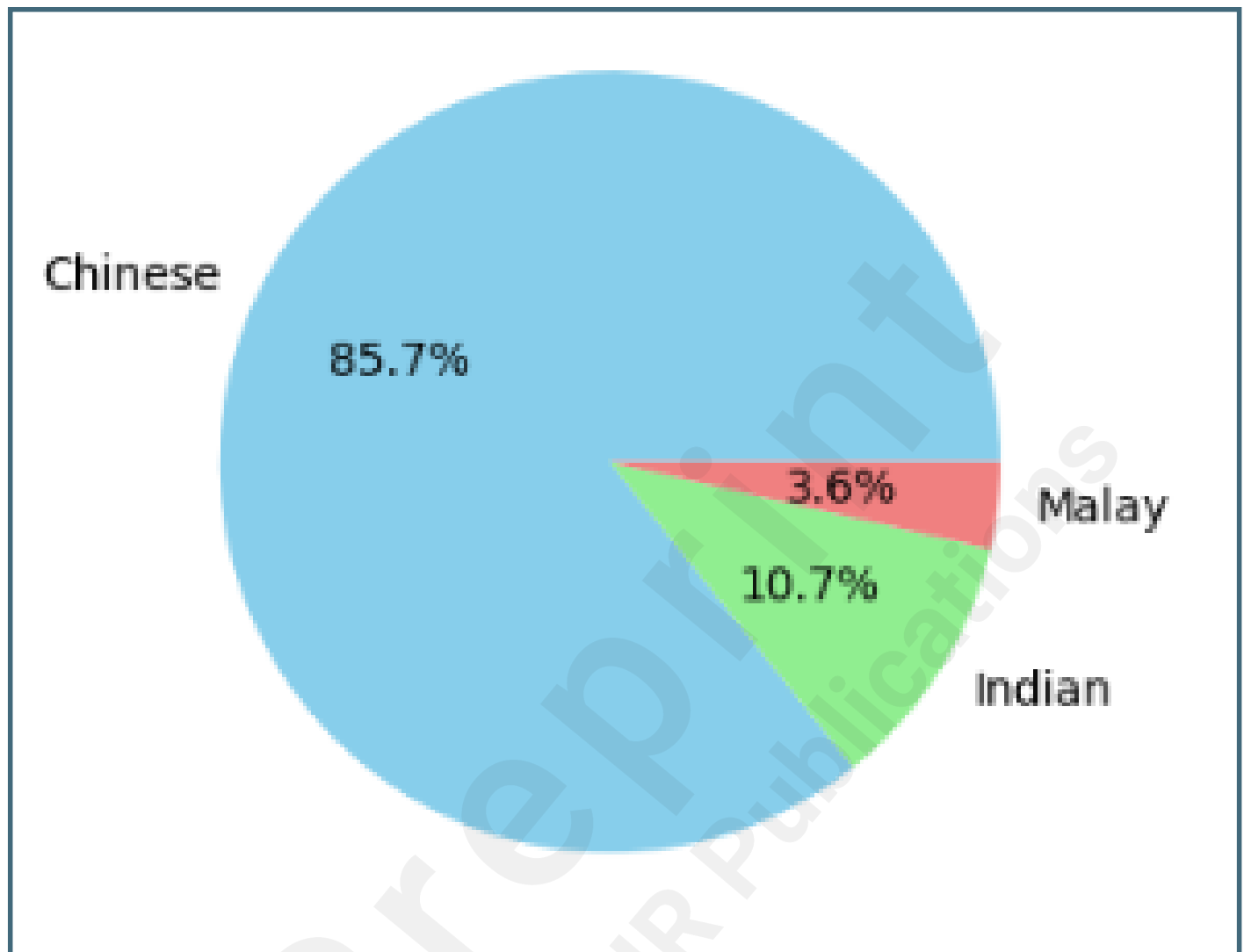


Fig. 15 Comparative Symmetric of Leg between Groups. (b) Elder.

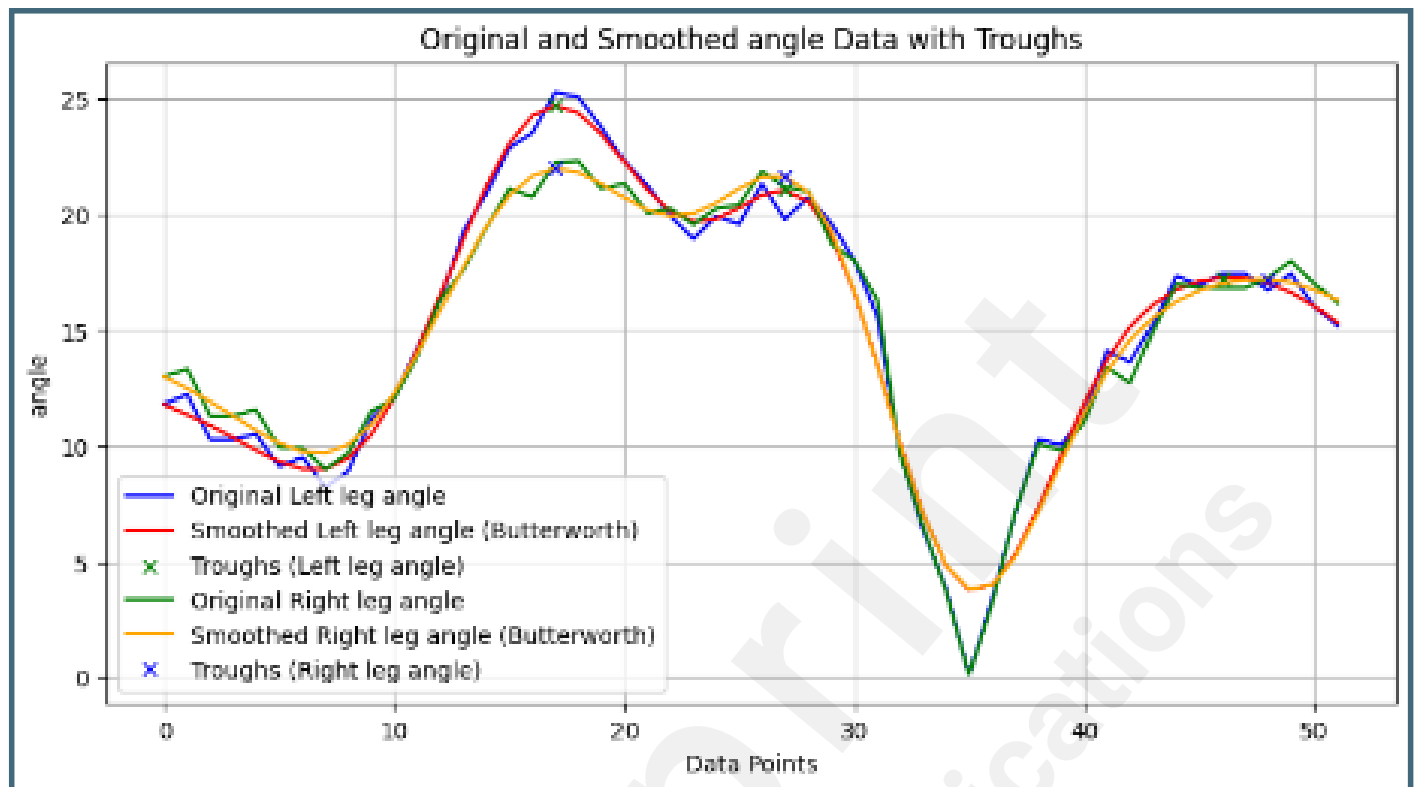


Fig. 1 TUG Shooting Site Configuration Plan.

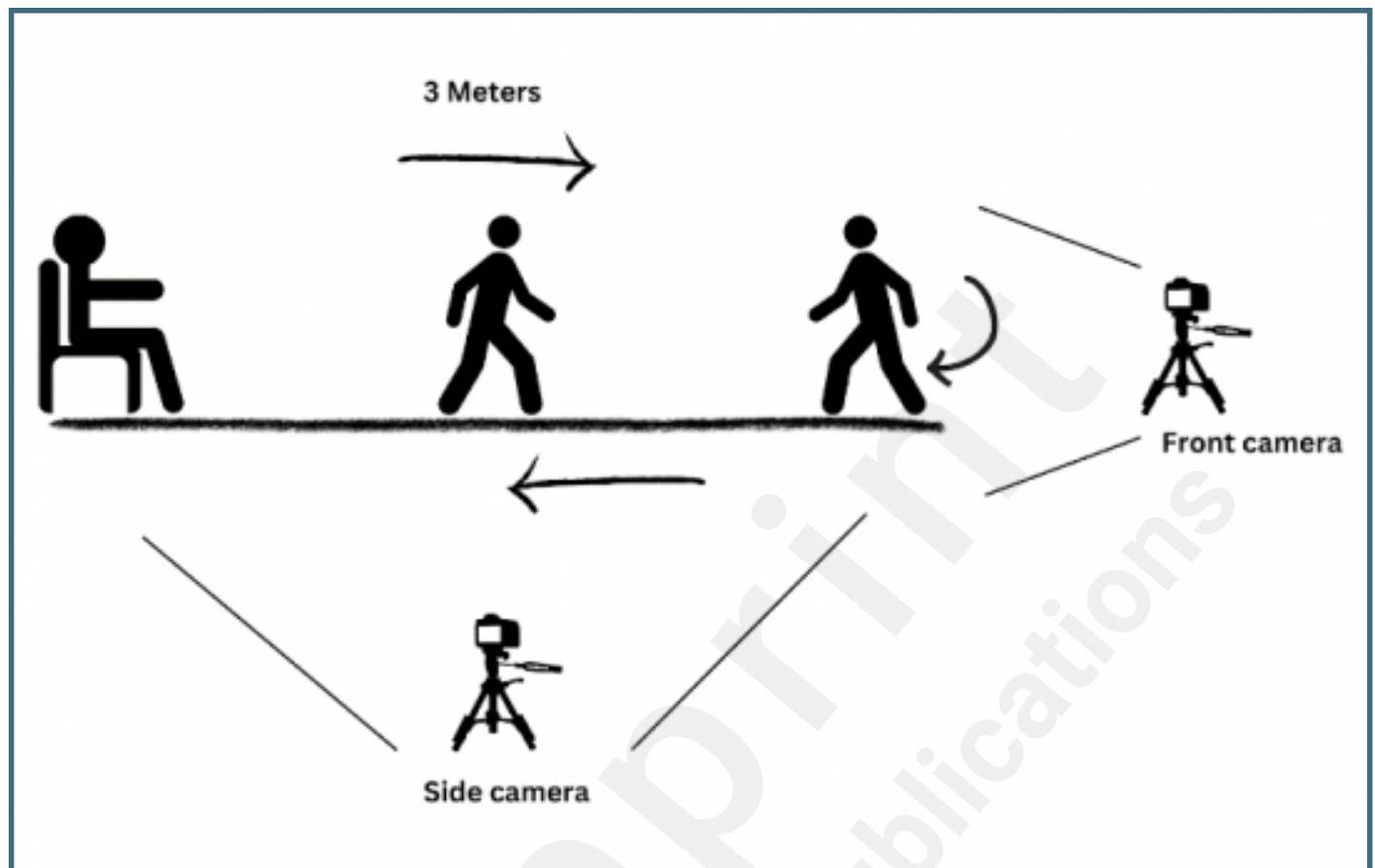


Fig. 2 Demographic Information of the Subjects. (a) Health Status Distribution.

