ROYAL SOCIETY OPEN SCIENCE

royalsocietypublishing.org/journal/rsos

Research

Cite this article: Elbatanouny H, Hussain A, Al-Shabi M, Khan W, Mahmoud S. 2026 Personalized prediction system for early prediction of freezing of gait in Parkinson's disease using explainable AI. *R. Soc. Open Sci.* **13**: 250818.
https://doi.org/10.1098/rsos.250818

Received: 3 June 2025
Accepted: 22 December 2025

Subject Category:
Computer science and artificial intelligence

Subject Areas:
artificial intelligence, pattern recognition, computer vision

Keywords:
freezing of gait, Parkinson's disease, machine learning, explainable AI, personalized Parkinson's disease

Author for correspondence:
Wasiq Khan
e-mail: W.Khan@ljmu.ac.uk

Personalized prediction system for early prediction of freezing of gait in Parkinson's disease using explainable AI

Hagar Elbatanouny[1], Abir Hussain[1], Mohammad Al-Shabi[2], Wasiq Khan[3] and Soliman Mahmoud[4]

[1]Department of Electrical Engineering, and [2]Department of Mechanical and Nuclear Engineering, University of Sharjah, Sharjah, United Arab Emirates
[3]Department of Computer Science, Liverpool John Moores University, Liverpool, UK
[4]University of Sharjah, Sharjah, United Arab Emirates

WK, 0000-0002-7511-3873

Freezing of gait (FOG) is a common symptom of Parkinson's disease (PD), characterized by sudden and temporary episodes of immobility, often resulting in falls and reduced quality of life. Early and accurate prediction of FOG can greatly improve patient outcomes by allowing timely intervention and tailored treatment strategies. This article presents a personalized system for the early prediction of gait freezing in PD patients using explainable artificial intelligence (XAI). The proposed system follows a subject-dependent approach, training models specifically for individual patients to enhance prediction accuracy. It utilized multiple explainability techniques to ensure transparency in the decision-making process, achieving an average accuracy of 98.67 ± 0.75% using the random forest (RF) model and a latency of 75.0 ± 31.1 ms across six patients. Feature importance analysis, including SHapley Additive exPlanations (SHAP) and local interpretable model-agnostic explanation (LIME) plots, revealed significant influences of features like the maximum value of the x-axis, the maximum of the accelerometer signal from the x-axis and the standard deviation of the gyroscope signal from the y-axis in classifying gait states.

1. Introduction

Parkinson's disease (PD) is a progressive neurological condition that significantly affects movement control. Tremors, stiffness, bradykinesia and postural instability are some of the motor and non-motor symptoms associated with this disorder [1,2]. Freezing of gait (FOG) is a symptom in which patients suffer a brief, involuntary inability to move their feet. This freezing frequently

THE ROYAL SOCIETY PUBLISHING

results in falls and a significant decrease in quality of life [3,4].

FOG is seen as a critical marker of advanced PD, and it can appear at any stage of the disease [5]. Its unpredictability makes it difficult to manage successfully, needing an early and reliable prediction system to assist clinicians in tailoring treatment options for individual patients [6]. Currently, drugs such as dopamine agonists and levodopa are used to treat PD symptoms, although they frequently fail to address FOG specifically. Furthermore, pharmaceutical effectiveness varies greatly among patients, necessitating tailored prediction and intervention tactics [7–10].

Lately, machine learning (ML) techniques have developed as potential tools for detecting and predicting FOG [11,12]. By exploiting enormous amounts of data, ML models can discover trends and provide insights that human observers might not recognize [8,13–16]. Nevertheless, the difficulty is in developing models that are both accurate and interpretable, particularly in medical applications where it is essential to understand the logic underlying predictions [17]. By making decision-making processes transparent, explainable artificial intelligence (XAI) provides clinicians the confidence to believe and act upon AI systems' predictions [18].

This article presents a personalized prediction system for the early prediction of FOG in PD using XAI. Our approach uses a subject-dependent methodology, where models are tailored specifically to six individual patients from two different datasets, incorporating the unique characteristics and conditions of these subjects to ensure more accurate and relevant predictions. The novelty of our approach lies in several key aspects:

- Explainability: The inclusion of explainability techniques ensures that clinicians can understand the model's reasoning, which supports informed decision-making and improves patient care.
- Reduced latency: The latency of predictions is significantly reduced, ensuring faster results.
- Enhanced performance: The approach boosts prediction performance, providing more accurate and reliable results for personalized FOG prediction.

The rest of the article is structured as follows. Section 2 provides a discussion of related work, highlighting existing approaches for FOG prediction and the challenges of integrating explainability. Section 3 details the methodology, including the datasets and patients used, the pre-processing techniques and the subject-dependent approach used to develop the prediction system. Section 4 presents the results and discusses the performance of the models, followed by an analysis of its latency and interpretability. Finally, §5 concludes the article and outlines future directions for improving the system's scalability and clinical applicability.

2. Related work

Zhang *et al.* [19] used impaired gait patterns to try to enhance FOG prediction in patients with PD. An accelerometer placed on the lower back was used to gather data from 12 patients with PD. After these patients completed a series of walking exercises that caused FOG, the data were processed to identify compromised gait characteristics such as heart rate, cadence and variability. They compared various feature sets, such as step-based impaired gait features, traditional FOG detection features, as well as customized pre-FOG labelling, using AdaBoost with pruned C4.5 trees. The top model achieved an accuracy of 82.7% in the patient-dependent test and 77.9% in the patient-independent test with a latency of 0.93 s.

In order to determine the best settings for the model, Sun *et al.* [20] focused on predicting FOG in patients with PD using a combination of deep learning techniques and manually selected features. The Daphnet dataset [21], which consists of gait recordings from 10 patients with PD, was used for evaluation. The suggested method, based on ResNeXt, achieved an accuracy of 95.40% with a Macro-F1 score of 0.89 and a kappa coefficient of 0.87. Their study showed that combining handcrafted features with deep learning features enhances prediction performance when compared with using either feature set alone.

Xia *et al.* [22] proposed a method for learning latent gait representations through self-supervised contrastive learning. Their approach utilized various augmented views of unlabeled acceleration data segments to develop invariant gait representations without requiring labelled data. For FOG prediction, they categorized the data into three classes: 'pre-FOG,' 'Walk' and 'FOG'. The effectiveness of their framework was validated using the DAPHnet dataset [21] and a multi-modal FOG dataset from Beijing Xuanwu Hospital in China (BXHC) [23]. Their model demonstrated a sensitivity of 84.61%, a specificity of 94.74% and an F1 score of 86.19% for the pre-FOG class on the DAPHnet dataset. Similarly, when

tested on the BXHC dataset, their model achieved a sensitivity of 80.00%, a specificity of 93.08% and an F1 score of 81.07%.

Elbatanouny *et al.* [17] provided a comprehensive review of the current state of ML techniques for detecting and predicting FOG. Their analysis explores the integration of wearable sensors with ML models, focusing on key components such as data sources, feature extraction methods, pre-processing strategies and evaluation metrics. It also addresses the use of cueing devices and compares ML-based approaches with traditional, non-ML methods. The authors highlight several important directions.

Filtjens *et al.* [24] proposed an interpretable deep learning architecture for detecting movements that precede FOG events. Their system utilizes a convolutional neural network to model the changes in movement prior to a FOG episode, using three-dimensional (3D) kinematic joint trajectories from the ankle, hip and knee joints. The system achieved an accuracy of 86.8%. In addition, the study incorporated layer-wise relevance propagation (LRP), a popular XAI technique, to identify the critical factors that the convolutional neural network considers when modelling the pathophysiology of FOG.

Priyadharshini *et al.* [25] presented a framework for diagnosing PD using ML and XAI methods. They utilized a dataset of T2-weighted 3D magnetic resonance imaging scans from the Parkinson's progression markers initiative, which includes data from 500 participants, comprising patients with PD, those at risk of developing PD and healthy controls. A variety of ML algorithms, such as GB, random forest (RF) and K-nearest neighbours (KNN), were tested, with GB achieving the highest accuracy of 96.8%. They also used several explainability techniques, including SHapley Additive exPlanations (SHAP) and local interpretable model-agnostic explanations (LIMEs), to enhance the interpretability of the predictions.

Recent XAI technologies, notably deep learning frameworks, have been investigated in the context of gait analysis and freezing in PD. Convolutional and recurrent neural networks have been explained using techniques such as LRP and integrated gradients, which highlight the most influential temporal segments or sensor channels contributing to FOG-related predictions [24,26]. Similarly, attention-based techniques have been implemented in sequence models to highlight important gait moments related to freeze onset [27]. While these methods enable fine-grained interpretability for deep architectures, they are frequently computationally demanding and necessitate huge amounts of training data, which may restrict their application in subject-dependent or real-time wearable-based FOG prediction systems.

There is a significant gap in the literature regarding the explainability of ML models used for FOG prediction in PD. Many studies focus primarily on improving model accuracy and performance without adequately addressing the transparency of model predictions [17], which is essential for clinical adoption. Furthermore, several papers overlook the importance of calculating prediction latency, a critical factor for real-time clinical decision-making. In addition, while some research has embraced the subject-dependent approach, there is a notable distinction in how it is implemented. Some studies train models on data from all other patients and test on data from the target patient [28,29]. This approach may not fully capture the patient-specific variations in FOG patterns, as each patient's gait behaviour can differ. In contrast, other studies train and test models on different data windows from the same patient, aiming to learn the unique gait patterns of the individual [19,29]; hence, we decided to follow this approach. Personalized solutions that account for factors such as age, gender and disease duration are critical, as these variables can significantly affect a patient's gait and FOG manifestations, emphasizing the need for subject-specific models that cater to these individual differences.

3. Methodology

This section outlines the methodological framework used to develop and evaluate the models used for predicting FOG. The approach involves the use of the DaphNet dataset. Detailed descriptions of the dataset and the subject used, along with the pre-processing steps, feature extraction and ML models applied, are provided. The section also explains the experimental set-up, including performance metrics, and discusses the tools used for model interpretability.

3.1. System overview

The system overview provided in figure 1 illustrates the framework for predicting FOG in patients with PD. The analysis focuses on subject-dependent prediction for individual patients. The data undergo pre-processing and feature extraction steps, such as filtering, calculating magnitude, windowing and extracting features, which serve as the foundation for model training.

