Early Frailty Detection through Gait Feature Analysis using Machine Learning Classifiers

Xuan Yang Lim1, a), Tee Connie2, 3, b), Vincent Tan Wei Sheng2, c), Michael Kah Ong Goh2, 3, d) and Nor Izzati Saedon4, e)

1Tong Garden Food (Malaysia) Sdn Bhd, 79200 Iskandar Puteri, Johor, Malaysia

2Faculty of Information Science and Technology, Multimedia University, 75450 Melaka, Malaysia  
3Centre for Image and Vision Computing, CoE for Artificial Intelligence, Multimedia University, 75450 Melaka, Malaysia

4Department of Medicine, Faculty of Medicine, Universiti Malaya, 50603, Kuala Lumpur, Malaysia

b) Corresponding author: tee.connie@mmu.edu.my  
a) limxuanyang@tonggarden.com.my  
c) 1181101841@student.mmu.edu.my

*d) michael.goh@mmu.edu.my  
e) izzati@um.edu.my*

**Abstract.** Frailty in the elderly, marked by declines in physical and cognitive functions, increases risks such as falls and hospital admissions. This study explores early frailty detection through gait analysis using computer vision and machine learning. Human pose estimation tracks subjects in video sequences, utilizing the GSTRIDE dataset (IMU-based gait parameters) and the MMU dataset (video-based TUG tests). Important gait phases like heel-strike and toe-off angles are captured and evaluated. Cross-validation with Recursive Feature Elimination (RFECV) determines the most important features for classification. The gaps between frail and healthy individuals are bridged by machine learning models such as Multilayer Perceptron and Support Vector Machines. The classifiers highlighted are the ones tested experimentally. This study illustrates the promise of leveraging gait analysis, computer vision, and machine learning to create more affordable and non-invasive systems for earlier detection of frailty among the elderly.

# INTRODUCTION

Frailty describes a widely prevalent condition with growing incidence in older people. Physically, frail seniors show a gradual reduction in strength, endurance, and overall functioning of various interrelated body systems. The increased vulnerability results in a heightened risk of falls, hospitalization, and even death [1]. Improving the quality of life while decreasing the costs associated with healthcare services can be achieved through proactive measures. The Timed Up and Go (TUG) test along with clinical methods such as the Frailty Phenotype and Frailty Index are useful in detecting advanced stages of frailty. However, these methods are notoriously slow, require a trained professional to conduct them, and lack scalability for population-level deployment. With the global elderly population rapidly increasing, there is an urgent need for efficient, accessible, and automated frailty detection methods.

Gait analysis stands out as a potentially viable option, given that walking patterns provide a good reflection of an individual’s physical well-being. Studies show that frail individuals tend to have shorter steps, slower walking speed, and increased stride variation [2]. Recent developments in computer vision and machine learning enable the detailed extraction of gait features from videos, thus providing a non-invasive and cost-effective means of large-scale screening.

In this research, we focus on the combination of gait feature analysis with machine learning algorithms to identify the preliminary signs of frailty. We make use of two datasets that complement each other: GSTRIDE, which captures gait parameters through Inertial Measurement Units (IMUs), and MMU, which contains video recordings of the TUG test for pose-based gait feature extraction. A number of machines learning classifiers, including Multilayer Perceptron (MLP) and Support Vector Machines (SVM), are applied together with Recursive Feature Elimination with Cross-Validation (RFECV) for optimized feature selection. The experiments conducted demonstrate the effectiveness of the proposed approach.

# MATERIALS AND METHODS

## GSTRIDE Dataset

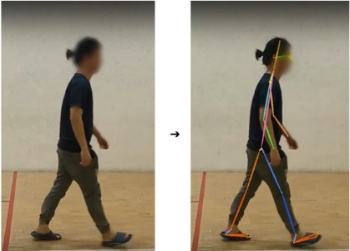
This study relies on the GSTRIDE dataset [2,3] created by Guillermo et al. which is an important database. It includes information from 163 patients covering demographics, body composition, physical and mental capabilities as well as cognitive condition. The IMU sensors that measure walking gait provide data on stride length, step speed, and cadence. The sensors offer raw and calibrated outputs. This dataset implements algorithms derived from the TUG test and the Short Physical Performance Battery (SPPB) test for classification into frail or non-frail categories. Frail individuals are marked with a TUG score exceeding 13 seconds alongside a Frailty Index greater than 2. The data encompasses a monitored duration of approximately 21.4 ± 7.1 minutes.