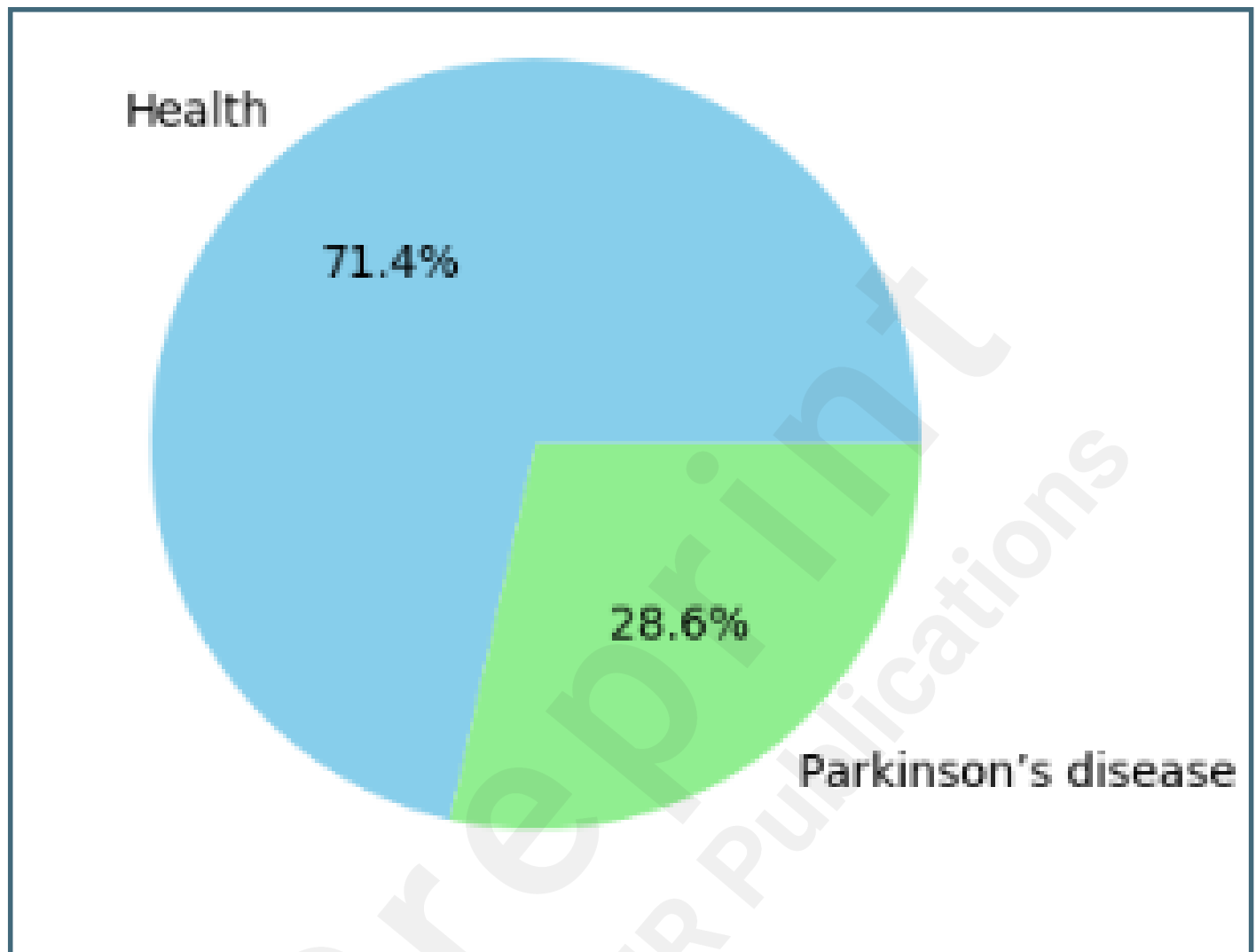


Fig. 3 Video Enhancement.

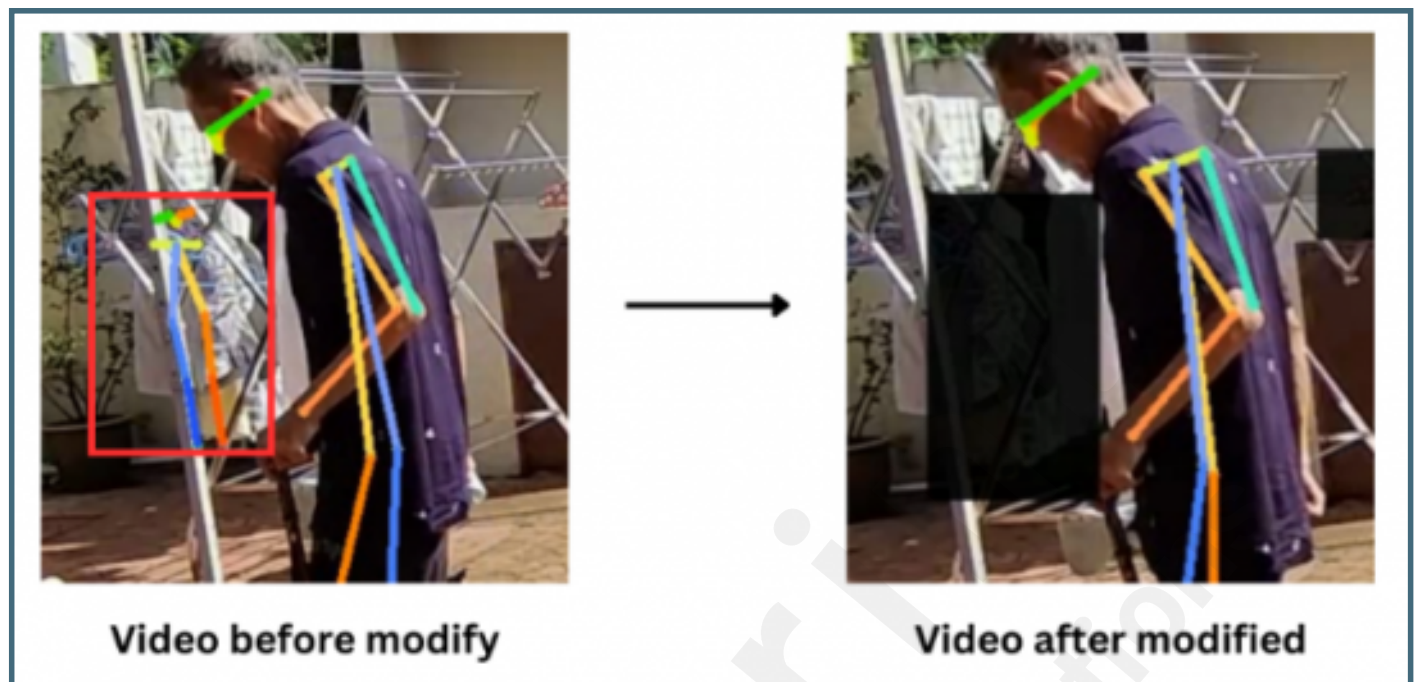


Fig. 4 Gait Analysis Using AlphaPose.

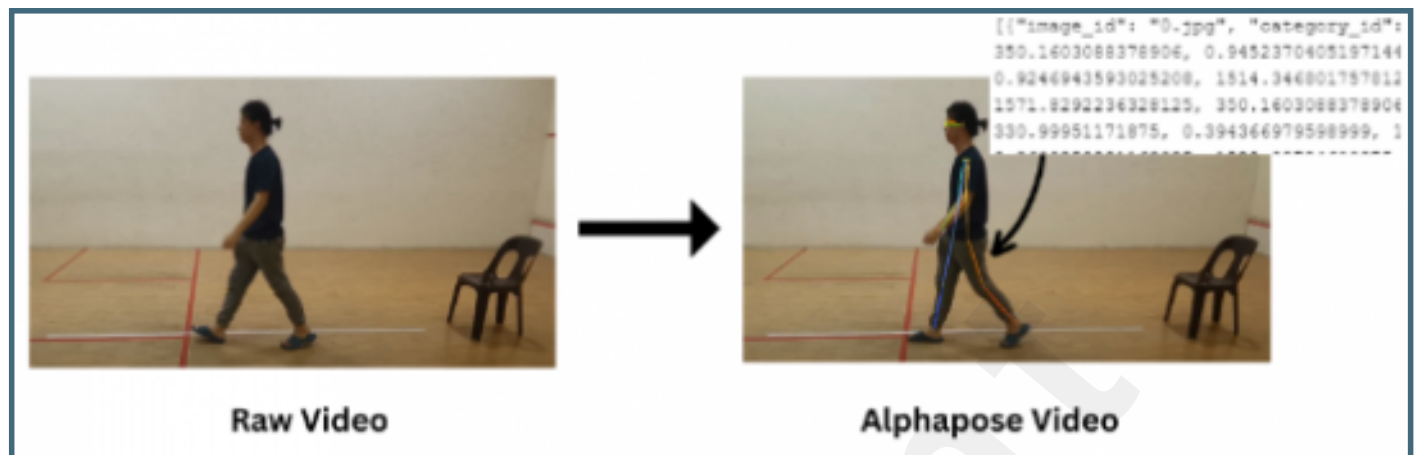


Fig. 5 Example of Graph by Butterworth with Peaks.

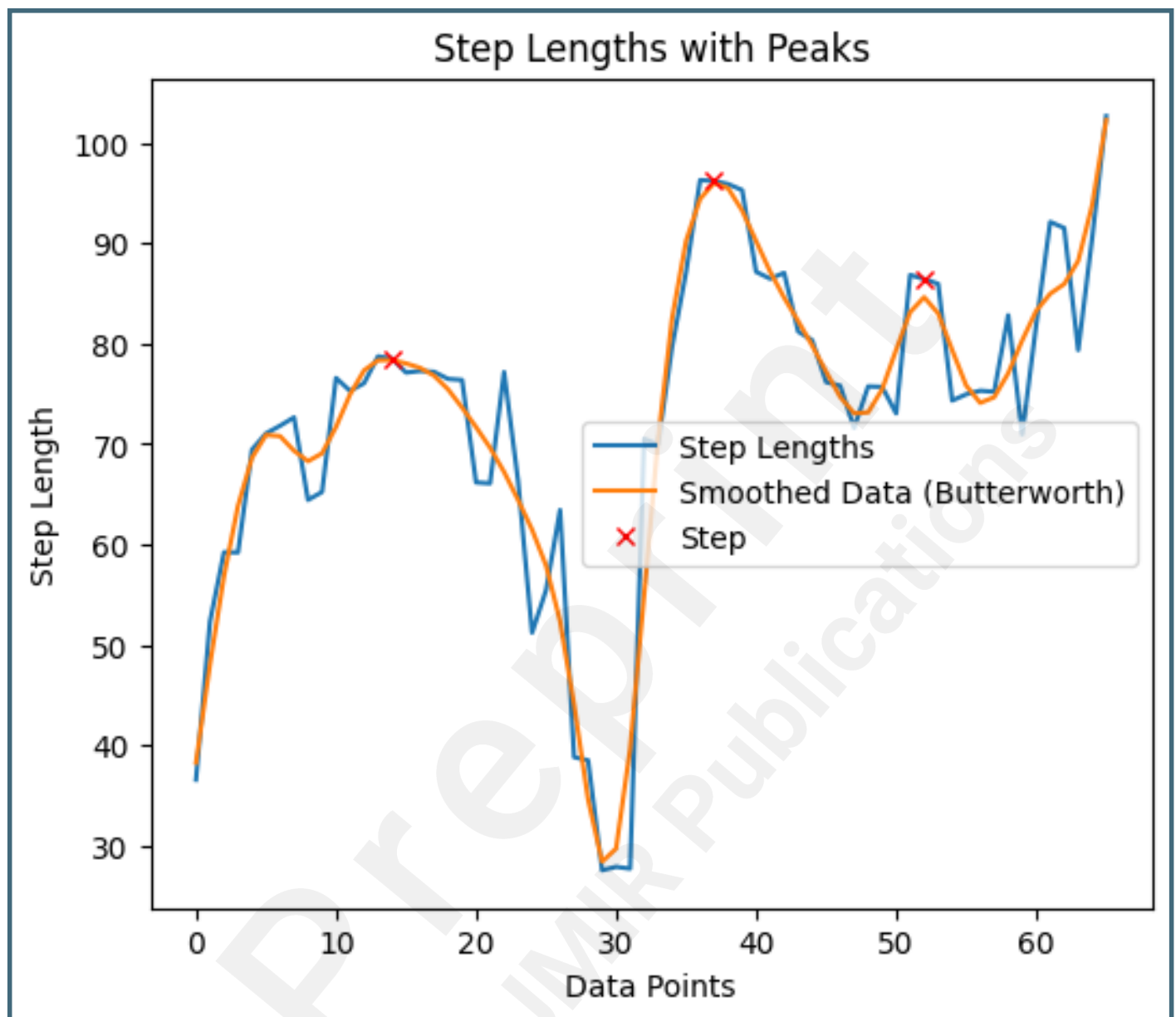


Fig. 6 Shoulder Distance Analysis with Peak Detection. (a) Graph before smoothed.