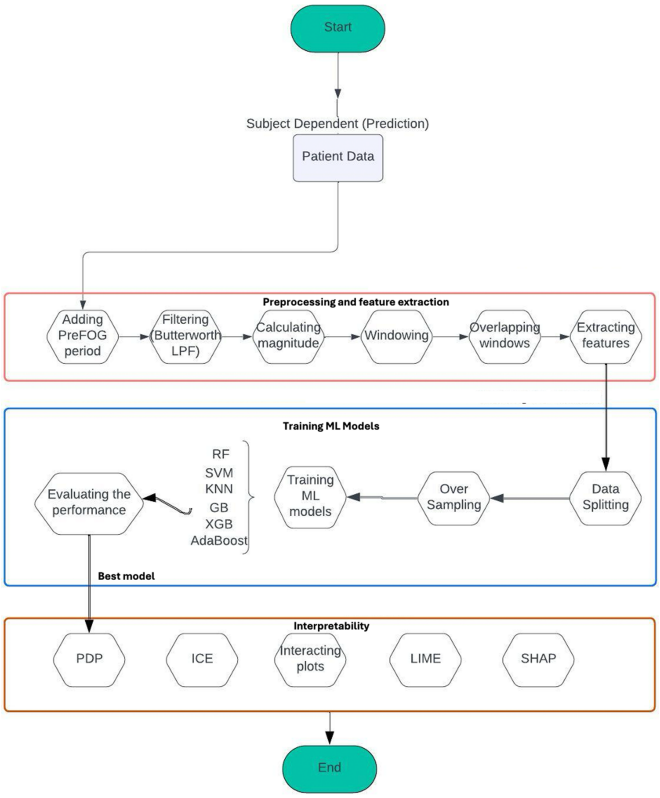


Figure 1. System overview.

The system also incorporates a variety of ML models. Once the best model(s) are identified, interpretability tools like LIME, SHAP, interaction plots, individual conditional expectation (ICE) and partial dependence plots (PDPs) are applied to provide insights into the model's decision-making processes. Detailed descriptions of each step, including the dataset, pre-processing techniques and interpretability methods, will be provided in detail.

3.2. Datasets

In this study, two publicly available datasets were utilized to develop and evaluate a personalized system for the early prediction of FOG in patients with PD. These datasets, DaphNet and inertial measurement unit (IMU)-PD, were selected owing to their rich temporal sensor recordings, diversity in patient demographics and detailed annotations of FOG episodes. Each dataset provides unique experimental set-ups and sensor configurations, allowing for a comparison of model performance under varying conditions. The following subsections present an overview of both datasets, including data collection protocols, sensor placements and participant characteristics. A separate subsection details the specific patients selected for this study and the reason behind their inclusion.

3.2.1. DaphNet dataset

The DaphNet dataset, created by Bachlin *et al.* [21], includes movement data from 10 patients with PD (seven males and three females), with an average age of 66.5 ± 4.8 years and varying disease severity (Hoehn and Yahr (H&Y) score: 2.6 ± 0.65). The patients, all diagnosed with a history of FOG and capable of walking unassisted, were tested during the 'OFF' phase of their medication cycle to ensure observable motor symptoms. The data collection consisted of two sessions for each patient: one without feedback cues and one with feedback cues, during which the patients performed three distinct walking tasks. The walking tasks included straight-line walking, random walking and simulated activities of daily living (ADL), which are visually represented in figure 2. Data were collected using three sensors placed on the patients' ankle, leg and lower back, with accelerometer data synchronized with video recordings.

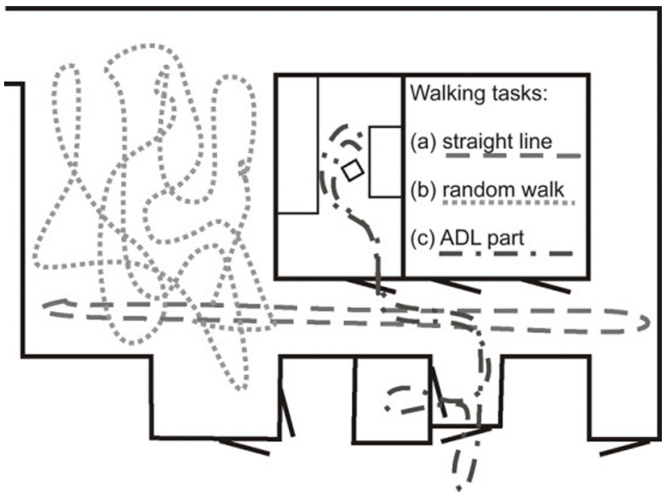


Figure 2. A representation of the subjects' path while collecting the data [21].

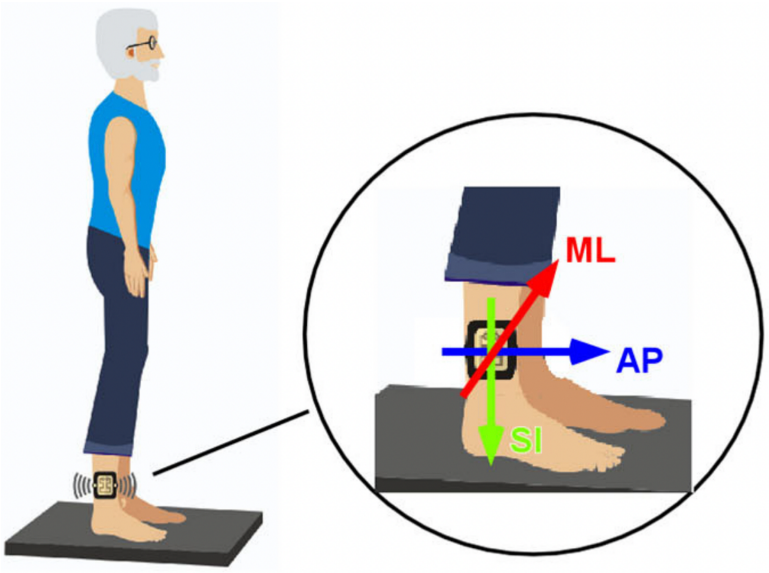


Figure 3. Wearable device placement on the participant's leg [30].

3.2.2. IMU dataset

The IMU dataset, collected by De Souza *et al.* [30], includes movement data from 35 patients with PD (16 females and 19 males) aged 44 to 84 years. The patients, all in the ON-medication state, were evaluated across three experimental sessions, each one month apart, with stable anti-Parkinsonian medication dosages. The patients performed a turning task, alternating 360° turns to the right and left for 2 min, while wearing an IMU sensor attached to the shank of their most affected leg. The IMU sensor, consisting of an accelerometer and gyroscope, measured triaxial linear acceleration and angular velocity, with data recorded at 128 Hz. Sensor placement is shown in figure 3, where the sensor is positioned on the participant's leg, with specific orientations for the axes to capture movement data during the task. After each session, accelerometer and gyroscope data were aligned with video recordings to identify FOG events.

3.2.3. Patients used in the study

Three patients from each dataset were chosen, making a total of six individuals, guaranteeing a range of ages, genders, disease durations and clinical stages. Figure 4 illustrates the selected patients, including their dataset origin, gender and identifier. Weight, height and body mass index (BMI) were taken into

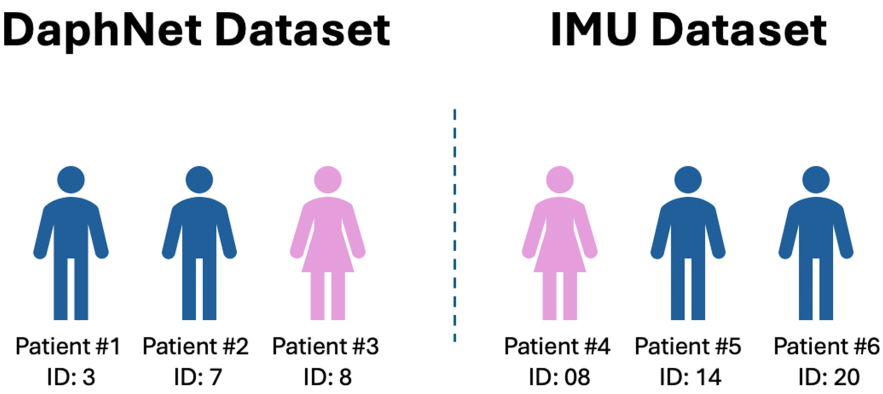


Figure 4. Overview of the selected patients.

consideration to contextualize motor function and gait variability, and the H&Y scale, which ranges from 1 (moderate) to 5 (severe impairment), was used to evaluate the severity of the disease.

Patients from the DaphNet dataset:

- Patient 1 (ID 3): a 59-year-old male with a disease duration of 30 years, the longest among DaphNet patients. His H&Y score was 2.5, which indicated modest bilateral symptoms and some postural instability. Despite having PD for the longest time, his clinical stage remained moderate, indicating that the condition progressed gradually. His findings are useful for studying long-term gait adaptations.
- Patient 2 (ID 7): a 66-year-old male had the lowest disease duration in the DaphNet dataset, at 2 years, and an H&Y score of 2.5, which was the same as Patient 1. Despite his brief disease history, he was already experiencing bilateral motor symptoms and balance difficulties. His story is important for understanding the early stages of motor impairment and how quickly symptoms might develop.
- Patient 3 (ID 8): a 68-year-old female. Her condition had been present for 18 years, and she had the highest H&Y score in the DaphNet group (4.0), suggesting severe motor dysfunction. Her participation sheds light on the advanced stages of PD, including FOG and postural instability.

Patients from the IMU Dataset:

- Patient 4 (ID 08): an 84-year-old female, the oldest of the IMU patients, with a disease duration of 18 years and an H&Y score of 4.0, indicating advanced PD. Given her height of 1.70 m and weight of 67 kg, her BMI of 23.2 is within the normal range. She is a notable case for evaluating FOG in older people with protracted progression because of her senior age and lengthy medical history.
- Patient 5 (ID 14): a 69-year-old man with the shortest disease duration of the IMU patients, 1.5 years, but had a constant H&Y score of 3.0, indicating moderate disease with obvious postural instability. He had a BMI of 25.4, which was marginally above normal, weighed 65 kg, and was 1.60 m tall.
- Patient 6 (ID 20): a 44-year-old male, the youngest among the IMU patients, with a 5 year disease history and an H&Y score of 3.0, the same as Patient 5. His height and weight were 1.66 m and 69 kg, respectively, giving him a slightly overweight BMI of 25.1. His profile makes it possible to examine younger people with mid-stage PD and evaluate the relationship between age and body composition and changes in gait.