## MMU Dataset

The recordings of the Timed Up and Go (TUG) Test are captured in the MMU Dataset, which utilizes two cameras set at a height of 1.25 meters for both frontal and lateral angles [4,5]. The subjects commenced from a seated position, proceeded to walk a short distance to turn around, and subsequently sat down. This movement sequence enabled comprehensive analysis of gait features. The dataset includes healthy participants and patients with Parkinson's disease, comprising 25 individuals between the ages of 20 and 88 years. This dataset is advantageous due to the diversity of age, health status, and walking patterns. Furthermore, all participants were health profiled through a frailty assessment using the Frailty Phenotype Score. Data collection occurred from May 21 to July 31 in 2024. For participation, all subjects received thorough explanations of the study's objectives, procedures, associated risks, and potential benefits during the informed consent process to ensure comprehension. Consent was documented by a co-signing researcher, securely stored in a digital system devoid of personally identifiable information (PII) which maintained confidentiality.

## Human Pose Tracking

AlphaPose (AP) [6-8] is used to capture detailed posture and body movements during the TUG test for the MMU dataset. AP is a reliable multi-person pose estimator that tracks key points of the human body across video frames. It consistently tracks the movement of the same individual through the video sequence. Figure 1 and Table 1 depicts the human body key points located by AP and their descriptions.



**Figure 1.** Keypoints estimated using Alphapose algorithm

**TABLE 1.** Description of body key points.

|  |  |
| --- | --- |
| **Body Key Points** | **Description** |
| Heels (LHeels, RHeels) | Represents the posterior feet, essential for stability and weight distribution |
| Knees (LKnee, RKnee) | Provides essential insights on knee joint dynamics during gait process along with the comprehension of stability and fluidity in the gait pattern |
| Ankles (LAnkle, RAnkle) | Capture ankle movement data that facilitates the push-off phase during gait process and contributes to overall propulsion and gait stability |
| Big Toes (LBigToe, RBigToe) | Prominent digits for the foot that provides toe-off and propulsion during gait cycle |
| Small Toes (LSmallToe, RSmallToe) | Prominent digits for the foot that provides toe-off and propulsion during gait cycle |

## Walking Phase Identification

We aim to differentiate between walking and non-walking frames in the video, keeping only the frames with walking patterns. Towards this end, key body points (head, hip, and big toe) were plotted across consecutive frames. These phases were detected by observing significant slopes in the key point positions. In particular, the head key points were used to detect different phases based on changes in graph trends and slope angles. Any graph segments with slopes greater than 3 degrees were flagged and removed. By plotting the inverted y-coordinates of the head and big toe key points, we measured the distance between them, representing the subject's height in pixels. This measurement was then used to calculate the calibration factor, which converts pixel data into meters. The calibration factor, in pixels per meter, is calculated using the Equation (1):

|  |  |
| --- | --- |
|  | (1) |

## Gait Extraction

Gait extraction is a pivotal step in obtaining gait data from raw positional data from body key points. Hence, the computation and conversion methods are required to be precise and accurate to capture detailed insights from the complex human gait dynamics.

### Steps and Strides

To identify steps taken during the TUG test, we analyzed the Euclidean distance of mid-toe positions for both legs over time, marking filtered peaks for clarity. The Euclidean distance is computed by Equation (2),

|  |  |
| --- | --- |
|  | (2) |

Peaks were identified by comparing their values to the median, removing those with values less than 20 units below the median to eliminate inconsistencies. The remaining peaks indicated specific frames where steps were taken. The graph in Figure 2 (a) shows a generally stable movement, with drops in values around the 100th and 200th frames due to turnaround phases. The number of steps was calculated by counting these peaks. The graph shows the Euclidean distance of mid-toe positions for both legs over time, with filtered peaks marked. The steps taken can be visualized as the peaks.