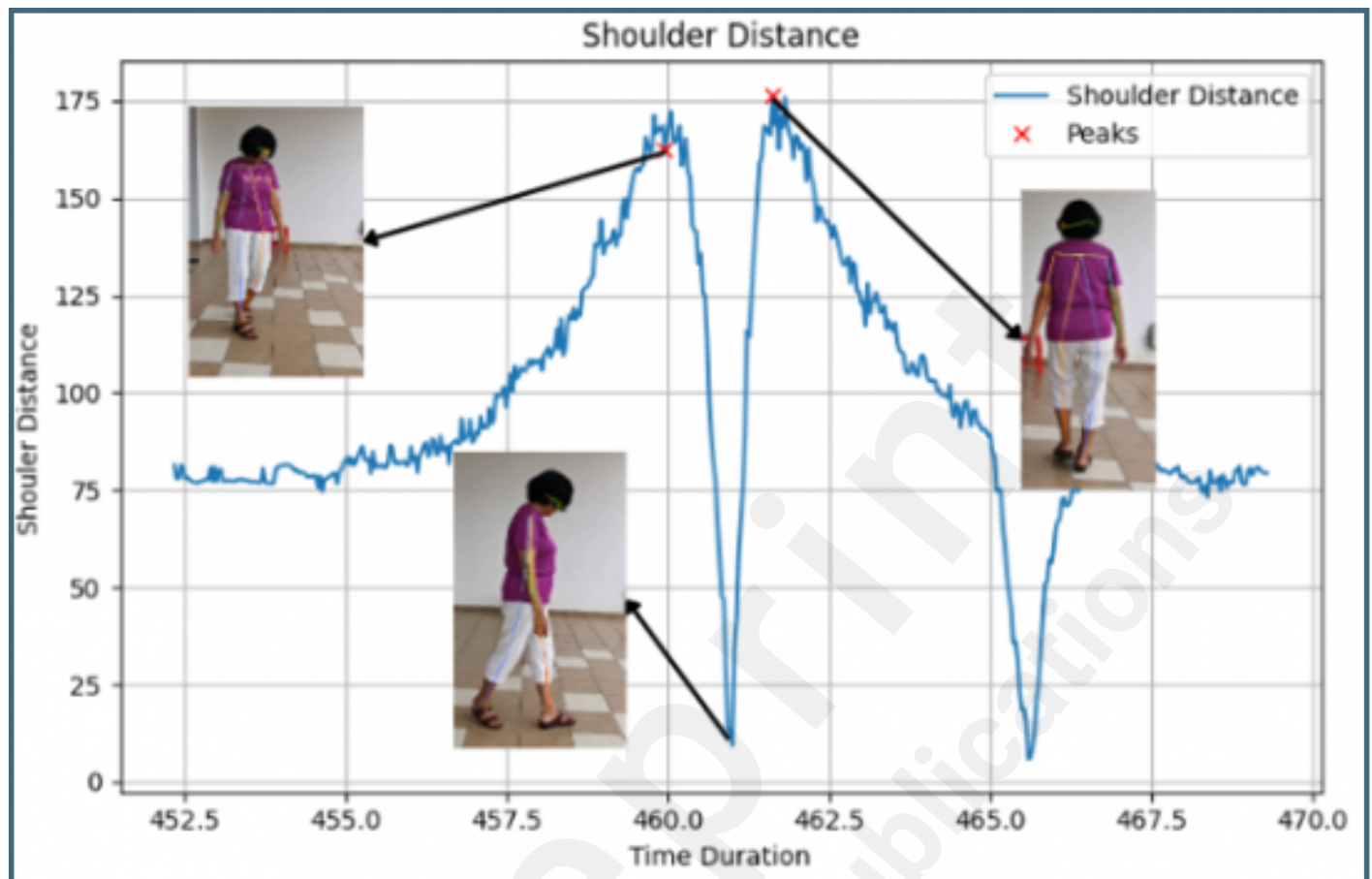


Fig. 7 Comparative Step Analysis by Step Length between Groups. (a) Young.

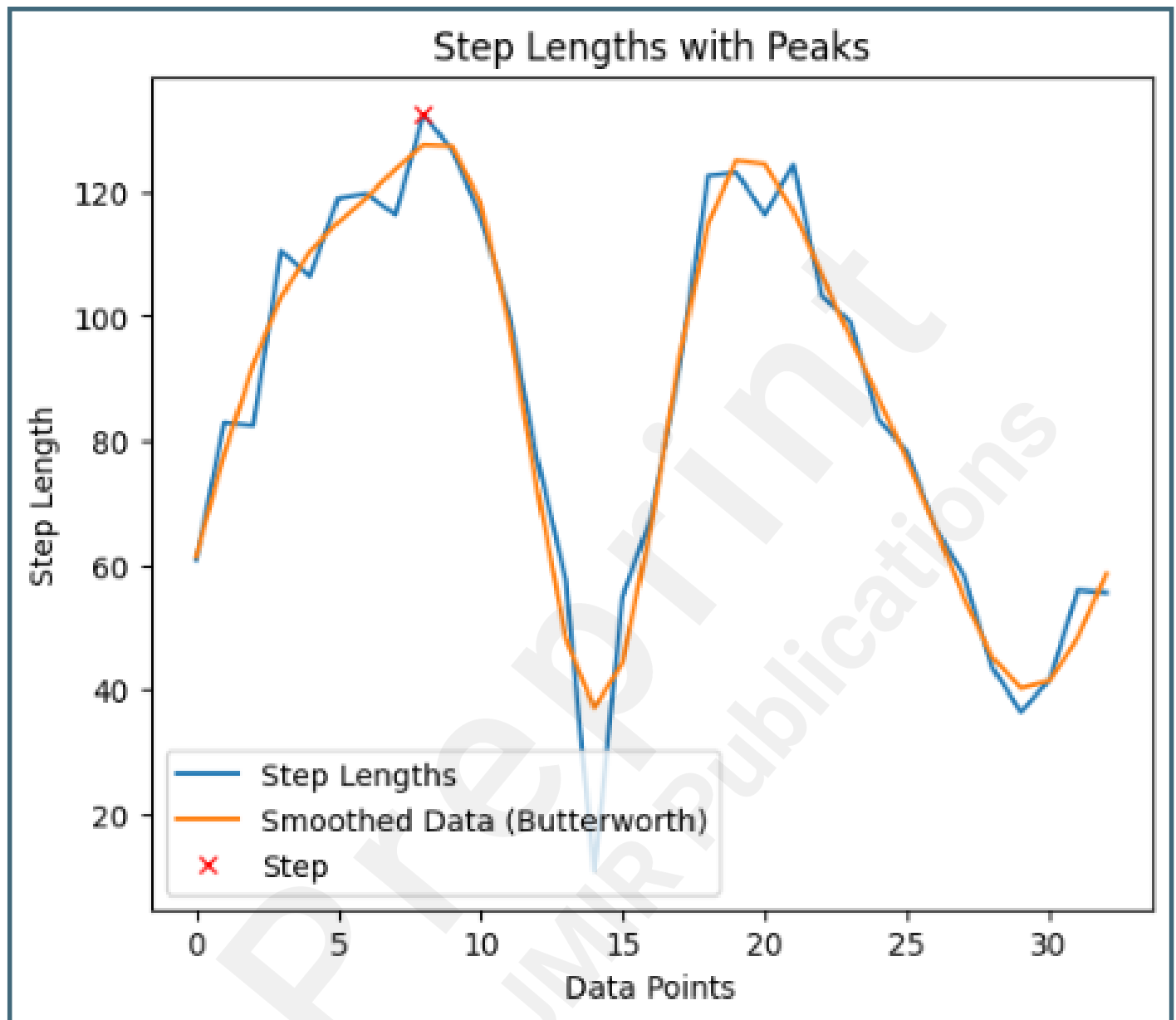


Fig. 8 Comparative Step Length Analysis between Groups. (a) Young.

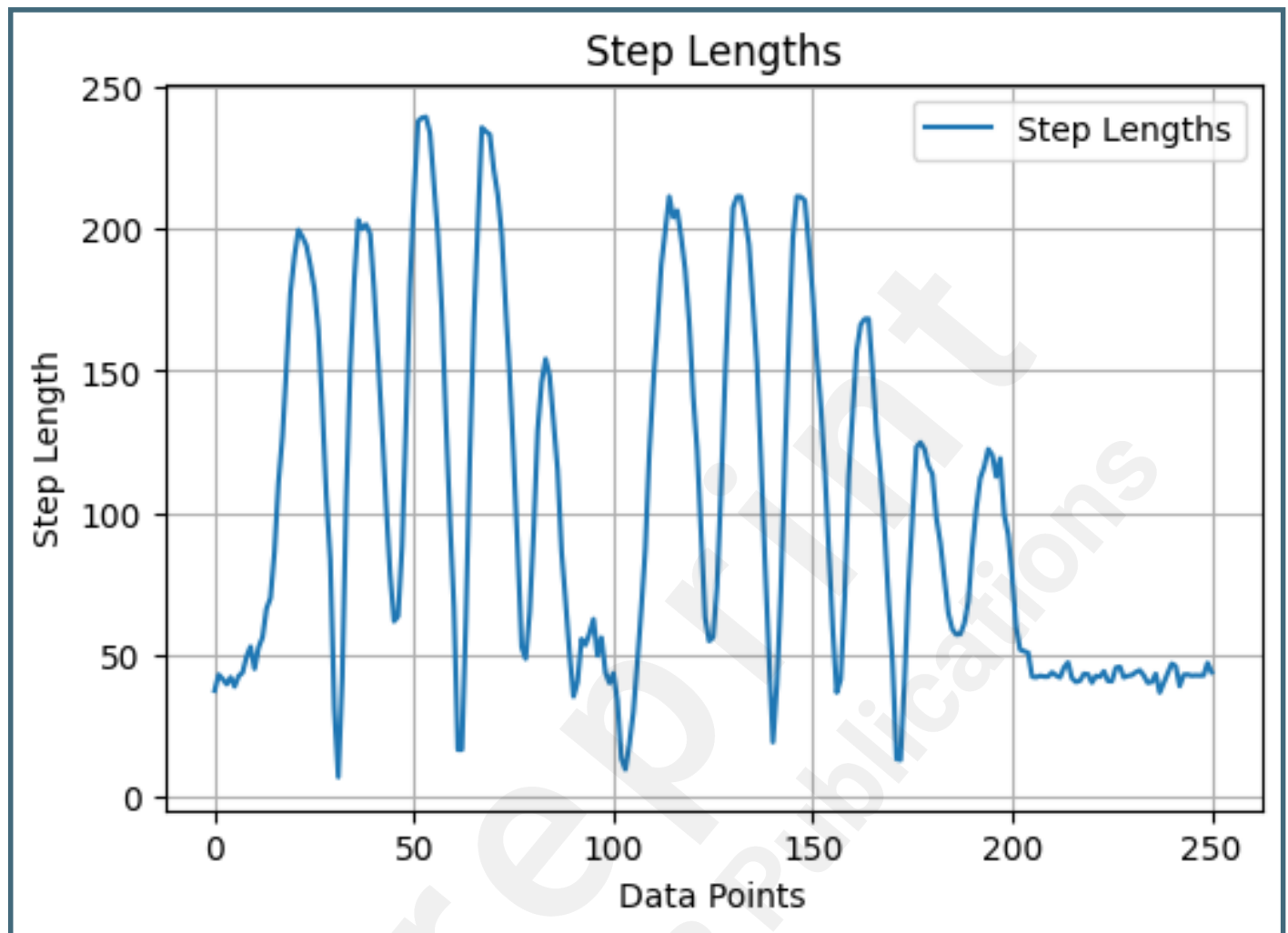


Fig. 9 Comparative Stride Length Analysis between Groups. (a) Young.

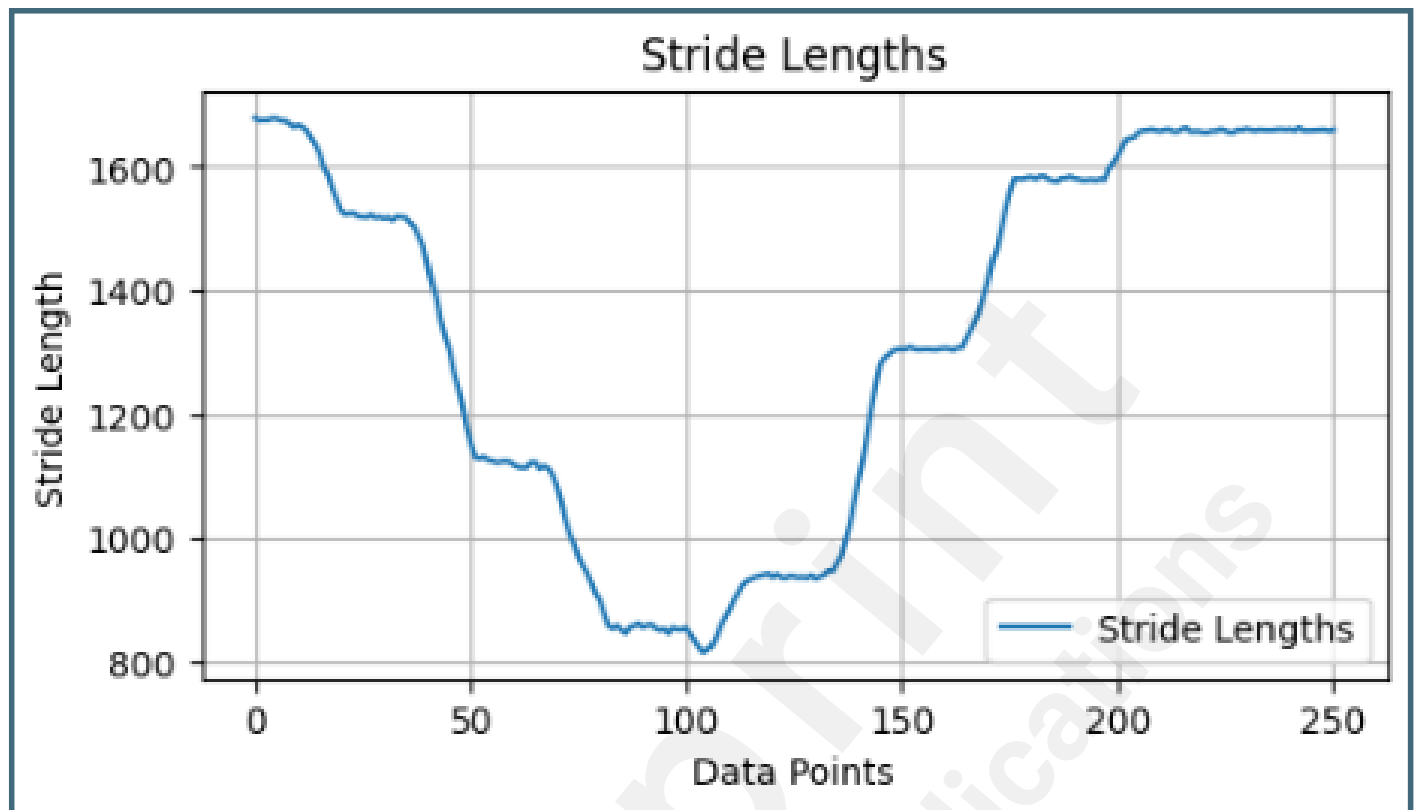


Fig. 10 Illustration of Knee Angle Position.