3.3. Pre-processing and feature extraction

To improve the early prediction of FOG, a 'pre-FOG' period was introduced as an additional event class, representing the transition between normal walking and the onset of FOG. This interval is crucial as it provides early indicators of FOG, aiding in the enhancement of predictive tasks. Figure 5 illustrates an annotated accelerometer signal, where different event states are colour-coded: walking (green), pre-FOG (blue) and FOG (purple).

Signal filtering is a crucial process used to eliminate noise from the signals, which can obscure important data and lead to false interpretations. Previous research has indicated that the significant frequency

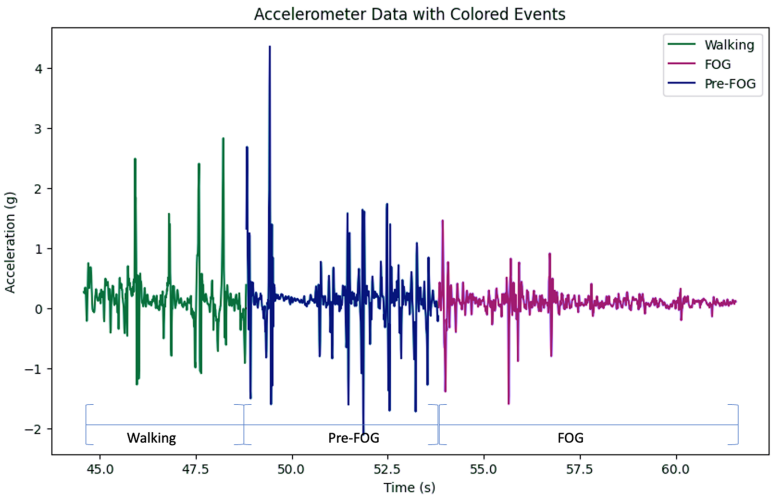


Figure 5. Accelerometer signal with colour annotations representing different states: FOG (purple), pre-FOG (blue) and walking (green).

components of human gait are typically found below 20 Hz, with the critical frequency range for FOG detection typically falling between 3 and 8 Hz. In line with this, many studies on FOG detection and prediction using accelerometer and gyroscope signals have employed low-pass Butterworth filters with cutoff frequencies of 10 and 15 Hz [31,32].

For this study, a third-order low-pass Butterworth filter with a cutoff frequency of 20 Hz was applied to reduce noise in the signals. This filter was selected to reduce high-frequency noise. After filtering, both the raw and filtered signals were visualized to ensure that the process provided a smoother signal without introducing significant distortion.

An extensive analysis was conducted under the subject-independent setting using various window sizes to determine the most effective configuration for capturing the relevant gait features. Based on this analysis, and as supported by [33], a 5 s window size was selected for this subject-dependent analysis. This choice was made to ensure alignment with the 5 s pre-FOG period, enabling the model to effectively learn the subtle changes in gait that typically precede a FOG episode. To enhance the granularity of data and improve temporal resolution, overlapping windows were employed with overlap percentages of 10%, 20% and 50%. For example, with a 10% overlap, each new window starts 10% after the previous one, resulting in a higher number of overlapping segments. This technique increases the continuity of the signal representation and provides a more detailed temporal mapping, making the analysis more sensitive to subtle and transient changes.

The label assigned to each window is determined by the majority label of the instances within that window. In cases where there is a tie between multiple labels, the first label encountered is considered the dominant one.

During data inspection, no missing or null values were detected, ensuring the integrity of the datasets. To enrich the analysis, additional features were introduced, specifically the magnitudes of accelerometer and gyroscope readings. This inclusion of magnitude data is especially useful for FOG prediction, as the intensity of movement often plays a more significant role than its direction [34]. The magnitude is computed using equation (3.1):

$$\text{magnitude} = \sqrt{x^2 + y^2 + z^2}, \tag{3.1}$$

where x, y and z are the sensor values in the three directions.

The features are extracted from the windows and include various descriptors such as zero-crossing rate (ZCR), median, interquartile range (IQR), peak-to-peak amplitude, minimum and maximum values and several statistical metrics like mean, standard deviation, variance, inverse coefficient of variation (ICV), skewness and kurtosis. Frequency-domain features were also extracted, including spectral energy, dominant frequency, peak power spectral density (PSD) and peak frequency within certain frequency bands.

In addition, key features from the literature were included, such as the locomotor band (0.5–3 Hz), the freezing band (3–8 Hz), and the freezing index (FI), as used by Moore *et al.* [35].

It is important to recognize that not all ML algorithms require feature normalization or standardization. Certain methods, such as decision trees and RFs, are largely unaffected by the scale of the input

features since they do not rely on distance-based calculations or gradient-based optimization. However, for algorithms that depend heavily on these principles, such as KNN and SVM, standardization becomes essential for optimal performance and reliable outcomes [36].

In this study, standardization was applied after the feature extraction process. This decision was made to ensure that all features have equal scale, which is crucial for models that utilize distance metrics or gradient-based optimization methods. Without standardization, features with larger ranges or values may influence the model, leading to skewed results where certain features dominate the learning process.

To implement this, the StandardScaler from the scikit-learn library in Python was used. This approach standardizes the features by eliminating the mean and scaling to unit variance.

3.4. Machine learning models

Different ML models were applied, chosen based on their specific strengths and their effectiveness as demonstrated in previous research [17]. The prediction model set included RF, KNN, support vector machine (SVM), adaptive boosting (AdaBoost), eXtreme gradient boosting (XGBoost) and gradient boosting (GBM).

These ML models were also chosen to reflect a wide range of learning paradigms appropriate for wearable sensor data and real-time clinical applications. Tree-based models like RF and GBM were chosen for their noise resistance, ability to represent nonlinear relationships and intrinsic feature importance estimate. Distance-based models, such as KNN, were examined for their simplicity and efficacy in subject-dependent scenarios, whereas SVM was included for its high-dimensional performance.

The data of each patient were divided into three parts: 80% for training, 10% for validation and 10% for testing. This split ensured that the models were trained and evaluated on distinct sets of data, preventing overfitting and providing an accurate assessment of their generalization capabilities.

To ensure robustness, fivefold cross-validation was applied during prediction. This cross-validation process helped in refining the models and assessing their stability across different data subsets.

To address class imbalance and enhance the model evaluation process, the synthetic minority oversampling technique (SMOTE) [37] was used to oversample minority class and downsample majority classes within the training dataset. Following the application of SMOTE, the performance of the models was evaluated using validation and testing sets composed of original, unaltered data. This strategy ensured that the evaluation metrics accurately reflected the models' predictive capabilities in realistic scenarios, as they would encounter in practical applications.

Optuna [38], a software framework for hyperparameter optimization, was utilized to fine-tune the hyperparameters of the best-performing model. The ranges for the hyperparameters were manually chosen and randomly tested to explore various combinations, allowing for a comprehensive evaluation of the model's performance across different settings.

Latency is the amount of time that passes between the time a prediction model receives input data and the time it produces an output or prediction when it comes to predicting FOG and pre-FOG in people with PD [39]. Reducing latency is essential since it has a big effect on the model's capacity to notify patients in a timely manner. FOG prediction's main goal is to alert people as soon as possible so that they may take precautions or get ready for an upcoming FOG episode [17].

Performance evaluation metrics play a crucial role in determining the reliability and clinical applicability of ML models. While several metrics, such as accuracy, precision, recall and the F1 score, offer valuable insights into model performance, precision and recall (sensitivity) become particularly significant. Since FOG is a serious medical symptom, it is imperative to minimize false negatives and false positives, ensuring that most of the identified FOG episodes are indeed real, to avoid unnecessary patient anxiety and ensure timely intervention.

3.5. Explainable techniques

Explainability plays a vital role in ML models, as understanding the rationale behind model predictions builds trust and provides transparency, which is critical for clinical adoption [3]. In this study, various explainability techniques were applied to interpret the model's decision-making process. PDPs were first used to visualize the relationship between individual features and the predicted outcome, showing how the predictions change when varying a single feature while keeping others constant. ICE lines were then used to provide a more granular, instance-level view of feature effects, revealing how feature values affect predictions for individual data points. Two-dimensional (2D) nteraction plots were used to assess

Table 1. Number of walking, FOG and pre-FOG events recorded for each patient.

patient ID	walking	FOG	pre-FOG
Patient 1	28	49	49
Patient 2	22	24	24
Patient 3	15	14	14
Patient 4	4	7	6
Patient 5	9	10	10
Patient 6	4	3	3

how combinations of features interact and influence predictions, providing deeper insight into the interdependencies between multiple features. LIME was applied next, generating local surrogate models to approximate the behaviour of the complex model for individual predictions, thus explaining specific results. Finally, SHAP was applied both locally and globally. The global SHAP approach provided an overall understanding of feature importance, quantifying how each feature contributed to the model's predictions across the entire dataset. The local SHAP method, on the other hand, offered insights into individual predictions, explaining the specific contributions of features for each data point [3].

While recent XAI approaches such as LRP, integrated gradients and attention-based explanations offer powerful interpretability for deep learning models, they were not used in this study because the emphasis was on classical ML algorithms optimized for low-latency and real-time deployment. Model-agnostic strategies, such as SHAP and LIME, are better suited to tree-based and ensemble models since they allow for both global and local explanations with minimal computational overhead. This makes them ideal for personalized, subject-dependent prediction systems that require transparency, consistency among patients, and real-time responsiveness for clinical application. Additional techniques such as PDP, ICE and interaction plots were employed to complement SHAP and LIME by visualizing marginal and joint feature effects.

4. Results and discussion

In this section, the subject-dependent prediction results for patients are presented from both datasets. The ML results offer a quantitative assessment of predictive accuracy and model reliability, followed by a detailed explanation of the model's reasoning using interpretability techniques.

Six patients were chosen for a detailed evaluation to assess the customized prediction approach. Each patient experienced a different number of walking, FOG and pre-FOG occurrences, with variable durations. For consistency in sensor placement and to facilitate fair comparison between datasets, the ankle sensor was selected from the DaphNet dataset, as prior analysis indicated that it provides the most discriminative patterns for identifying FOG and pre-FOG episodes. This choice is also consistent with the IMU-PD dataset, in which sensors were placed on the lower leg. A summary of the recorded events for each patient is provided in table 1.