|  |  |
| --- | --- |
| A graph of a graph  Description automatically generated with medium confidence  (a) | (b) |

**Figure 2.** (a) Smoothed euclidean distance of MidToe for both legs over time with filtered peaks marked. (b) MidToe X key points for both legs

The frames for each step are extracted by identifying and marking the peaks where the steps occur. A graph visualizing mid-toe X key points over time is plotted. This graph contains information on the horizontal distance covered by each leg over the duration of the video. In the analysis, each leg’s step frames are marked and the distance between steps and strides is calculated. Step and stride lengths are calculated, and mean values are derived for further analysis (Figure 2 (b)). Each peak shows the movement pattern of each leg and the consistent peaks show stepping events. As described above, the graph shows the horizontal distance traveled by each leg, while noticeable stride length is also marked between each two consecutive peaks. The positions of mid-toe X vertically yield a wave-like pattern during the TUG test which demonstrates the gait cycle. The results also describes some of the dynamics of the participant’s walks like the consistency of steps and stride length. The gait features extracted by the computation of steps and strides provide the ability to derive speed, time and distance traveled to calculate other features. These computed features enhance the analysis by adding time, distance and speed as variables during the assessment of walking dynamics. These features help in understanding the participants gait as well as their stability.

## Feature Analysis

Feature analysis is a crucial step in studying and analyzing frailty. Understanding the gait patterns of different groups of individuals, i.e. healthy (both young and old) and frail, provides valuable insights into identifying frailty. This is clearly illustrated in Figure 3.

A group of graphs showing different types of data

AI-generated content may be incorrect.

**Figure 3.** Comparison of steps taken of different individuals during the TUG test. The steps of a younger person are represented in the top-left graph while the top-right graph shows an older person’s steps. The bottom graph in the figure depicts the steps of a frail individual

The top-left graph depicts the steps of a young healthy adult and the top-right graph illustrates the older healthy adult’s steps. The bottom graph shows a frail person's steps. Unlike the other graphs, the frail individual's graph was complex, containing a lot of fluctuation between steps, which meant more steps with an irregular gait. In addition, the x-axis which shows the frames taken to accomplish the TUG test, reveals that healthy persons perform the test in a very short duration while the frail person takes much longer, which is illustrated by the greater number of frames.

## Joint Dataset

Considering the available features extracted in the MMU dataset, a subset of features related to gait is chosen to merge with the GSTRIDE Database to create an all-inclusive and varied dataset. Table 2 displays the features contained in the Joint Dataset.

**TABLE 2.** Gait features of Joint Dataset.

|  |  |
| --- | --- |
| **Gait Features** | **Unit** |
| Strides | Strides |
| Mean stride time | Seconds |
| Mean step speed | ms-1 |
| Mean stride length | Meters |
| Cadence | Strides/min |
| Heel-strike angle | Degrees |
| Toe-off angle | Degrees |

## Feature Selection

This section explores the rationale for feature selection using the GSTRIDE and Joint datasets to ensure the chosen gait features are strongly correlated with the frailty score. The goal is to select an optimal number of features that enhance model performance. Recursive Feature Elimination with Cross-Validation (RFECV) is employed, utilizing Logistic Regression (LogR) to iteratively filter features and determine the optimal set. The selected features are based on model coefficients and important scores. Figure 4 (a) shows the cross-validation results of the Logistic Regression (LogR) model tested with different numbers of features. For the GSTRIDE database, the optimal number of features selected is 11. Strides, Foot Flat, Push, Swing, Toe-off Angle, Mean Load, Heel-strike Angle, Step Speed, Mean 3D Path, and Clearance are features yielding the highest scores. The cross-validation results of the Logistic Regression model from the Joint dataset with varying number of features used are shown in Figure 4(b). Selecting 4 features was the best result. The highest scored features are Stride Time, Toe-off Angle, Step Speed, and Stride Length.

|  |  |
| --- | --- |
| (a) | A graph with a line  Description automatically generated  (b) |

**Figure 4.** (a) Number of optimal features selected for GSTRIDE Database. (b) Number of optimal features selected for Joint dataset