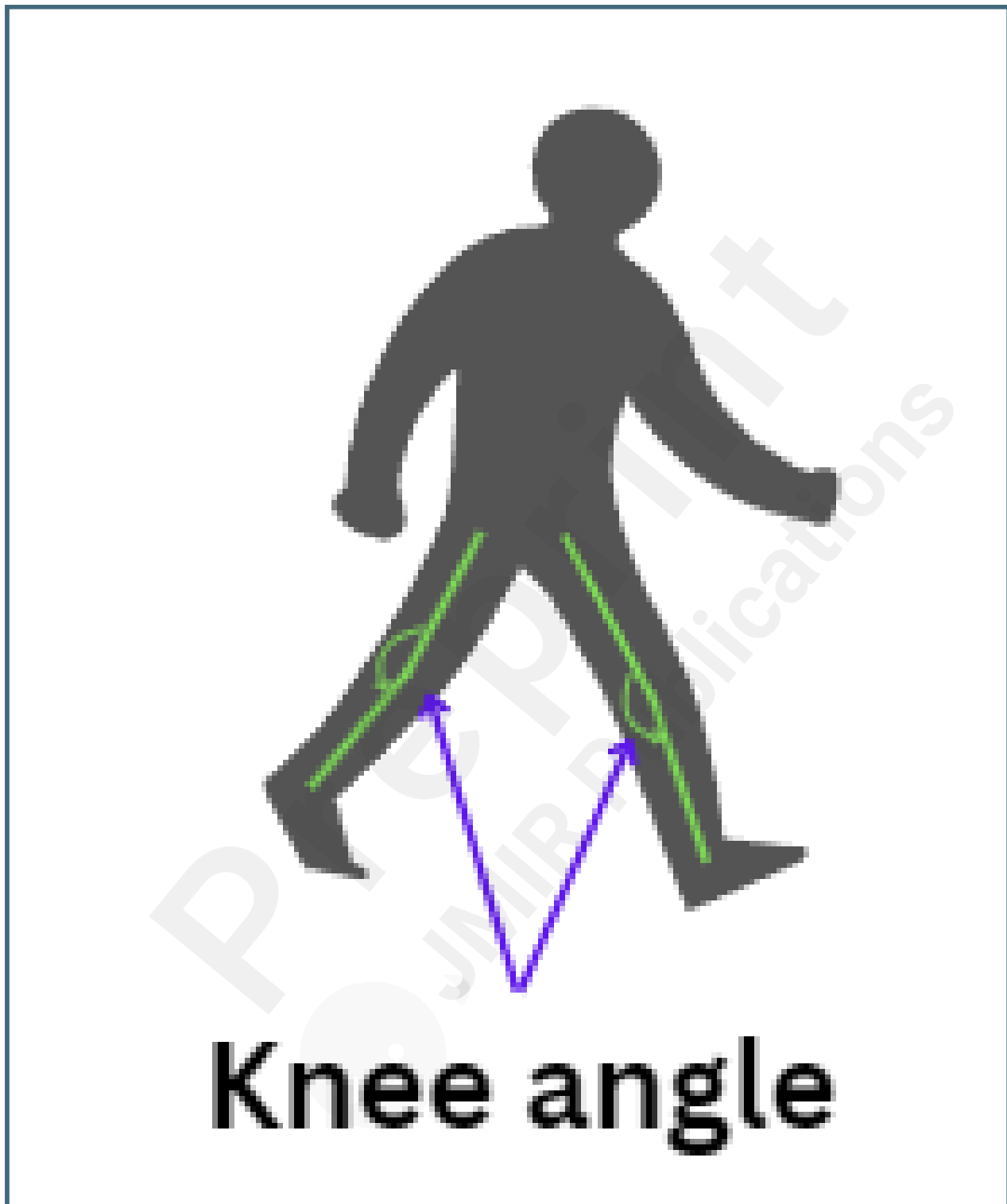


Fig. 11 Comparative Knee Angle between Group. (a) Young.

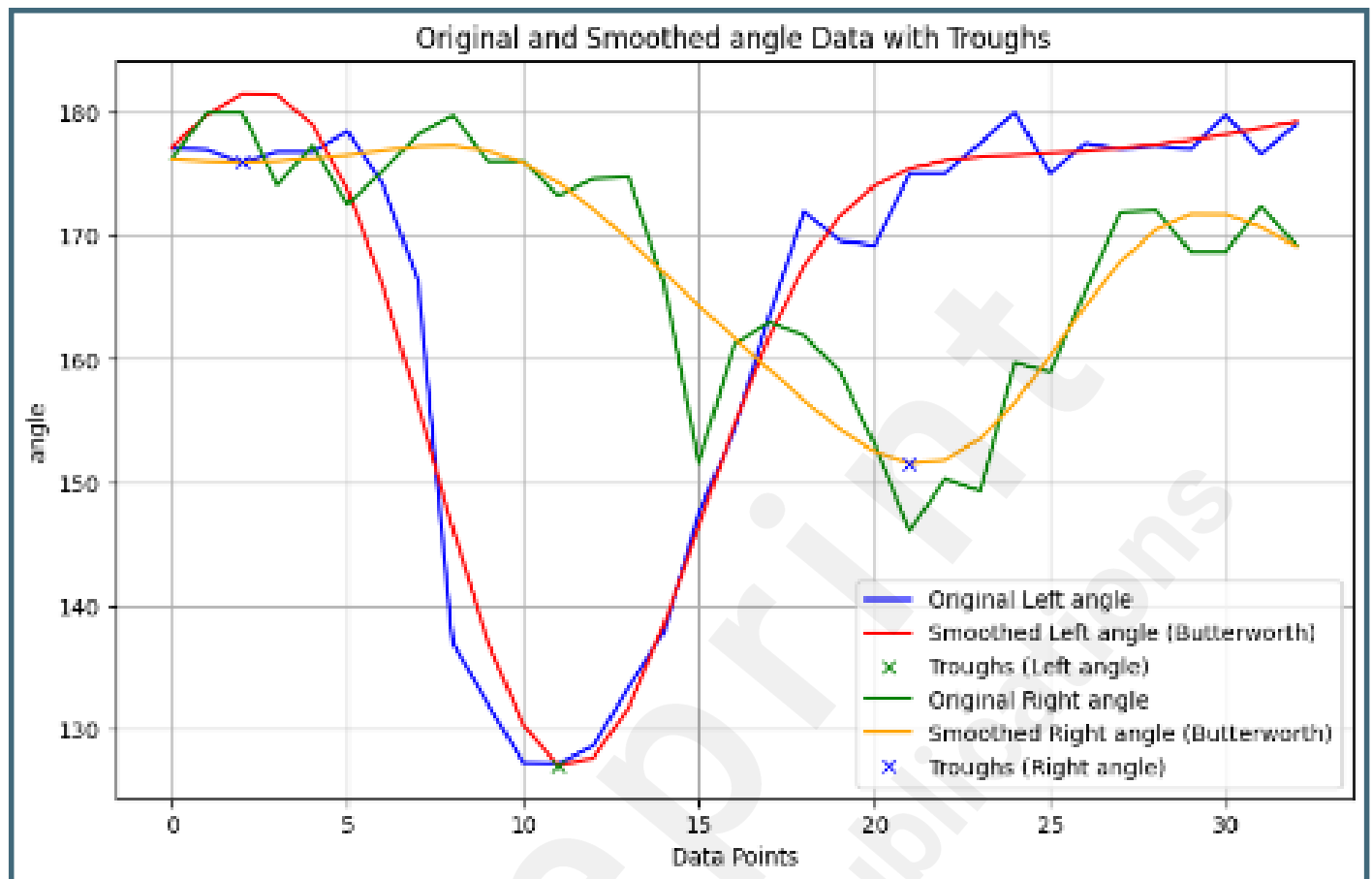


Fig. 12 Illustration of Hip Angle Position.

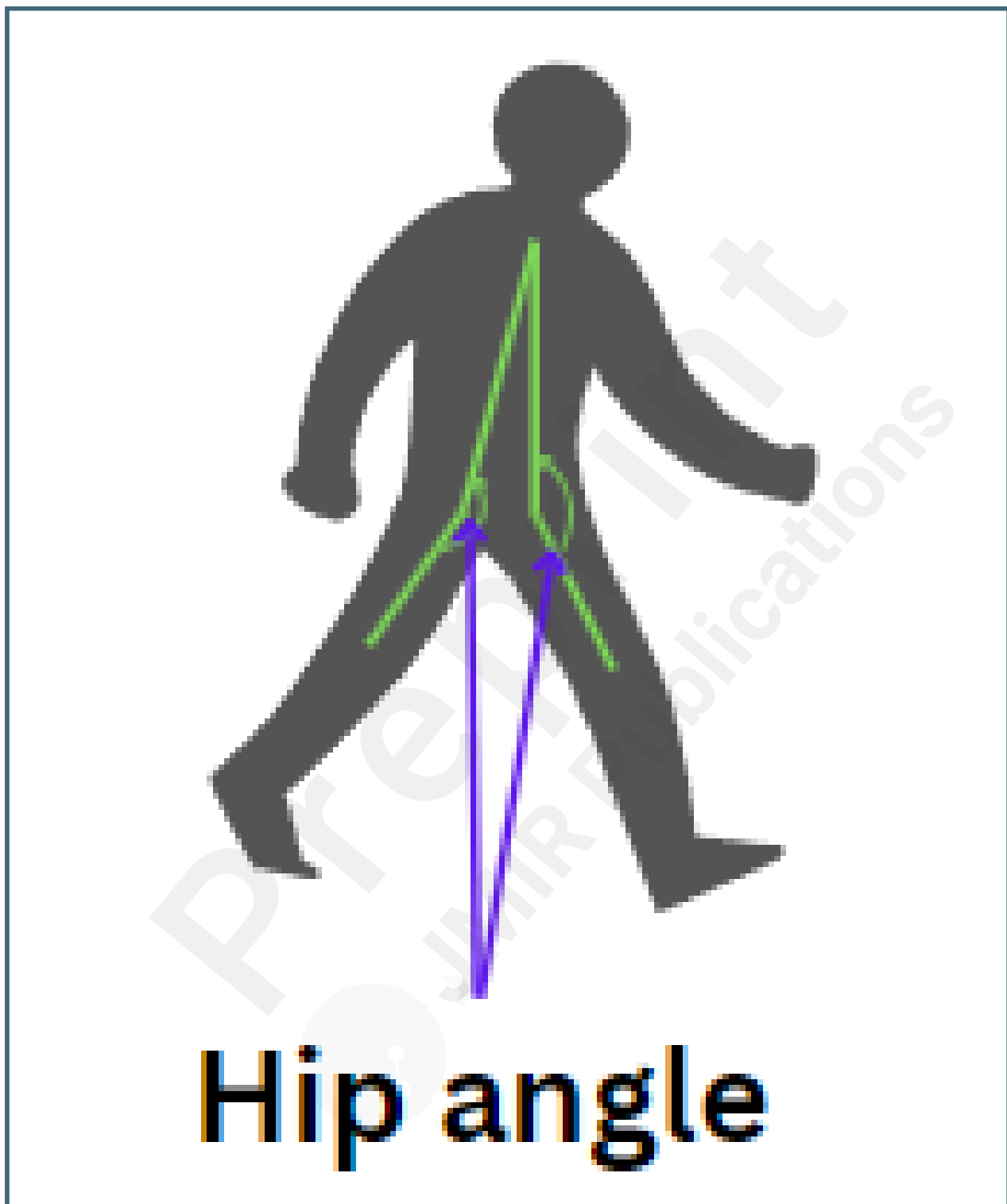


Fig. 13 Comparative Hip Angle between Groups. (a) Young.