Overlapping windows were applied to this data, with table 2 detailing the class distribution across each overlap percentage.

Table 3 summarizes the classification performance of the best-performing model per patient, using SMOTE oversampling. For five of the six patients, the RF model outperformed the others on most parameters. Patient 4 was an exception, since the GB model outperformed the RF model by a little margin.

High precision, recall and F1 scores were consistently found throughout the walking, pre-FOG and FOG classes, demonstrating the model's ability to discern across gait modes. Notably, pre-FOG F1 scores ranged from 0.93 to 1.00, indicating that the model could predict freezing occurrences early on. The accuracy scores were continuously high, ranging from 97.14% to 99.41%, with an average of 98.67 ± 0.75%.

Model latency was used to assess the system's responsiveness for real-time use. The latency values varied from 46.6 to 134.1 ms, with an average of 75.0 ± 31.1 ms. Given that the prediction is provided for a 5 s window that ends 2.5 s before the actual onset of freezing, the effective lead time for action is

Table 2. The number of windows for each overlapping percentage, where W, walking; P, pre-FOG and F, FOG.

patient	10%			20%			50%		
	W	P	F	W	P	F	W	P	F
Patient 1	7793	1004	1469	3897	501	735	1559	200	294
Patient 2	7230	634	354	3614	318	177	1444	128	72
Patient 3	2533	360	1024	1266	180	513	505	74	205
Patient 4	91	274	813	45	137	407	18	55	163
Patient 5	393	583	202	196	292	101	79	116	41
Patient 6	986	173	19	493	86	10	198	34	4

Table 3. Evaluating the performance of the RF model with different feature selection techniques.

		DaphNet dataset			IMU dataset		
		Patient 1	Patient 2	Patient 3	Patient 4	Patient 5	Patient 6
walking	precision	1.00	1.00	0.98	1.00	1.00	0.99
	recall	0.99	1.00	0.98	1.00	0.96	1.00
	F1 score	1.00	1.00	0.98	1.00	0.98	1.00
pre-FOG	precision	0.96	0.99	0.93	1.00	0.97	1.00
	recall	0.98	0.96	0.93	0.94	1.00	1.00
	F1 score	0.97	0.97	0.93	0.97	0.99	1.00
FOG	precision	0.96	0.94	0.96	0.98	1.00	1.00
	recall	0.98	1.00	0.96	1.00	1.00	0.67
	F1 score	0.97	0.97	0.96	0.99	1.00	0.80
accuracy		98.87%	99.41%	97.14%	98.64%	98.63%	99.32%
latency		54.1 ± 0.37 ms	81.3 ± 0.41 ms	81.3 ± 0.72 ms	46.6 ± 0.11 ms	134.1 ± 0.21 ms	52.8 ± 0.59 ms

estimated using equation (4.1):

$$\text{lead time} = 2.5\,\text{s} - \text{latency}. \tag{4.1}$$

The average lead time for patients is 2.425 ± 0.031 s, allowing for opportune alerts or actions before the freezing episode.

RF-based feature selection was used to assess the effect of lowering the amount of input features. Figures 6 and 7 show that the RF model was evaluated on gradually smaller subsets (e.g. top 10, 30, 50, 100 and 120 features), and the results were compared with those obtained using the whole feature set.

The findings indicate that the complete feature set constantly provided the best or near-highest accuracy across all patients. When using the whole feature set, accuracy in the IMU dataset (Patients 4–6) reached 99.32%. Similarly, with the DaphNet dataset (Patients 1–3), the entire feature set performed best, with Patient 2 achieving 99.41% accuracy. Nonetheless, in other situations, decreased feature sets (e.g. top 30 or 50 features) produced comparable results, particularly in the IMU cohort, where the accuracy reduction was minimal.

While RF-select did not regularly outperform the whole feature set in terms of accuracy, it does provide a useful trade-off by allowing for more computationally efficient models. Smaller input sizes can considerably reduce inference time, resulting in lower latency and a longer lead time before FOG onset when used in real-time systems. This economy is especially useful for wearable or embedded applications, where computing power and time are limited. Although the whole feature set provides the best accuracy, RF-select with limited subsets may be chosen in future implementations when low latency is crucial.

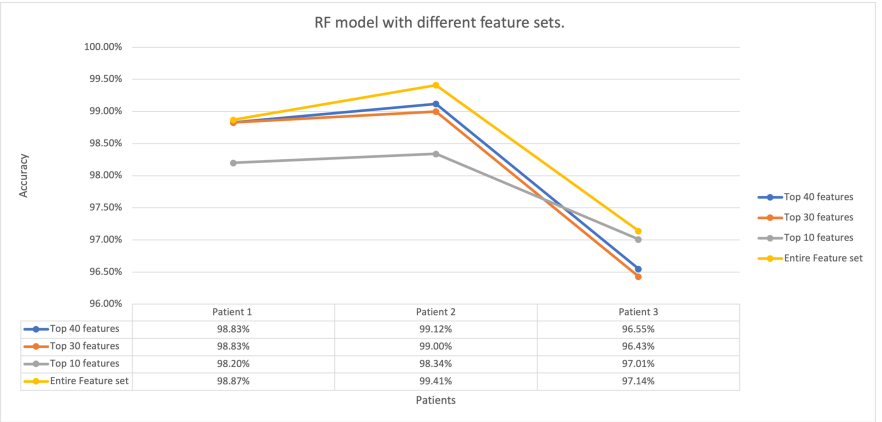


Figure 6. Accuracy of the RF model for Patients 1–3 (DaphNet) using different RF-select feature subsets.

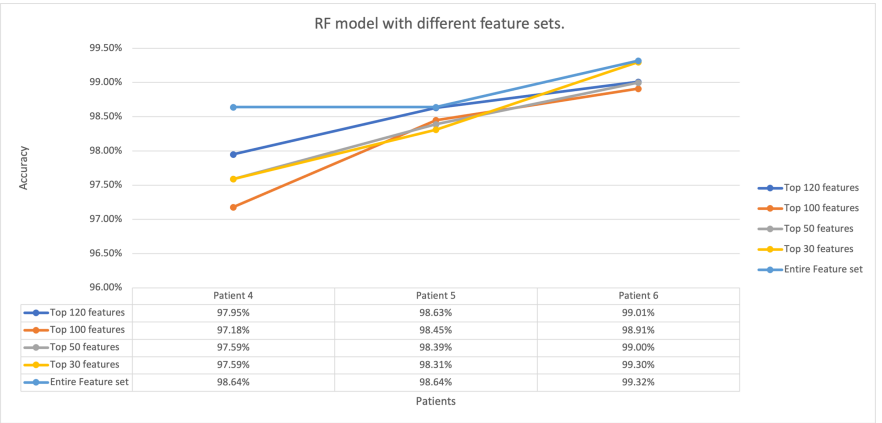


Figure 7. Accuracy of the RF model for Patients 4–6 (IMU) using different RF-select feature subsets.

Table 4 lists the top 10 features chosen by RF-based feature selection for each of the six patients. The findings show significant variability in the most discriminating features between people, reflecting the uniqueness of PD symptoms and patient profiles.

Statistical time-domain variables such as maximum, minimum, variance and peak-to-peak values are most common across all patients, emphasizing their importance in capturing gait dynamics and freezing transitions.

Certain features appear to be particularly individualized. For example, spectral energy and max PSD are particularly critical for Patient 4 (an elderly participant with severe illness), possibly representing more irregular or powerful movement bursts while turning. Younger patients, such as Patient 6, displayed various acceleration-based features from both the x and z axes, including peak-to-peak and IQR values, which could indicate faster or more varied steps.

Table 5 lists the best-performing hyperparameters for each patient based on Optuna optimization for the RF model, along with their related test set accuracy. The table shows the efficacy of patient-specific tuning, with each model independently adjusted to capture distinct gait patterns and freezing properties.

The validation results suggest that false positives and false negatives are modest across all patients, with total error rates ranging from 0.58% to 2.45%, as illustrated in figure 8. The pre-FOG class showed the most misclassifications, including both false positives (e.g. walking projected as pre-FOG) and false negatives (e.g. pre-FOG predicted as walking or FOG). This reflects the transitional and less distinguished characteristics of pre-FOG windows.

False negatives in FOG prediction were rare, indicating that actual FOG occurrences were consistently identified. Similarly, false positives for FOG were rare, decreasing the possibility of unneeded intervention. The errors that did occur were often close in class (e.g. pre-FOG versus FOG), indicating that the model had strong applicability even in borderline circumstances.

Table 4. Top 10 Features using RF-feature selection for the patients.

	Patient 1	Patient 2	Patient 3	Patient 4	Patient 5	Patient 6
feature #1	peak to peak (mag)	kurtosis (mag)	maximum (x)	variance (Gy)	peak to peak (Axe)	maximum (Gx)
feature #2	std dev (mag)	maximum (z)	peak frequency (y)	spectral energy (Az)	median (Axe)	minimum (Gmag)
feature #3	icv (mag)	skewness (mag)	minimum (y)	std dev (Gy)	minimum (Amag)	iqr (Ay)
feature #4	peak frequency (y)	dominant frequency (y)	dominant frequency (x)	max psd (Gmag)	minimum (Gz)	zcr (Axe)
feature #5	variance (mag)	icv (z)	median (x)	icv (Axe)	maximum (Gy)	icv (Gmag)
feature #6	maximum (y)	peak frequency (z)	minimum (x)	minimum (Gy)	minimum (Az)	iqr (Az)
feature #7	dominant frequency (y)	zcr (x)	maximum (y)	median (Az)	minimum (Gmag)	skewness (Axe)
feature #8	max psd (y)	variance (z)	median (mag)	zcr (Gy)	peak to peak (Amag)	iqr (Axe)
feature #9	maximum (mag)	dominant frequency (x)	max psd (x)	maximum (Gy)	minimum (Ay)	maximum (Az)
feature #10	minimum (y)	kurtosis (y)	mean (x)	mean (Axe)	maximum (Axe)	peak to peak (Gmag)

Table 5. Best parameters obtained by Optuna for each patient using the RF model.