## Classifiers

This study explores various classifiers in relation to the efficacy of frailty detection among the elderly using gait characteristics. Logistic Regression (LR) is a binary classification algorithm which is a commonly used and simple technique. It employs a weighted summation method to forecast outcomes, using features and components of the model as predictors [11]. Random Forest (RF) is an ensemble approach that improves accuracy by building multiple decision trees and aggregating their predictions. It performs well on multifaceted datasets and provides important information regarding feature relevance RF is an ensemble approach that improves accuracy by constructing multiple decision trees and aggregating their outputs [10]. The Multilayer Perceptron (MLP) is foundational in artificial neural networks and solves intricate classification challenges through the detection of non-linear interconnections across numerous neuron layers.[11] Support Vector Machines (SVM) fall into the category of supervised classifiers, but they also excel in regression tasks. There are recent modifications such as Bayesian optimization that set SVMs parameters for classifying Parkinson’s disease [12] and the GL-TSVM variant which increases robustness to outliers in biomedical datasets [13],[14].

# Results and Discussions

Analyzing the classification results from accuracy, F1 score, sensitivity, and recall, the classifiers performance based on confusion matrices [9]. The models’ capabilities to classify frailty were evaluated using extracted gait features and their corresponding confusion matrices. The best-selected features and highest scoring models were evaluated with tuned hyperparameters that were tailored to the structure of the dataset to achieve the best results.

## Experiment with GSTRIDE Database

As seen in Figure 5 (a), the classifiers result from the GSTRIDE Database have been outlined. Each of the classifier’s strengths and weaknesses for frailty prediction through gait features are described. MLP has the best accuracy and recall among the classifiers at 0.87 and 1.00, respectively. This is because the architecture of MLPs or multi-layered perceptron enables them to model complex non-linear dependencies in data. He still overfitting might be a problem. SVM performs reasonably as well, with accuracy and recall of 0.80 and 0.82, respectively, which can be attributed to its utility in finding optimal hyperplanes for classification. Logistic Regression (LR) also achieves reasonable accuracy and recall numbers of 0.77 and 0.82, respectively, which he obtained from appropriately modeling binary outcomes and in this case, capturing the relationship between gait features and frailty. Random Forest (RF) was also competitive with accuracy and recall numbers of 0.73 and 0.82, respectively as he incorporates multiple decision trees to improve stability RF, however, struggles with high dimensional data. These findings suggest that, although there are general preferences towards MLP and SVM, LR and RF present strong alternatives which sometimes outperform their more sophisticated peers due to the inherent properties of the algorithms and the data.

|  |  |
| --- | --- |
| A group of colored bars  AI-generated content may be incorrect.  (a) | A group of colored bars  AI-generated content may be incorrect.  (b) |

**FIGURE 5.** (a) Performance metrics of classifiers using GSTRIDE Database. Each performance metric is illustrated in an individual graph for all 4 classifiers. (b) Performance metrics of classifiers using Joint dataset. Each performance metric is illustrated in an individual graph for all 4 classifiers

## Experiment with Joint Dataset

In Figure 5 (b), the performance metrics for the four classifiers are displayed based on the experiment conducted with the Joint Dataset. Their specific advantages and disadvantages in determining frailty from gait features are underscored. LR and SVM outperform the other classifiers in accuracy and recall (0.69 and 0.78), primarily due to SVM's strong performance with binary classification problems and the interplay of gait features and frailty. SVM's strong performance is largely due to its ability to find optimal hyperplanes, and LR excels because of it’s strong modeling of binary-depending variables. MLP's accuracy and recall, while lower (0.66 and 0.72, respectively), may stem from overfitting despite its moderately strong results due to the ability of neural networks to capture non-linear relationships. RF’s moderate frailty recall of 0.67 and accuracy of 0.66 stems from its ensemble nature of combining multiple decision trees. While these results indicate the overall superiority of LR and SVM, the poorer performance of MLP and RF illustrates the competitive nature of these algorithms and their dependence on data characteristics.