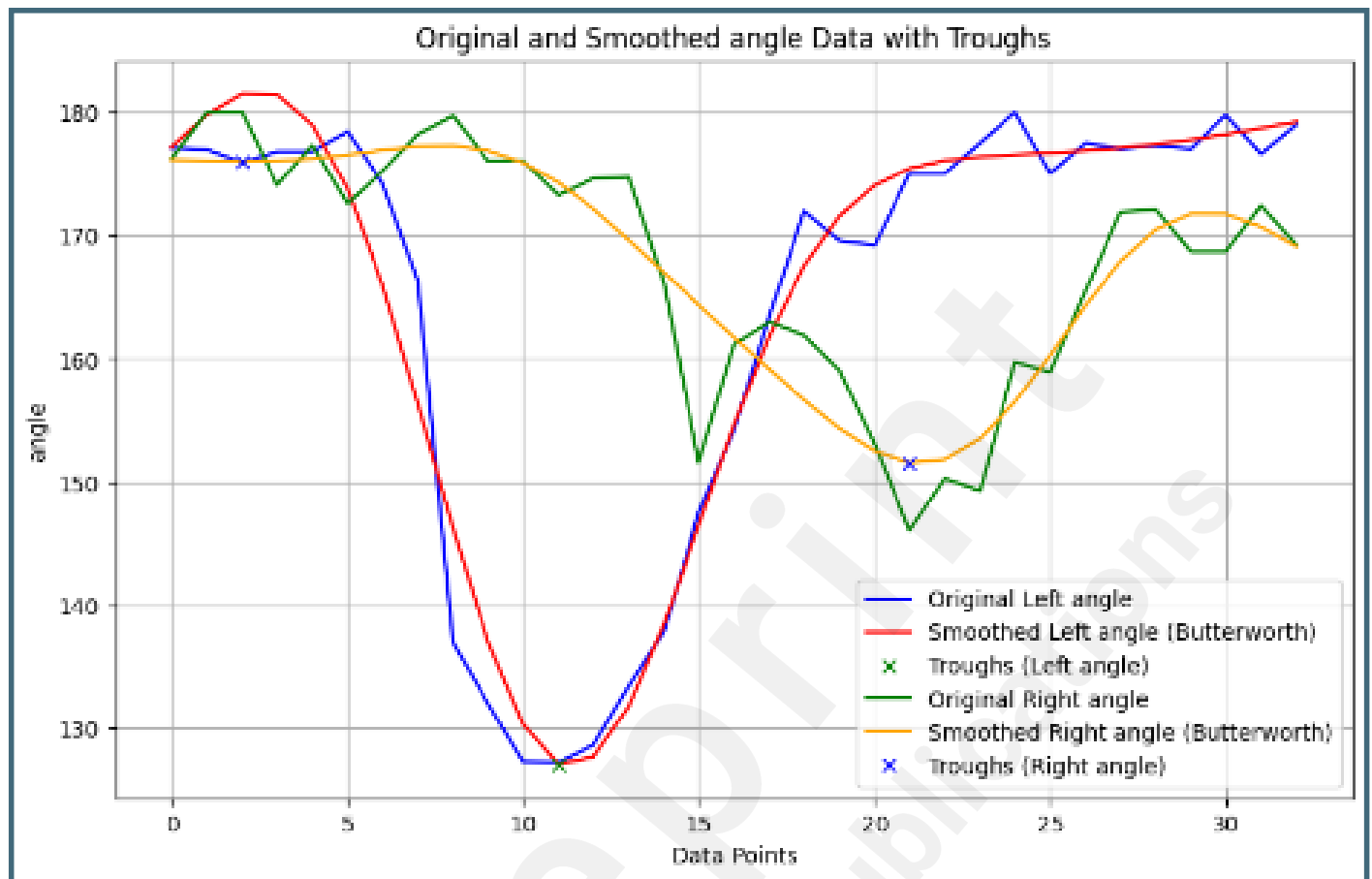


Fig. 14 Illustration of Symmetric Leg Position.

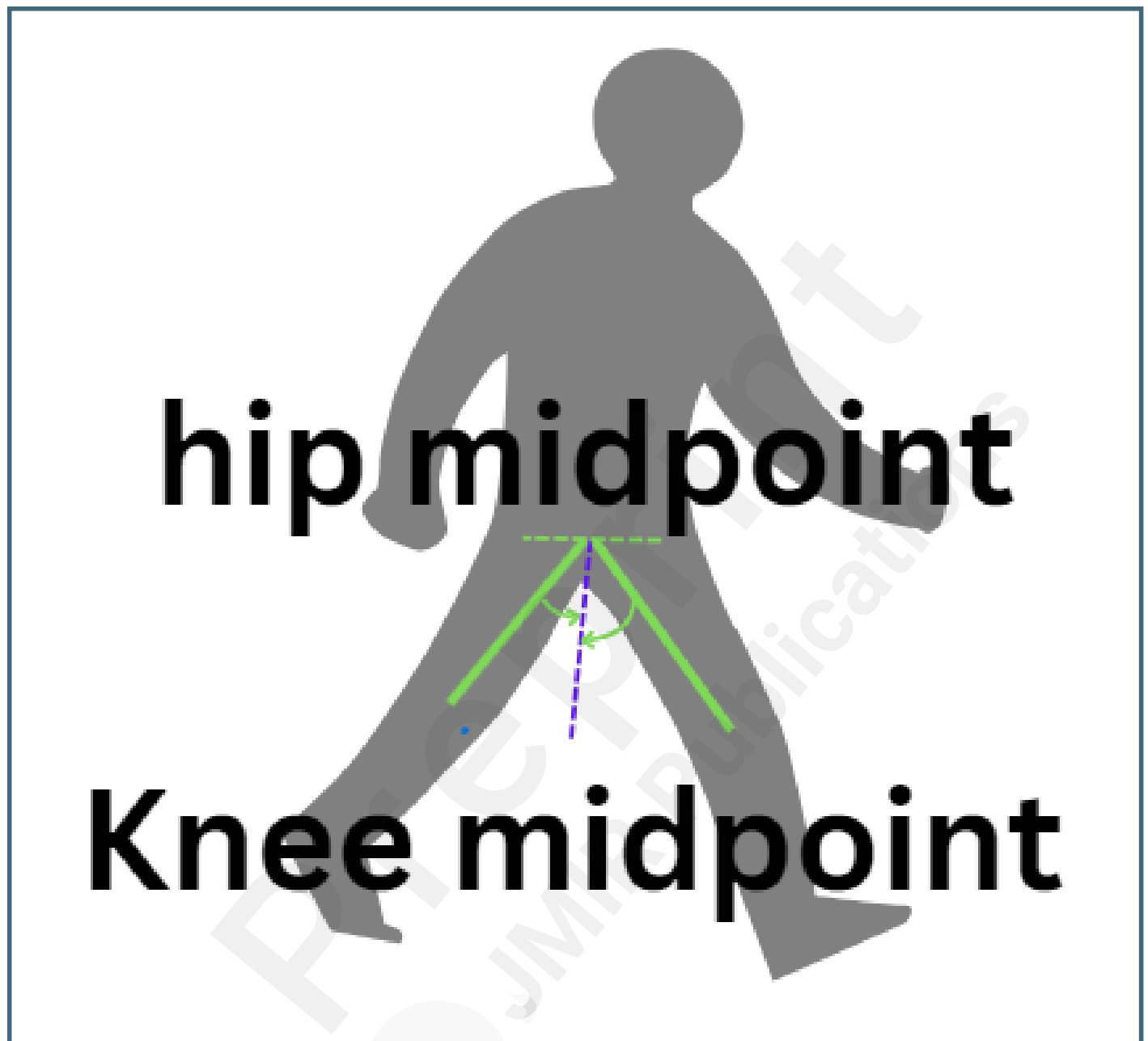


Fig. 15 Comparative Symmetric of Leg between Groups. (a) Young.

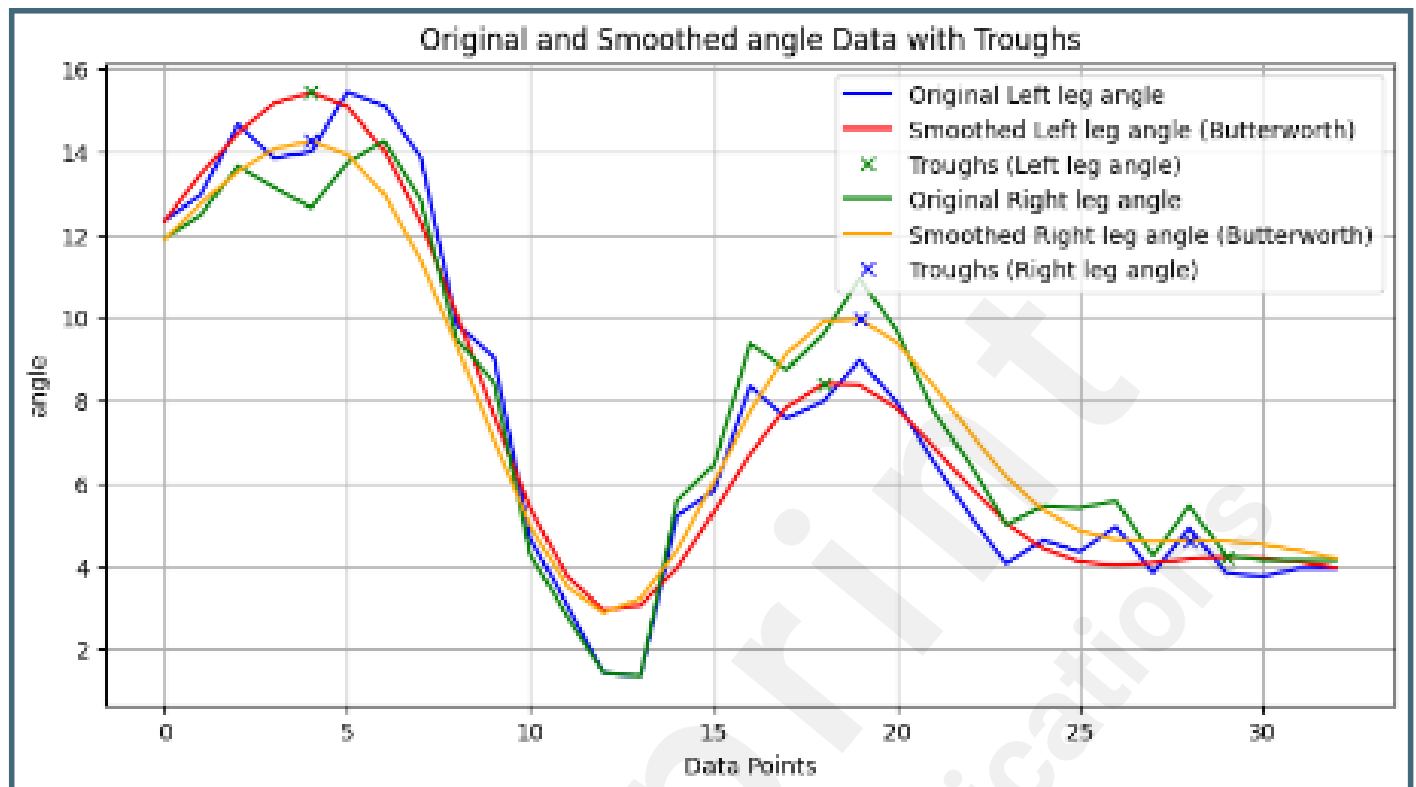


Fig. 16 Illustration of Symmetric Arm Position.