parameter	Patient 1	Patient 2	Patient 3	Patient 4	Patient 5	Patient 6
*n*_estimators	263	151	151	106	243	58
max_depth	24	28	28	12	28	14
min_samples_split	15	4	4	10	7	13
min_samples_leaf	5	9	9	1	1	2
max_features	sqrt	sqrt	sqrt	log2	log2	log2
bootstrap	false	false	false	false	false	false
accuracy (%)	98.6760	99.4163	97.1429	97.2789	98.6395	99.3197

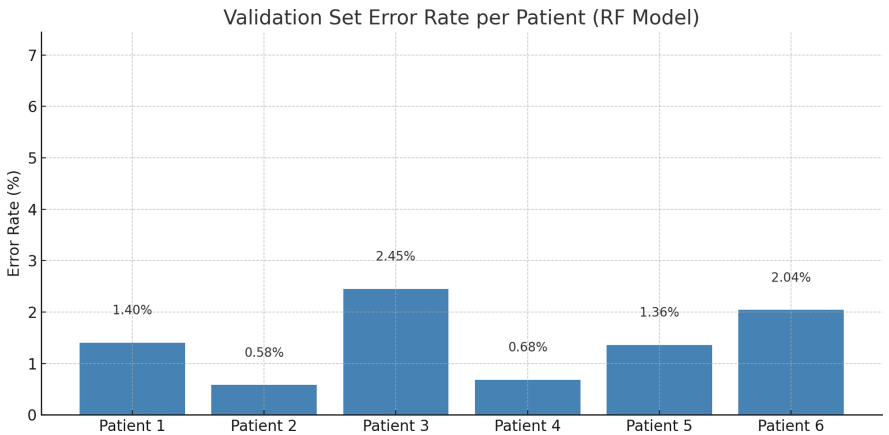


Figure 8. Error rate per patient.

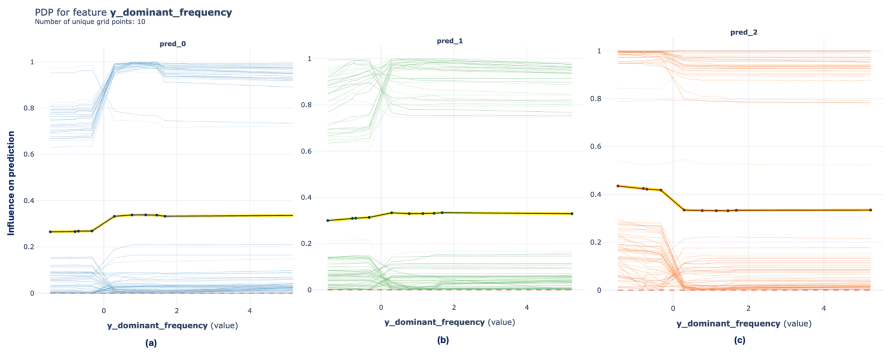


Figure 9. PDP plot with 100 ICE for feature 'ydominantfrequency' (Patient 2), where (a) is FOG class, (b) is pre-FOG class and (c) is walking class.

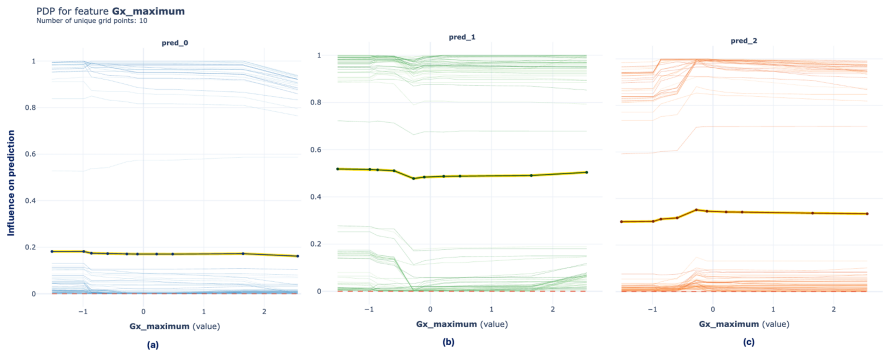


Figure 10. PDP plot with 100 ICE for feature 'Gx_maximum' (Patient 5), where (a) is FOG class, (b) is pre_FOG class and (c) is walking class.

4.1. Explainability results

To further understand the decision-making process of the personalized models and assure clinical interpretability, different explainability techniques were applied to the selected patients. These include global and local SHAP, LIME, ICE-lined PDPs and 2D PDP interaction graphs. The goal is to investigate the effect of individual characteristics on prediction outcomes, both globally and per-instance (window) basis.

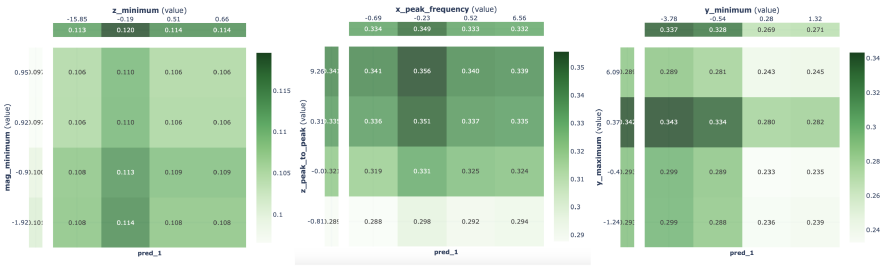
Figures 9 and 10 show the PDP plots for the features 'y_dominant_frequency' (Patient 2) and 'Gx _maximum' (Patient 5), which demonstrate how changes in these feature values affect the model's class probabilities of FOG, pre-FOG and walking.

In figure 9, the model shows a nonlinear response to changes in 'y_dominant_frequency'. For class 0 (figure 9a), the influence on the model prediction increases when the feature value is approximately zero, indicating a threshold-like behaviour. The pre-FOG class in figure 9b shows moderate sensitivity with slight variations, suggesting that changes in this feature's values have a minimal effect on the prediction since all values exert nearly similar effects. The ICE lines show consistent individual trends, with notable positive jumps around zero for the FOG class and similar negative changes for the walking class.

Figure 10 illustrates the marginal effect of 'Gx_maximum' on Patient 5. The plots in figure 10b,c show that the influence on walking and pre-FOG classes is most notable between −1 and 0. This effect is evident in the ICE lines, where negative jumps in the pre-FOG class and positive shifts in the walking class support the model's dependence on this interval to distinguish between transitional and steady walking states.

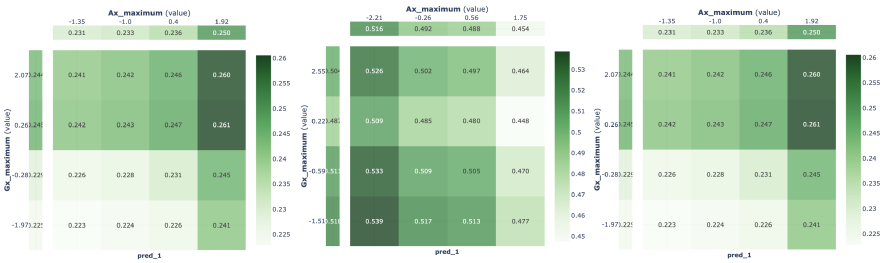
Figures 11 and 12 illustrate the 2D PDP interaction plots of selected feature pairs for the pre-FOG class in all patients. To capture nuanced relationships that may influence the model's decision boundaries, features were chosen either based on similarities across sensor axes—such as 'mag_minimum' and 'z_min', or from top-ranked features identified earlier.

Figure 11 shows the interaction effects for Patients 1–3. Figure 11a indicates that the interaction between 'z_minimum' and 'mag_minimum' for Patient 1 has a little increase in influence when 'mag _minimum' values are lower, but the overall effect remains moderate. For Patient 2, figure 11b shows that increasing 'z_peak_to_peak' has more effect on pre-FOG prediction than the 'x_peak_frequency'.



(a) Patient 1 (b) Patient 2 (c) Patient 3

Figure 11. 2D interacting plots (pre-FOG class) for different features in three cases: a) Patient 1, b) Patient 2, and c) Patient 3.



(a) Patient 4 (b) Patient 5 (c) Patient 6

Figure 12. 2D interacting plots (pre-FOG class) for the features 'Axmaximum' and 'Gxmaximum' in three cases: a) Patient 4, b) Patient 5, and c) Patient 6.

Table 6. Interaction ranges of high influence for feature pairs across pre-FOG, FOG and walking classes, where Feature A is 'Ax _maximum' and Feature B is 'Gx_maximum'.

	Patient 4		Patient 5		Patient 6	
	feature A	feature B	feature A	feature B	feature A	feature B
walking	low -135 - (-1.0)	all values _	all values _	low -1.51 - (-0.59)	low -135 - (-1.0)	all values _
pre-FOG	high 0.4 -1.92	high 0.26 - 2.07	low -2.21 - (-0.26)	all values _	high 0.4 - 1.92	high -0.28 - 2.07
FOG	low -135 - 0.4	low 1.97 - (-0.28)	high 0.56 - 1.75	high 0.22 - 2.55	low -135 - 0.4	low 1.97 - (-0.28)