# Conclusion

This research explores the use of gait analysis and advanced machine learning techniques to assess frailty among the elderly. For this study, the GSTRIDE and MMU datasets were utilized to extract specific gait parameters that differentiate healthy individuals from frail elderly participants. Out of the classifiers that were tested, Multilayer Perceptron (MLP) and Support Vector Machines (SVM) performed best, particularly MLP which outperformed the others in accuracy and recall. Logistic Regression (LR) and Random Forest (RF) classifiers also showed a moderate level of accuracy. The application of Recursive Feature Elimination with Cross-Validation (RFECV) to enhance the models by eliminating irrelevant features enhanced the performance of all the models. Further work will focus on augmenting the datasets and refining the models to evaluate this unobtrusive and low-cost approach towards improving eldercare.

# References

1. J. M. Pritchard, C. C. Kennedy, S. Karampatos, G. Ioannidis, B. Misiaszek, and S. Marr, “Measuring frailty in clinical practice: A comparison of physical frailty assessment methods in a geriatric out-patient clinic,” BMC Geriatrics **17**(1), 1 (2017).
2. S. García-de-Villa, G. G. V. Neira, M. N. Álvarez, E. Huertas-Hoyas, L. R. Ruiz, and A. J. del-Ama, “A database with frailty, functional and inertial gait metrics for the research of fall causes in older adults,” Scientific Data **10**(1), 1 (2023).
3. M. N. Álvarez, C. Rodríguez-Sánchez, E. Huertas-Hoyas, G. García-Villamil-Neira, M. T. Espinoza-Cerda, and L. Pérez-Delgado, “Predictors of fall risk in older adults using the G-STRIDE inertial sensor: an observational multicenter case–control study,” BMC Geriatrics **23**(1), 1 (2023).
4. B. Friedrich, S. Lau, L. Elgert, J. M. Bauer, and A. Hein, “A deep learning approach for TUG and SPPB score prediction of (Pre-) frail older adults on real-life IMU data,” Healthcare **9**(2), 169 (2021).
5. E. Barry, R. Galvin, C. Keogh, F. Horgan, and T. Fahey, “Is the Timed Up and Go test a useful predictor of risk of falls in community dwelling older adults: A systematic review and meta-analysis,” BMC Geriatrics **14**(1), 14 (2014).
6. H. S. Fang, J. Li, H. Tang, C. Xu, H. Zhu, and Y. Xiu, “AlphaPose: Whole-body regional multi-person pose estimation and tracking in real-time,” (2022).
7. J. Li, C. Wang, H. Zhu, Y. Mao, H. S. Fang, and C. Lu, “CrowdPose: Efficient crowded scenes pose estimation and a new benchmark,” (2018).
8. J. Li, Z. Wang, Z. Wang, S. Qiu, D. Peng, and K. Zhang, “A comprehensive evaluation method for frailty based on semi-supervised learning and transfer-learning,” Information Fusion **102**, 102504 (2024).
9. N. Mohammed, M. Mouhajir, and Y. Sedjari, “High performance computing applied to logistic regression: A CPU and GPU implementation comparison,” arXiv:2308.10037 (2023).
10. Y. Shan, L. Hui, H. Zheng, Y. Qiaoling, and W. Yali, “Research on imbalanced classification problem based on optimal random forest algorithm,” Proceedings of the 2024 International Conference on Machine Learning and Intelligent Computing, PMLR **245**, 383–392 (2024).
11. M. Liu, L. Zhang, and Y. Zhao, “A multilayer perceptron-based model applied to histopathology image classification of lung adenocarcinoma subtypes,” Frontiers in Oncology **13**, 1172234 (2023).
12. A. M. Elshewey, M. Y. Shams, and N. El-Rashidy, “Bayesian optimization with support vector machine model for Parkinson disease classification,” Sensors **23**(4), 2085 (2023).
13. M. Akhtar, M. Tanveer, and M. Arshad, “GL-TSVM: A robust and smooth twin support vector machine with guardian loss function,” arXiv:2408.16336 (2024).
14. V.W.S. Tan, W.X. Ooi, Y.F. Chan, C. Tee, and M.K.O. Goh, “Vision-Based Gait Analysis for Neurodegenerative Disorders Detection,” Journal of Informatics and Web Engineering **3**(1), 136–154 (2024)