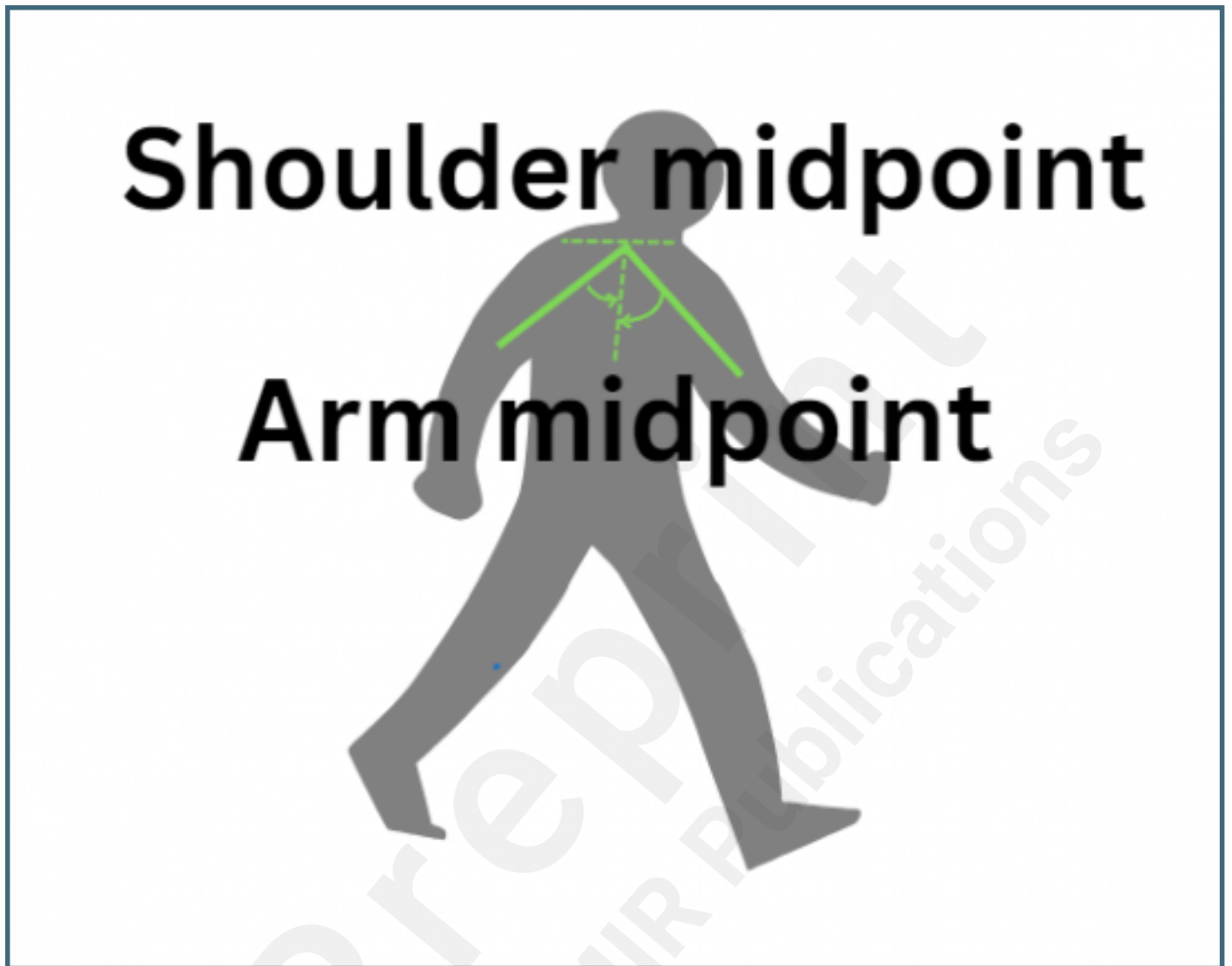


Fig. 17 Illustration of Trunk Angle I Vertical Position. (a) Young.

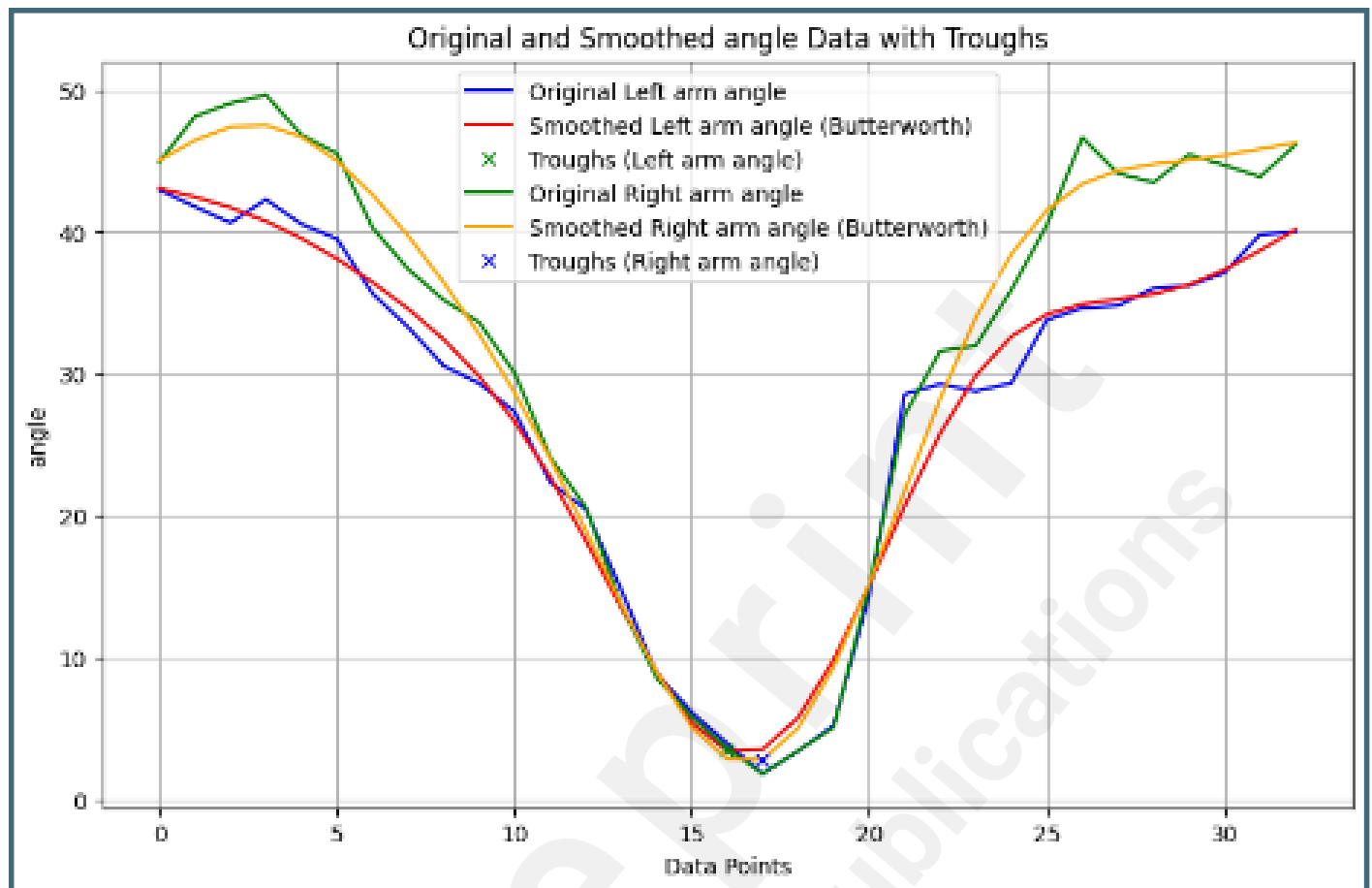


Fig. 18 Illustration of Trunk Angle I Vertical Position.

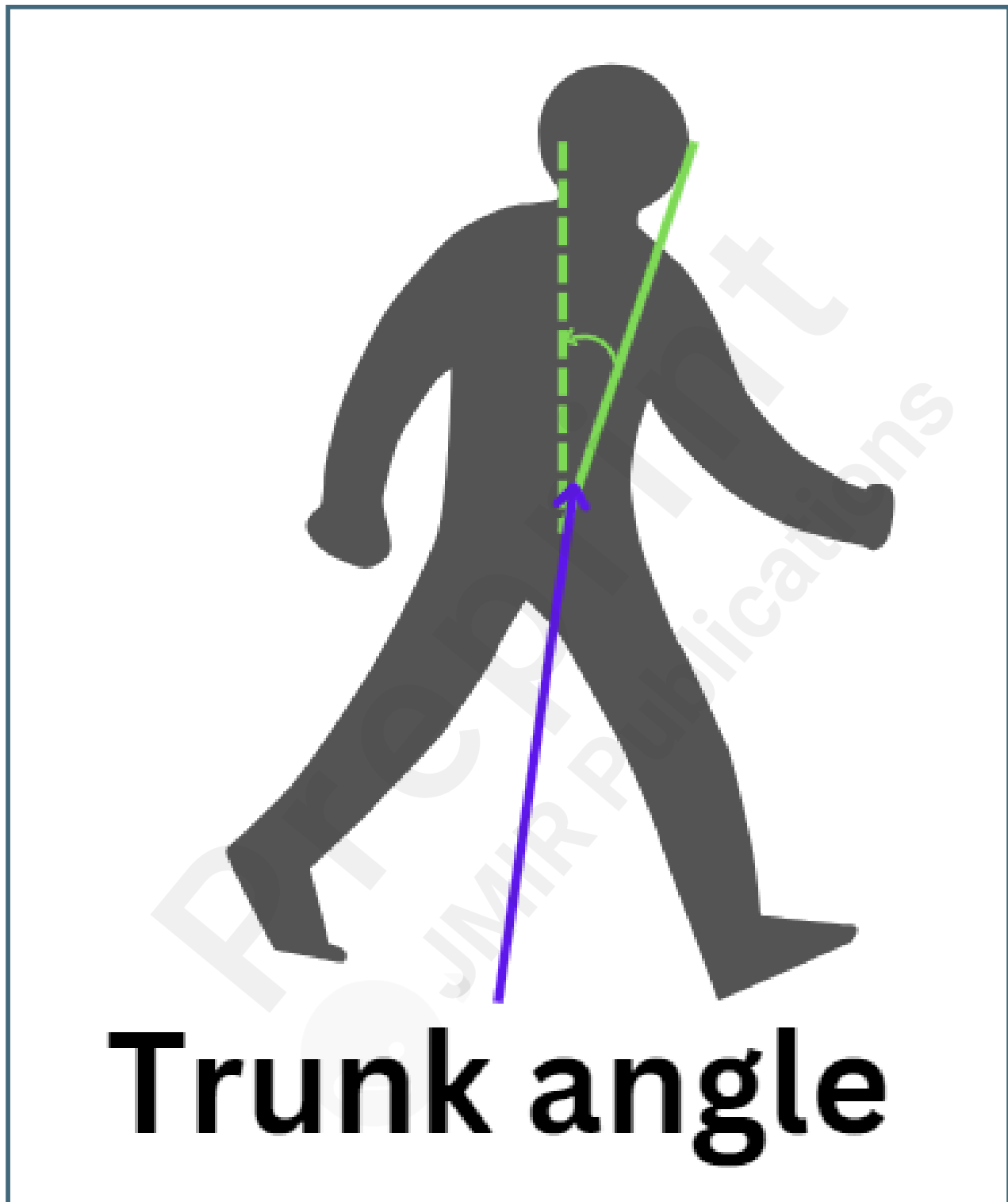


Fig. 19 Comparative Trunk Angle I between Groups. (a) Young.