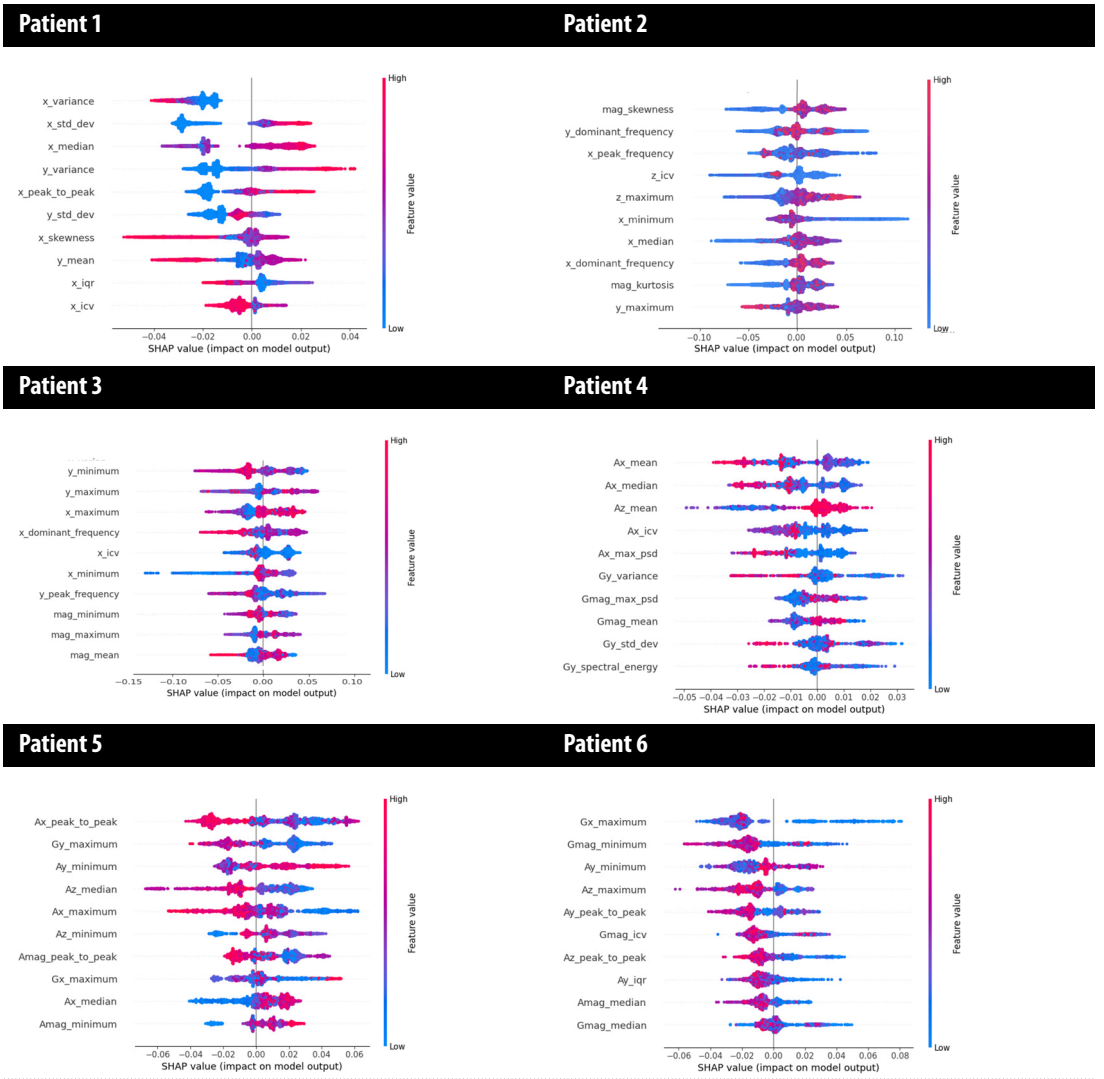
In figure 11c, Patient 3 demonstrates a zone of higher influence when 'y_minimum' is low and 'y _maximum' falls between approximately −0.4 and 6.09, suggesting this range is critical in shaping pre-FOG predictions.

Figure 12 and table 6 summarize the interaction patterns for Patients 4–6 using the pair 'Ax _maximum' and 'Gx_maximum'. As shown in the table, for Patient 4, higher values of both features are linked to greater influence on the pre-FOG class, while lower values correspond to FOG predictions. Patient 6 behaves similarly to Patient 4, confirming the consistency of these interactions between these two patients.

In contrast, Patient 5 shows a distinct pattern. High values of both 'Ax_maximum' and 'Gx_maximum' strongly influence the prediction of the FOG class, whereas low values of 'Ax_maximum' alone significantly increase the prediction probability for pre-FOG, regardless of the value of 'Gx_maximum', as shown in table 6. This interaction pattern is reflected in figure 12b, where the influence exceeds 0.5, indicating a particularly strong model response. Furthermore, this is supported by the PDP plot in figure 10b, where 'Gx_maximum' shows high influence on pre-FOG classification when values range between approximately -1.5 and -0.5.

The underlying biomechanical variations among patients may be reflected in these interaction trends. For instance, Patient 5's borderline overweight BMI and relatively recent diagnosis may have contributed to his sharp interaction contrast, which may have affected gait dynamics and amplified sensor signals in particular feature ranges.

Table 7. Global SHAP plots for the pre-FOG class.



The top 10 features and their influence on the model's prediction for the pre-FOG class across the six patients are listed in table 7, which shows how SHAP was used. This global analysis helps in determining which features had the biggest effect on the model's ability to predict pre-FOG events.

High SHAP values for Patient 1 were linked to low levels of features, including 'x_variance', 'y_std_div', and 'x_iqr', suggesting a greater push toward pre-FOG predictions. Higher 'x_median' and 'y_variance' values, on the other hand, had less effect on pre-FOG prediction. These results imply that decreased IQR and variability in the x and y axes are important markers for identifying the patient's pre-FOG start.

Low 'x_minimum' values for Patient 2 were associated with high SHAP values, and globally important features like 'z_maximum', 'y_dominant_frequency', 'mag_skewness' and 'mag_kurtosis' were also crucial. The predictive importance of these features is further supported by the fact that they overlap with the top four features chosen by RF, listed in table 4.

For Patient 3, the SHAP analysis revealed that lower 'x_maximum' values were related to decreasing SHAP scores, whereas low 'y_peak_frequency' resulted in higher SHAP values. Six features from the SHAP summary matched the top RF-selected features, including 'x_maximum', 'y_peak_frequency', 'y_minimum', 'x_dominant_frequency', 'x_minimum' and 'y_maximum', confirming consistency in feature importance across the two methods.

In Patient 4, the features 'Ax_mean', 'Gy_variance', 'Gy_std_div' and 'Gy_spectral_energy' were related to low SHAP values, indicating little influence on pre-FOG prediction. Low values of 'Ax_icv' resulted in greater SHAP values, leading predictions to the pre-FOG class. The top RF feature list in

table 4 included five critical features supported by SHAP: 'Ax_mean', 'Ax_icv', 'Gy_variance', 'Gmag _max_psd' and 'Gy_std_div'.

Patient 5's SHAP plot showed that high values of 'Ax_peak_to_peak' were mostly associated with low SHAP values, but low values of 'Ax_maximum' contributed favourably to pre-FOG prediction. These findings confirm the significance of variable and peak features in early gait prediction. Seven top SHAP features matched RF-selected features in table 4: 'Ax_peak_to_peak', 'Gy_maximum', 'Ay_minimum', 'Ax_maximum', 'Amag_peak_to_peak', 'Ax_median' and 'Amag_minimum'.

For Patient 6, low values across the majority of the top SHAP-ranked features were related to higher SHAP scores, implying that suppressed or reduced feature values affected pre-FOG prediction. Some notable features include 'Gx_maximum' and 'Ay_iqr'. The top RF-selected list in table 4 shared five features: 'Gx_maximum', 'Gmag_minimum', 'Az_maximum', 'Gmag_icv' and 'Ay_iqr'.

Overall, the SHAP global summary is interpretable and well aligned with the RF feature selection results, highlighting the stability and significance of important features for individualized pre-FOG prediction. The findings also indicate individual diversity in which parameters have the most effect, underlining the importance of XAI in tailoring prediction models to patient-specific traits.

The local explanations produced by SHAP and LIME for all patients with respect to the pre-FOG class are listed in table 8. Some patients, like Patients 1 and 4, have explanations that correlate to distinct windows between the two methodologies, whereas other patients, like Patients 2 and 3, had the same window investigated.

A low chance of pre-FOG is indicated by the SHAP explanation for window #8 for Patient 1. While characteristics like 'y_zcr' (= -0.5132) and 'y_std_div' (= -0.6454) had a stronger opposing influence, driving the prediction towards non-pre-FOG, features like 'y_minimum' (= 0.6213) and 'x_kurtosis' (= 1.65) contributed positively to the pre-FOG prediction. On the other hand, window #4's LIME explanation for Patient 1 supports a positive pre-FOG prediction based on features like 'x_maximum' and 'mag _variance'. While certain variables like 'y_max_psd' and 'y_mean' indicate non-pre-FOG behaviour, their influence is less apparent.

The SHAP explanation for window #200 in table 8 indicates a significant likelihood of pre-FOG for Patient 2 with features, including 'z_dominant_frequency' (= -0.7229) and 'mag_skewness' (= 1.452) having significant positive contributions. Although there are some opposing forces, the LIME explanation for the same window similarly confirms the pre-FOG prediction. In particular, 'z_dominant_frequency' was considered non-indicative of pre-FOG in LIME owing to its value being marginally higher than the threshold (= -0.76); 'y_dominant_frequency', on the other hand, was in line with pre-FOG characteristics, ranging from 0.58 to 1.39.

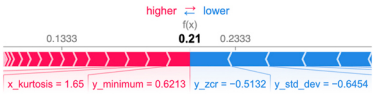
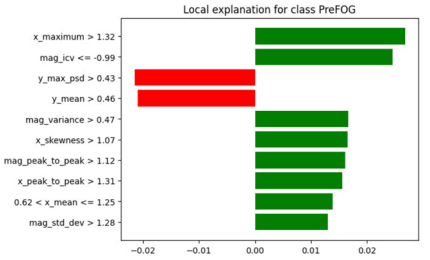
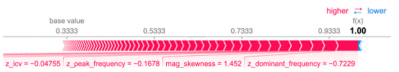
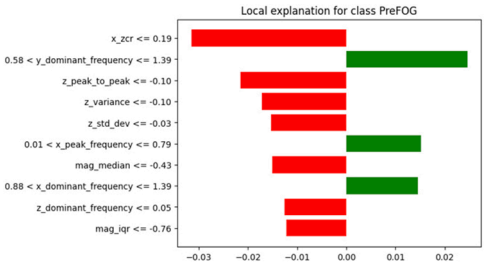
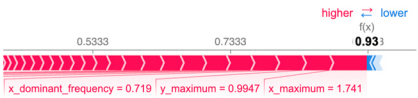
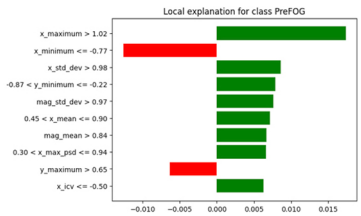
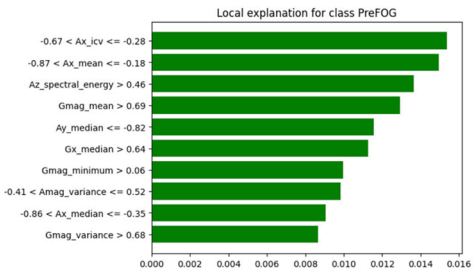
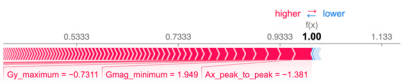
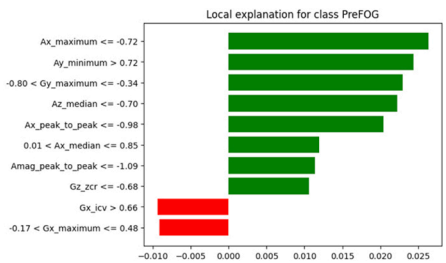
The high pre-FOG probability in Patient 3's SHAP explanation for window #178 is mainly driven by 'x _maximum' (= 1.741) and 'y_maximum' (= 0.9947). The classification is also supported by the LIME plot for the same window, yet with a slightly distinct interpretation. For example, unlike SHAP, 'y_maximum' seems to work against pre-FOG in the LIME output. Nonetheless, both plots attest to the importance of 'x_maximum'.