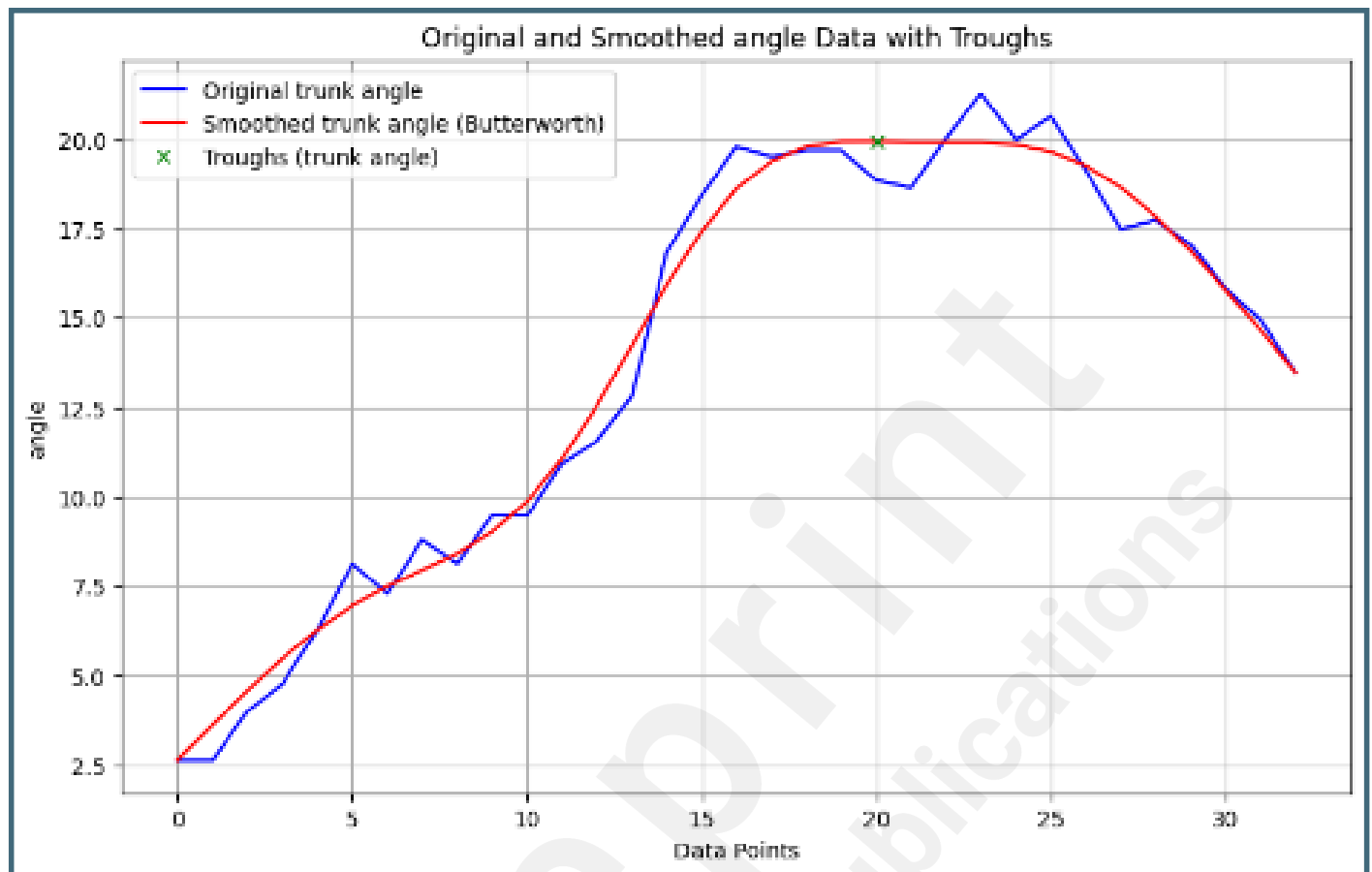


Fig. 20 Illustration of Trunk Angle II Horizontal Position.

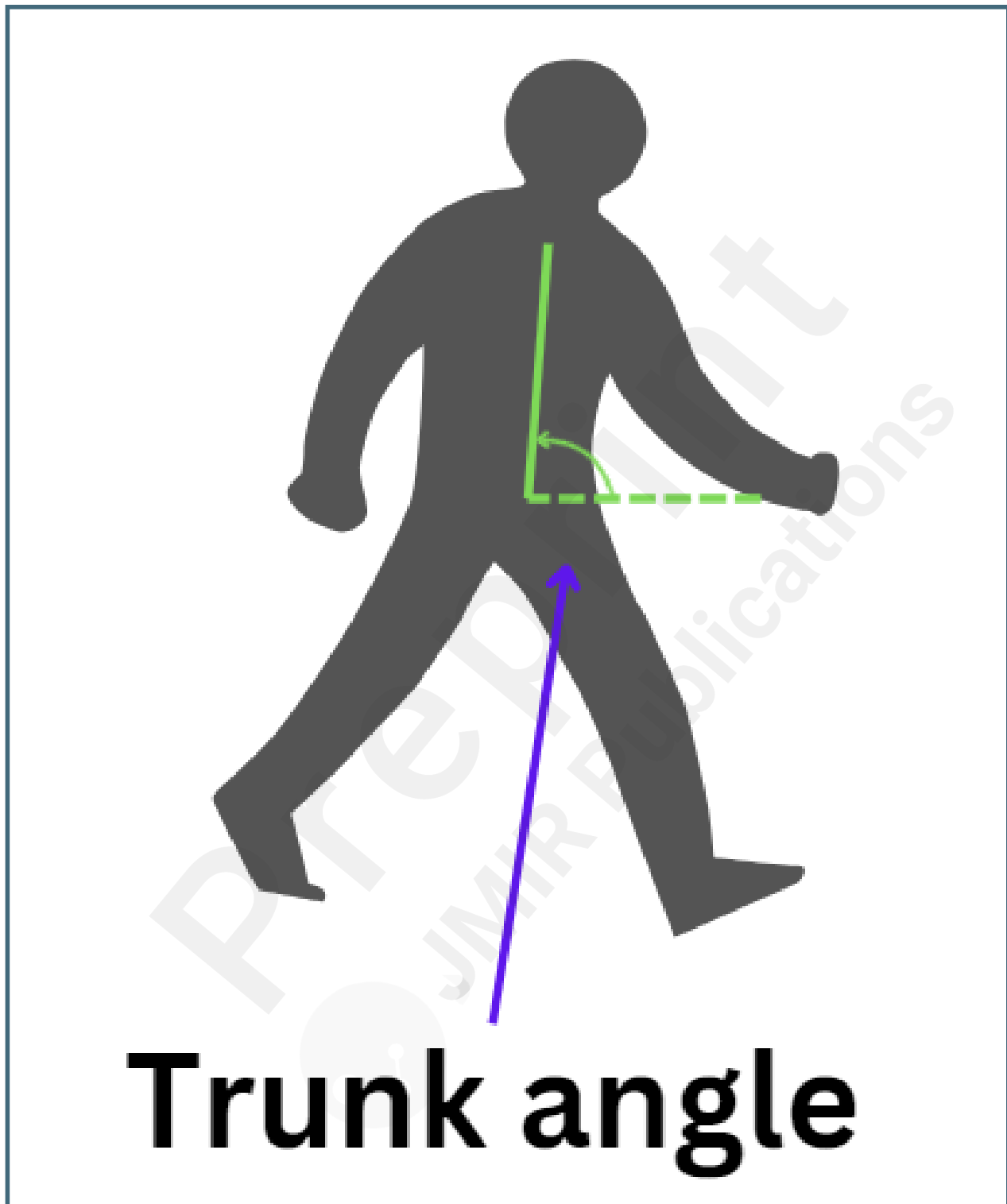


Fig. 21 Comparative Trunk Angle II Horizontal between Groups. (a) Young.

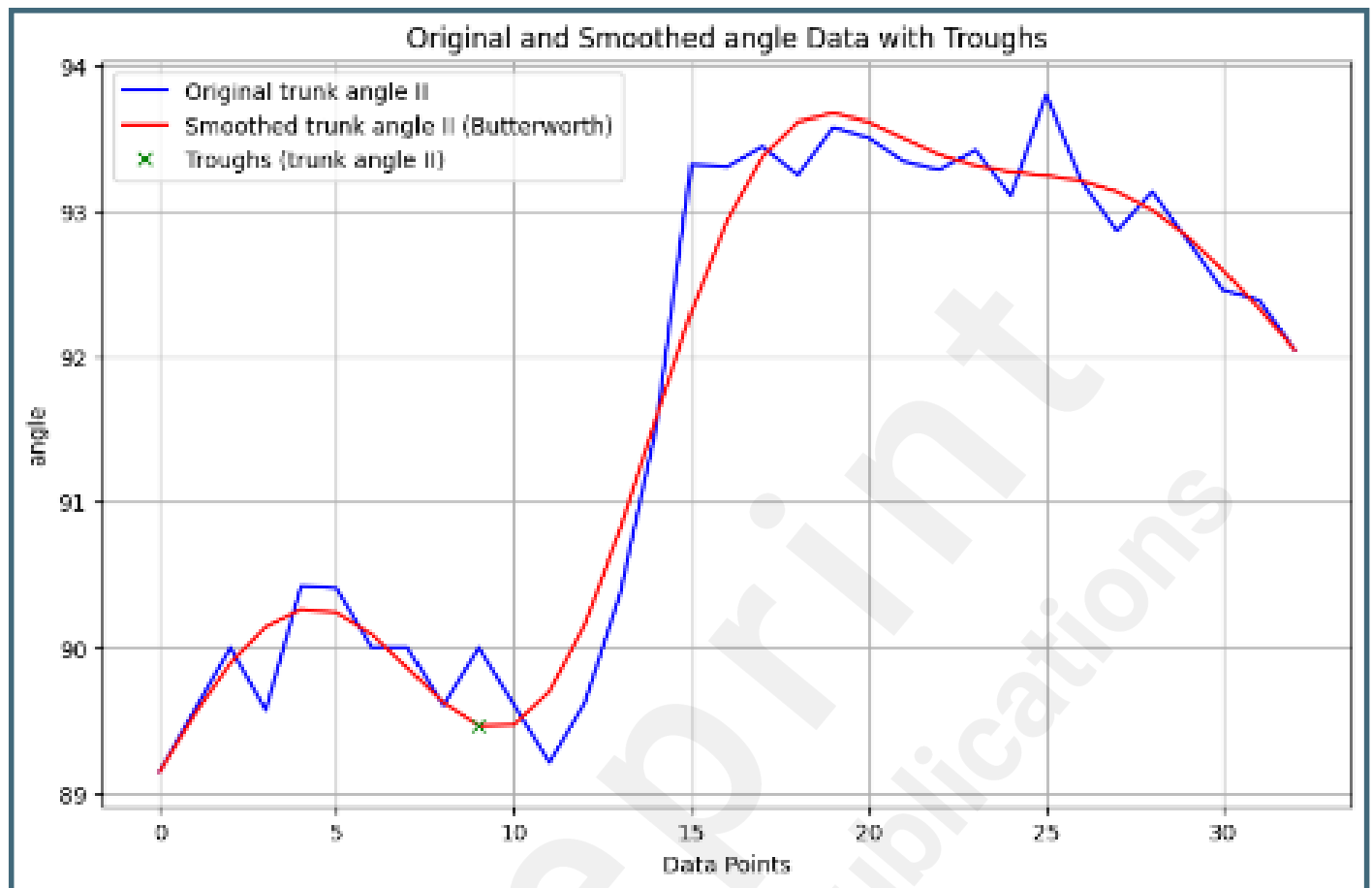


Fig. 22 Illustration of Shank Angle Position.

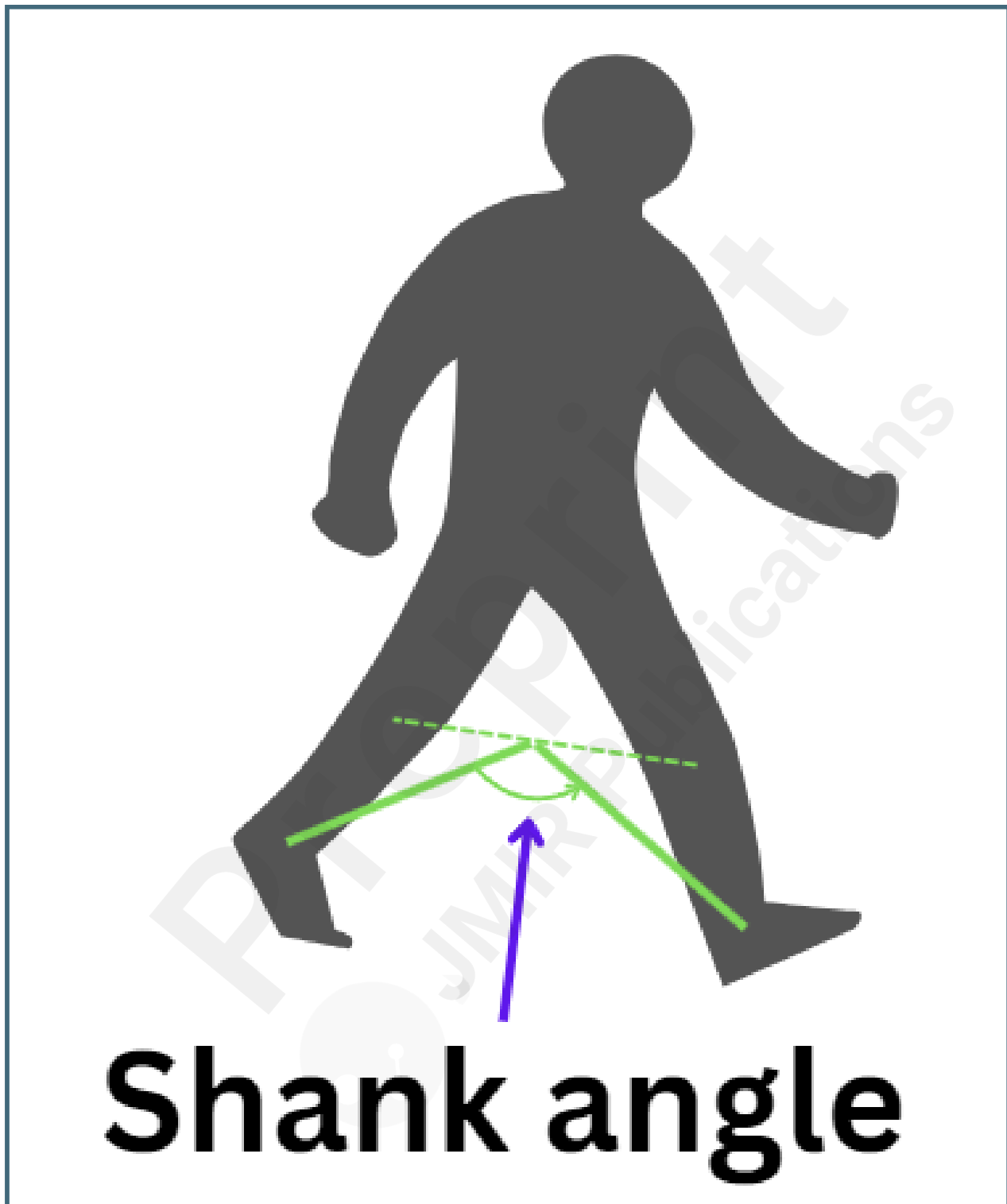


Fig. 23 Comparative Shank Angle between Groups. (a) Young.