The LIME rule set indicates that Patient 3 has pre-FOG when 'x_maximum' values are more than 1.02, whereas Patient 1 has a higher threshold (1.32). This variation may be explained by the clinical setting and patient-specific gait profiles. Patient 3, who has a higher H&Y score and more severe motor symptoms, shows more sensitivity to slight changes in gait features, while Patient 1, who has had the disease for the longest (30 years), may need a more apparent deviation in sensor features to trigger pre-FOG classification.

LIME further shows that 'y_minimum' at −0.63 contributes to pre-FOG because it is inside the interval [-0.87, -0.22]. In contrast, 'y_maximum' (= 0.99), which is over the LIME threshold of 0.65, does not seem predictive on its own. Interestingly, this is not quite consistent with the 2D interaction plot in figure 11c, which indicates that the effect of the model is greatest when 'y_maximum' is approximately between -0.4 and 6.09 and 'y_minimum' is low (between -3 and -0.54), supporting the local SHAP interpretation and partially confirming the LIME insights.

The SHAP explanation for window #383 in table 8 indicates a high chance of pre-FOG for Patient 4, the oldest participant (84 years), with advanced PD and a high H&Y score of 4.0. Features that greatly contribute to the prediction include 'Gy_zcr' (= 0.7567) and 'Gx_skewness' (= 2.164), indicating significant variability and asymmetry in gyroscope signals, which are typical in older individuals with advanced gait problems. The pre-FOG categorization is supported by the LIME explanation for window #600, with values of 'Ax_icv' (between -0.67 and -0.28) and 'Ax_mean' (between -0.87 and -0.18) playing a significant role. These characteristics probably represent motion abnormalities and a decrease in lower-limb movement that are commonly observed in long-term progression.

Table 8. Local explainability plots (SHAP and LIME) for each patient—pre-FOG class.

patient	local SHAP plot	LIME plot
Patient 1	Window #8	Window #4
		
Patient 2	Window #200	Window #200
		
Patient 3	Window #178	Window #178
		
Patient 4	Window #383	Window #600
		
Patient 5	Window #900	Window #600
		

(*Continued.*)

Table 8. (*Continued.*)

patient	local SHAP plot	LIME plot
Patient 6	Window #383	Window #600

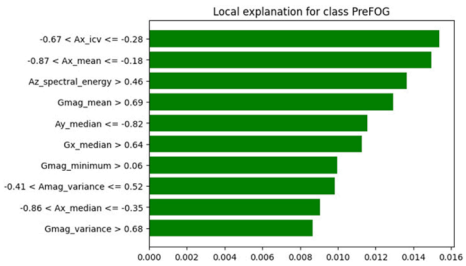


Table 9. Comparison of ML models for subject-dependent FOG prediction.

paper	model	pre-FOG	performance	latency	XAI
[19]	AdaBoost	5 s	accuracy = 82.7%	0.93 ± 0.089 s	NO
[40]	SVM	_	sensitivity: 88.09% (FOG)	_	NO
[28]	LDA	2 s	AUC (Area Under the Curve) = 90%	_	_
[29]	LSTM	_	accuracy = 88% (FOG)	_	NO
this work	RF	5 s	average accuracy = 98.67 ± 0.75%	average: 75.0 ± 31.1 ms	PDP, ICE, SHAP, LIME

Patient 5 showed a high pre-FOG likelihood in SHAP for window #900. He was a 69-year-old male with the shortest disease duration (1.5 years) in the IMU group, yet he had a moderate H&Y score of 3.0. Dominant contributors included features like 'Ax_peak_to_peak' (= -1.381) and 'Gmag_minimum' (= 1.949). 'Ax_peak_to_peak' and 'Gy_maximum' are highlighted as important supporting indicators in the LIME output for window #600, which supports the SHAP explanations. The LIME plot is further supported by the interaction plot in figure 12b and table 6, which indicates that the influence of pre-FOG prediction increases with 'Ax_maximum' values between -2.21 and -0.26. On the other hand, 'Gx_maximum' seems to have low effect on the prediction, suggesting that axial variability, rather than lateral or rotational gait dynamics, is what drives the model's sensitivity for this patient.

According to table 8, Patient 6 had a high pre-FOG probability in SHAP for window #383. Important characteristics include 'Gy_zcr' (= 0.7567) and 'Gx_skewness' (= 2.164), which can indicate quick signal transitions or subtle asymmetries that are common in younger individuals with reduced motor coordination but retained strength. According to the LIME description for window #600, most features support the prediction, with 'Ax_icv' and 'Ax_mean' appearing as crucial features. These findings provide insight into how age and disease trajectory affect pre-FOG prediction patterns by indicating that predictive patterns in younger people with mid-stage PD are driven by sudden neuromuscular alterations rather than recurrent instability.

4.2. Comparison of related work on subject-dependent FOG prediction

Table 9 provides a comparison of several studies for subject-dependent FOG prediction, highlighting the models used, pre-FOG window sizes, performance metrics, latency and the application of XAI techniques. The studies presented use a range of ML algorithms, including AdaBoost, SVM, linear discriminant analysis (LDA), long short-term memory (LSTM) and RF.

One of the most notable aspects of table 9 is the reported latency times. Most studies do not explicitly mention latency, or their latency is not clearly defined, which is a critical factor in real-time clinical applications. For example, Zhang *et al*. [19] used AdaBoost and reported a latency of 0.93 ± 0.089 s, which is acceptable for some applications but may not be optimal for real-time use. In contrast, the approach presented in this work demonstrates a significant improvement in latency, with an average latency of

75.0 ± 31.1 ms across six patients, which is much faster and more suitable for real-time monitoring of patients.

In addition, the inclusion of XAI techniques sets this work apart from others. Although most studies do not incorporate any explainability methods, this work applies a range of XAI techniques such as PDP, ICE lines, SHAP and LIME. These methods provide transparency into how the model makes predictions, allowing clinicians to understand the rationale behind the system's output.

Improvements in both performance (average prediction accuracy of 98.67% across six patients) and latency (75 ms), along with the integration of XAI techniques, make this approach a promising step forward in the personalized and transparent prediction of FOG in PD.

5. Conclusion and future work

FOG presents a significant challenge in managing PD, as its unpredictable nature can lead to serious falls and mobility issues. In this work, we developed a personalized prediction system for early prediction of FOG in patients with PD using a subject-dependent approach coupled with XAI techniques. The RF model demonstrated an average accuracy of 98.67 ± 0.75% and a latency of 75.0 ± 31.1 ms across six patients, making it suitable for real-time clinical applications. The integration of XAI techniques such as SHAP, LIME and PDPs improved the model's interpretability and ensured transparency in the decision-making process, enhancing clinical trust. The explainability analysis revealed significant feature interactions, such as 'mag_peak_to_peak' and 'x_mean', influencing the predictions for FOG, pre-FOG and walking states. Both SHAP and LIME methods provided transparent insights into the model's behaviour, confirming the importance of features like 'x_maximum' and 'Ax_maximum' in predicting pre-FOG.

Future work for the personalized prediction system should focus on incorporating data from multiple sensor set-ups and longer recording periods to enhance the reliability of FOG predictions. In addition, integrating the prediction system into wearable devices for continuous real-time monitoring will enable timely alerts and support personalized interventions, making the system more scalable and accessible for personalized patient care.

Ethics. This work did not require ethical approval from a human subject or animal welfare committee.

Data accessibility. This study uses only publicly available datasets that have been previously published. All datasets referenced are properly cited within the manuscript. No new data were generated or analysed during the current study. The following linked DOI leads to both datasets we used: https://doi.org/10.3389/fnins.2022.832463, [41,42].

Declaration of AI use. We have not used AI-assisted technologies in creating this article.

Authors contributions. H.E.: conceptualization, methodology, resources, writing—original draft; A.H.: project administration, supervision, writing—review and editing; M.A.-S.: investigation, resources, writing—review and editing; W.K.: formal analysis, methodology, supervision, writing—review and editing; S.M.: resources, supervision, writing—review and editing.

All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Conflict of interest declaration. We declare we have no competing interests.

Funding. No funding has been received for this article.

Acknowledgements. The authors would like to thank the contributors of the publicly available DAPHNet and IMU-PD datasets for making their data accessible for research purposes. The authors also acknowledge the valuable support and feedback provided by colleagues during the development of this work.