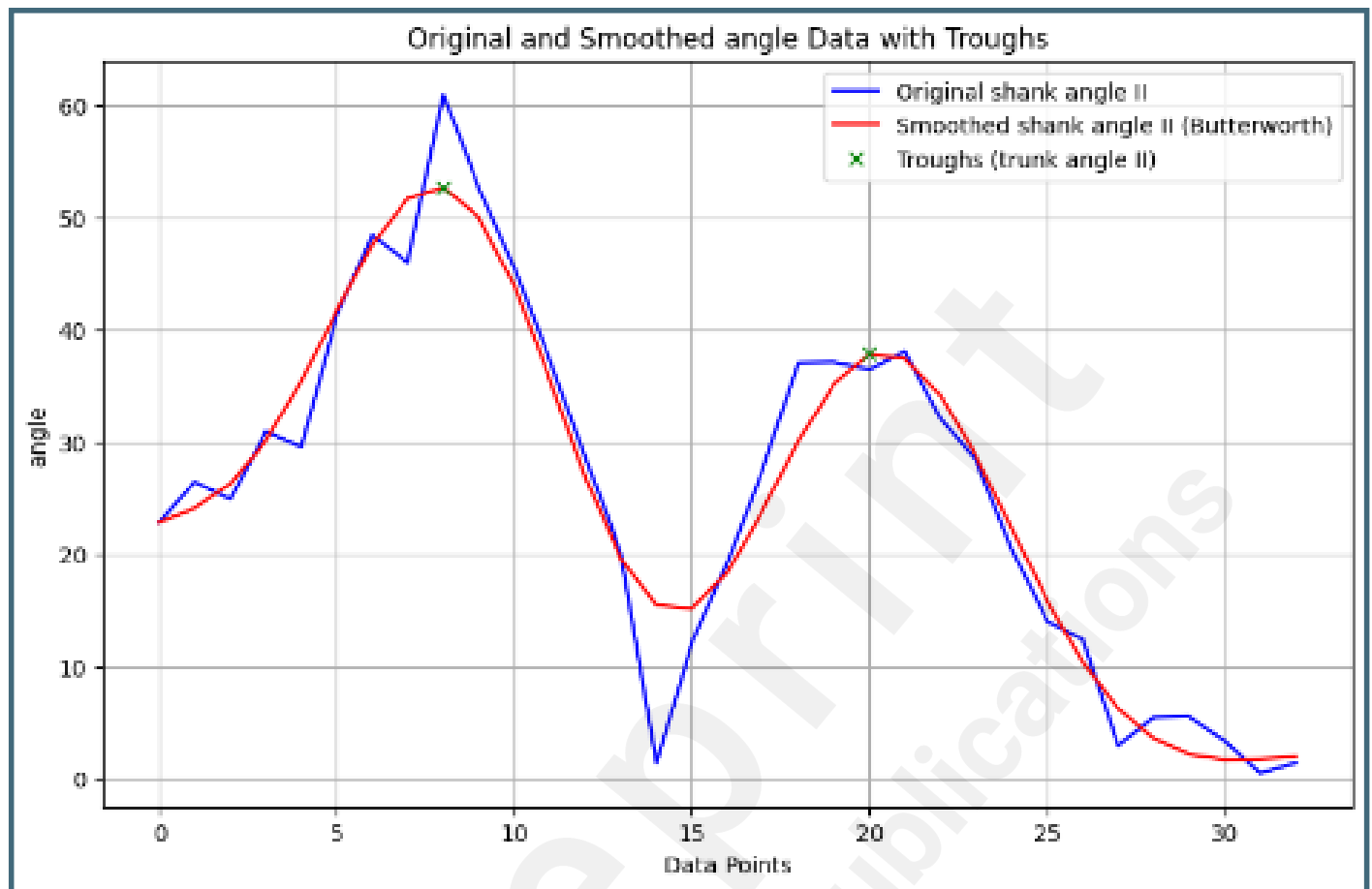


Fig. 24 Receiver Operating Characteristic (ROC) Curve.

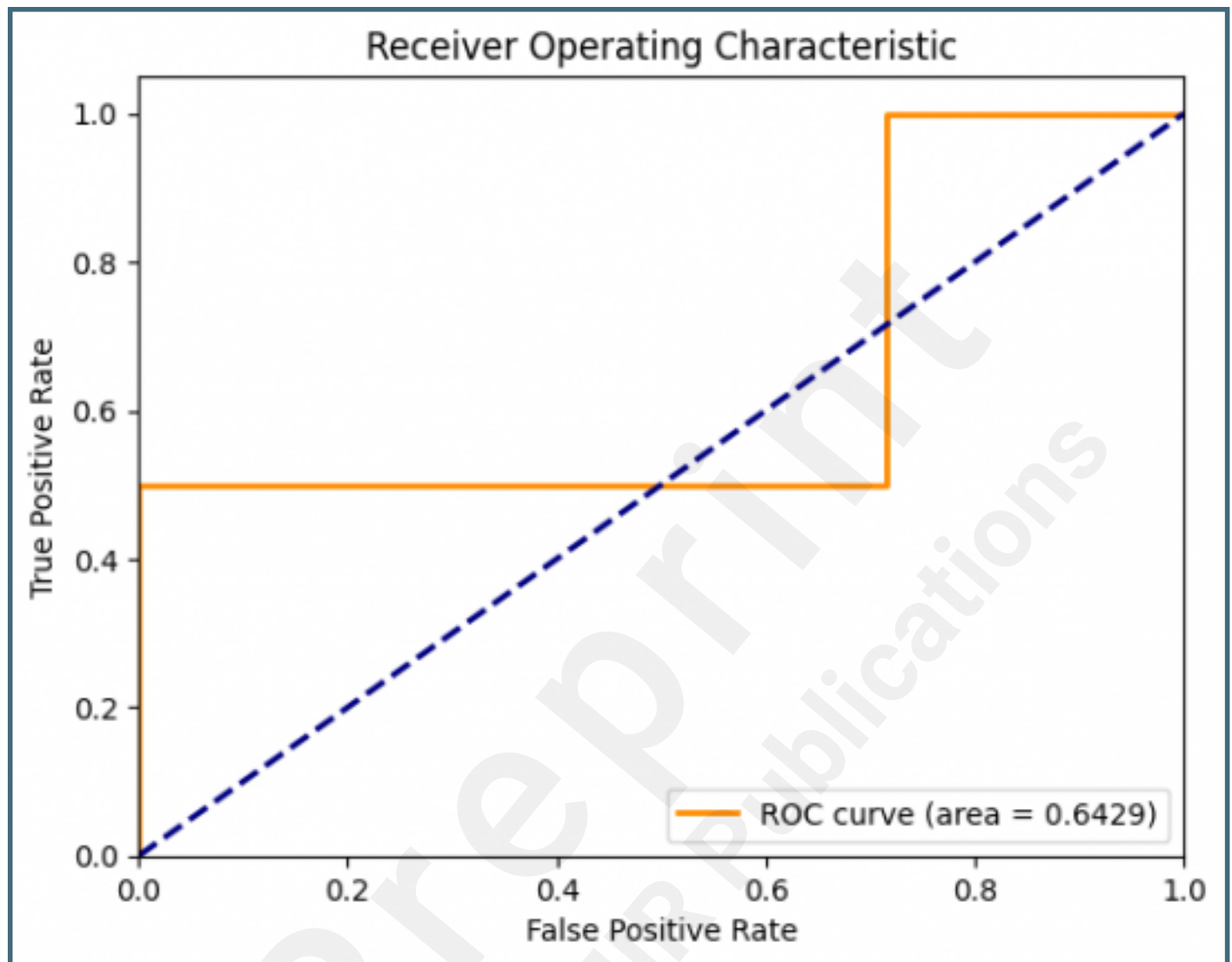


Fig. 25 Train/Validation Loss Curve.

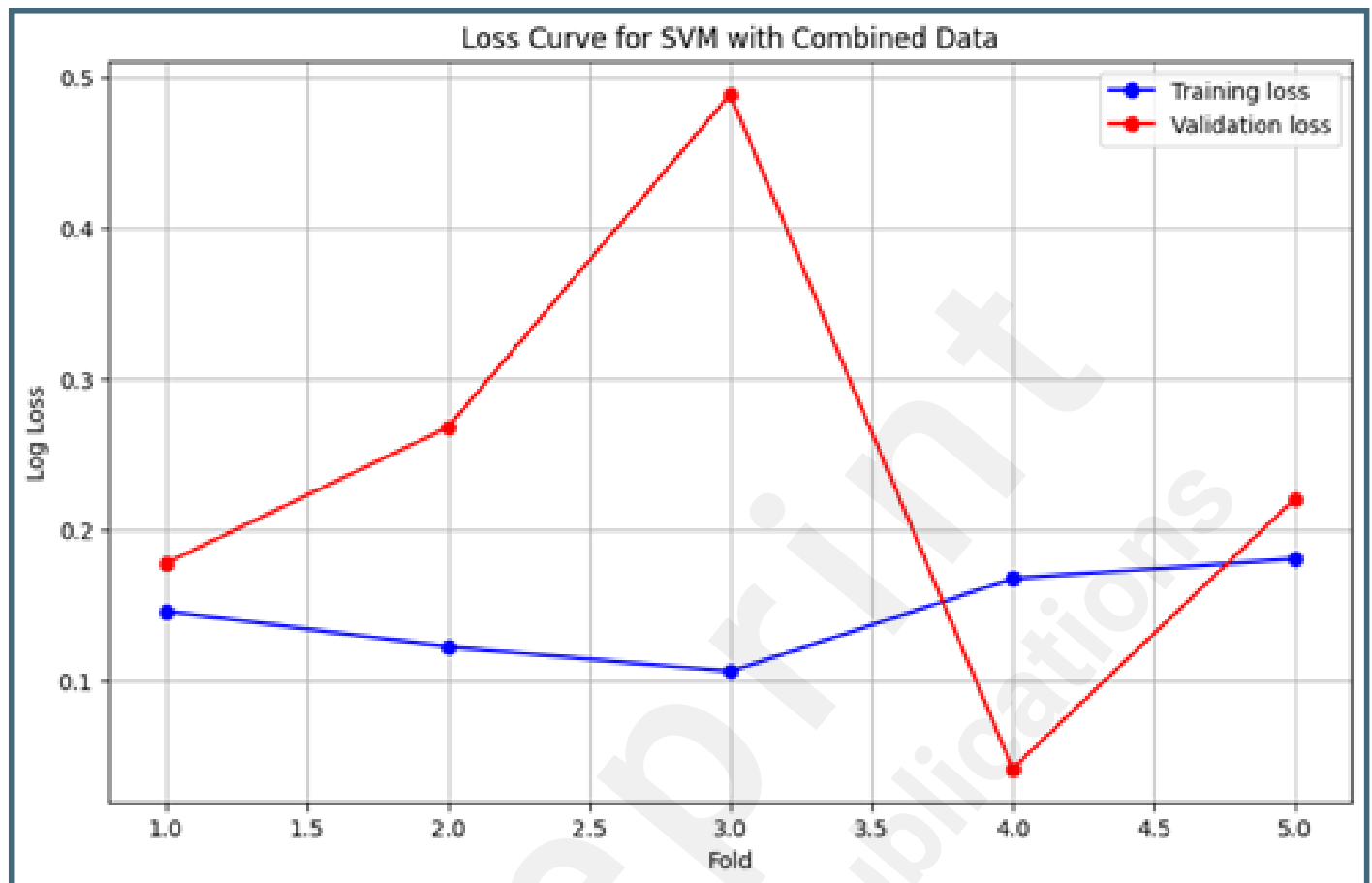


Fig. 26 Testing Set of Confusion Matrix for Model-based Features.