References

1. Bloem BR, Okun MS, Klein C. 2021 Parkinson's disease. *Lancet* **397**, 2284–2303. (doi:10.1016/S0140-6736(21)00218-X)
2. Hayes MT. 2019 Parkinson's disease and parkinsonism. *Am. J. Med.* **132**, 802–807. (doi:10.1016/j.amjmed.2019.03.001)
3. Rea P. 2015 *Essential clinical anatomy of the nervous system*. Elsevier Inc. (doi:10.1016/B978-0-12-802030-2.00001-7)
4. Sveinbjornsdottir S. 2016 The clinical symptoms of Parkinson's disease. *J. Neurochem.* **139**, 318–324. (doi:10.1111/jnc.13691)
5. Amboni M *et al*. 2015 Prevalence and associated features of self-reported freezing of gait in Parkinson disease: the DEEP FOG study. *Parkinsonism Relat. Disord.* **21**, 644–649. (doi:10.1016/j.parkreldis.2015.03.028)
6. Giladi N, Hausdorff JM. 2006 The role of mental function in the pathogenesis of freezing of gait in Parkinson's disease. *J. Neurol. Sci.* **248**, 173–176. (doi:10.1016/j.jns.2006.05.015)
7. Nutt JG, Bloem BR, Giladi N, Hallett M, Horak FB, Nieuwboer A. 2011 Freezing of gait: moving forward on a mysterious clinical phenomenon. *Lancet Neurol.* **10**, 734–744. (doi:10.1016/S1474-4422(11)70143-0)

8. Nonnekes J, Snijders AH, Nutt JG, Deuschl G, Giladi N, Bloem BR. 2015 Freezing of gait: a practical approach to management. *Lancet Neurol*. **14**, 768–778. (doi:10.1016/S1474-4422(15)00041-1)
9. Giladi N, Nieuwboer A. 2008 Understanding and treating freezing of gait in parkinsonism, proposed working definition, and setting the stage. *Mov. Disord*. **23**, S423–5. (doi:10.1002/mds.21927)
10. Okuma Y. 2014 Practical approach to freezing of gait in Parkinson's disease. *Pract. Neurol*. **14**, 222–230. (doi:10.1136/practneurol-2013-000743)
11. Elbatanouny H, Kleanthous N, Alusi S, Mahmoud S, Hussain A. 2024 Navigating the freeze: a machine learning approach to detect freezing of gait in Parkinson's patients. In *2024 IEEE Int. Conf. on Omni-layer Intelligent Systems (COINS)*, pp. 1–4. Khorfakan, UAE: IEEE. (doi:10.1109/COINS61597.2024.10622129)
12. Elbatanouny H, Kleanthous N, Salah S, Mahmoud S, Hussain A. 2024 Enhancing freezing of gait prediction in Parkinson's disease using machine learning and explainable AI. In *2024 17th Int. Conf. on Development in eSystem Engineering (DeSE)*, pp. 236–241. Khorfakan, UAE: IEEE. (doi:10.1109/DeSE63988.2024.10912039)
13. Borzì L, Mazzetta I, Zampogna A, Suppa A, Olmo G, Irrera F. 2021 Prediction of freezing of gait in Parkinson's disease using wearables and machine learning. *Sensors* **21**, 614. (doi:10.3390/s21020614)
14. Ferster ML, Mazilu S, Tröster G. 2015 Gait parameters change prior to freezing in Parkinson's disease: a data-driven study with wearable inertial units. *EAI Endorsed Trans. Ambient Syst.* **3**, 1–8. (doi:10.4108/eai.28-9-2015.2261411)
15. Mazilu S, Blanke U, Dorfman M, Gazit E, Mirelman A, Hausdorff JM, Tröster G. 2015 A wearable assistant for gait training for Parkinson's disease with freezing of gait in out-of-the-lab environments. *ACM Trans. Interact. Intell. Syst.* **5**, 1–31. (doi:10.1145/2701431)
16. Al-Tayeb KAS, Elbatanouny H, Hussain A. 2024 Effective noise reduction in biomedical speech signals: a case study on Parkinson's disease. In *2024 17th Int. Conf. on Development in eSystem Engineering (DeSE)*, pp. 37–41. Khorfakan, UAE: IEEE. (doi:10.1109/DeSE63988.2024.10912057)
17. Elbatanouny H, Kleanthous N, Dahrouj H, Alusi S, Almajali E, Mahmoud S, Hussain A. 2024 Insights into Parkinson's disease-related freezing of gait detection and prediction approaches: a meta analysis. *Sensors* **24**, 3959. (doi:10.3390/s24123959)
18. Barredo Arrieta A *et al*. 2020 Explainable artificial intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI. Inf. Fusion. **58**, 82–115. (doi:10.1016/j.inffus.2019.12.012)
19. Zhang Y, Yan W, Yao Y, Ahmed JB, Tan Y, Gu D. 2020 Prediction of freezing of gait in patients with Parkinson's disease by identifying impaired gait patterns. *IEEE Trans. Neural Syst. Rehabil. Eng*. **28**, 591–600. (doi:10.1109/TNSRE.2020.2969649)
20. Sun H, Ye Q, Xia Y. 2024 Predicting freezing of gait in patients with Parkinson's disease by combination of manually-selected and deep learning features. *Biomed. Signal Process. Control* **88**, 105639. (doi:10.1016/j.bspc.2023.105639)
21. Bachlin M, Plotnik M, Roggen D, Maidan I, Hausdorff JM, Giladi N, Troster G. 2010 Wearable assistant for Parkinson's disease patients with the freezing of gait symptom. *IEEE Trans. Inf. Technol. Biomed*. **14**, 436–446. (doi:10.1109/TITB.2009.2036165)
22. Xia Y, Sun H, Zhang B, Xu Y, Ye Q. 2024 Prediction of freezing of gait based on self-supervised pretraining via contrastive learning. *Biomed. Signal Process. Control* **89**, 105765. (doi:10.1016/j.bspc.2023.105765)
23. Li H. 2021 Multimodal dataset of freezing of gait in Parkinson's disease. *Mendeley Data* **3**. (doi:10.17632/r8gmbtv7w2.3)
24. Filtjens B, Ginis P, Nieuwboer A, Afzal MR, Spildooren J, Vanrumste B, Slaets P. 2021 Modelling and identification of characteristic kinematic features preceding freezing of gait with convolutional neural networks and layer-wise relevance propagation. *BMC Med. Inform. Decis. Mak*. **21**, 341. (doi:10.1186/s12911-021-01699-0)
25. Priyadharshini S, Ramkumar K, Vairavasundaram S, Narasimhan K, Venkatesh S, Amirtharajan R, Kotecha K. 2024 A comprehensive framework for Parkinson's disease diagnosis using explainable artificial intelligence empowered machine learning techniques. *Alex. Eng. J.* **107**, 568–582. (doi:10.1016/j.aej.2024.07.106)
26. Fearon C, Butler JS, Waechter SM, Killane I, Kelly SP, Reilly RB, Lynch T. 2021 Neurophysiological correlates of dual tasking in people with Parkinson's disease and freezing of gait. *Exp. Brain Res*. **239**, 175–187. (doi:10.1007/s00221-020-05968-8)
27. Fahad MS, Ranjan A, Kumar G. 2025 Enhancing patient-independent detection of freezing of gait in Parkinson's disease with deep adversarial network. In *Neural computing and applications*, pp. 1–21. UK: Springer.
28. Palmerini L, Rocchi L, Mazilu S, Gazit E, Hausdorff JM, Chiari L. 2017 Identification of characteristic motor patterns preceding freezing of gait in Parkinson's disease using wearable sensors. *Front. Neurol*. **8**, 394. (doi:10.3389/fneur.2017.00394)
29. Torvi VG, Bhattacharya A, Chakraborty S. 2018 Deep domain adaptation to predict freezing of gait in patients with Parkinson's disease. In *2018 17th IEEE Int. Conf. on Machine Learning and Applications (ICMLA)*, pp. 1001–1006. Orlando, FL: IEEE. (doi:10.1109/ICMLA.2018.00163)
30. Ribeiro De Souza C *et al*. 2022 A public data set of videos, inertial measurement unit, and clinical scales of freezing of gait in individuals with Parkinson's disease during a turning-in-place task. *Front. Neurosci*. **16**, 832463. (doi:10.3389/fnins.2022.832463)
31. Antonsson EK, Mann RW. 1985 The frequency content of gait. *J. Biomech*. **18**, 39–47. (doi:10.1016/0021-9290(85)90043-0)
32. Zhou H, Hu H. 2008 Human motion tracking for rehabilitation—a survey. *Biomed. Signal Process. Control* **3**, 1–18. (doi:10.1016/j.bspc.2007.09.001)
33. Elbatanouny H, Kleanthous N, Salah S, Mahmoud S, Hussain A. 2024 Enhancing freezing of gait prediction in Parkinson's disease using machine learning and explainable AI. In *2024 17th Int. Conf. on Development in eSystem Engineering (DeSE)*, pp. 236–241. Khorfakkan, UAE: IEEE. (doi:10.1109/DeSE63988.2024.10912039)
34. Orphanidou NK, Hussain A, Keight R, Lishoa P, Hind J, Al-Askar H. 2018 Predicting freezing of gait in Parkinsons disease patients using machine learning. In *2018 IEEE Congress on Evolutionary Computation (CEC)*, pp. 1–8. Rio de Janeiro, Brazil: IEEE. (doi:10.1109/CEC.2018.8477909)
35. Moore ST, MacDougall HG, Ondo WG. 2008 Ambulatory monitoring of freezing of gait in Parkinson's disease. *J. Neurosci. Methods* **167**, 340–348. (doi:10.1016/j.jneumeth.2007.08.023)

36. Elen A, Avuçlu E. 2021 Standardized variable distances: a distance-based machine learning method. *Appl. Soft Comput*. **98**, 106855. (doi:10.1016/j.asoc.2020.106855)
37. Fernandez A, Garcia S, Herrera F, Chawla NV. 2018 SMOTE for learning from imbalanced data: progress and challenges, marking the 15-year anniversary. *J. Artif. Intell. Res.* **61**, 863–905. (doi:10.1613/jair.1.11192)
38. Shekhar S, Bansode A, Salim A. 2021 A comparative study of hyper-parameter optimization tools. In *2021 IEEE Asia-Pacific Conf. on Computer Science and Data Engineering (CSDE)*, pp. 1–6. Brisbane, Australia: IEEE. (doi:10.1109/CSDE53843.2021.9718485)
39. Khosla A, Kumar N, Khera P. 2024 Machine learning approach for predicting state transitions via shank acceleration data during freezing of gait in Parkinson's disease. *Biomed. Signal Process. Control* **92**, 106053. (doi:10.1016/j.bspc.2024.106053)
40. Rodríguez-Martín D *et al*. 2017 Home detection of freezing of gait using support vector machines through a single waist-worn triaxial accelerometer. *PLoS One* **12**, e0171764. (doi:10.1371/journal.pone.0171764)
41. Bachlin M *et al*. 2010 Wearable assistant for Parkinson's disease patients With the freezing of gait symptom. In *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, pp. 436–446, IEEE. (doi:10.1109/TITB.2009.2036165)
42. Ribeiro De Souza C. 2022 A Public Data Set of Videos, Inertial Measurement Unit, and Clinical Scales of Freezing of Gait in Individuals With Parkinson's Disease During a Turning-In-Place Task. *Front. Neurosci* **16**, 832463. (doi:10.3389/fnins.2022.832463)