British Journal of Multidisciplinary and Advanced Studies,6(6)1-20, 2025
Engineering and Technology
Print ISSN: 2517-276X
Online ISSN: 2517-2778
https://bjmas.org/index.php/bjmas/index

Early Prediction of Freezing of Gait in Parkinson's Disease Using Wearable Sensors and Machine Learning

Shreyan Khanna
Westwood High School, Austin, Texas, United States

doi: https://doi.org/10.37745/04972 Published November 03, 2025

Citation: Khanna S. (2025) Early Prediction of Freezing of Gait in Parkinson's Disease Using Wearable Sensors and Machine Learning, *British Journal of Multidisciplinary and Advanced Studies*,6(6)1-20

Abstract: *Freezing of Gait (FoG) is a disabling motor symptom of Parkinson's Disease (PD) that causes loss of mobility and increases fall risk. Most wearable-based FoG systems detect freezing at or after onset, limiting their preventive use. This study presents a machine learning framework to predict FoG several seconds before onset. The publicly available Daphnet dataset was used, containing tri-axial accelerometer data segmented with sliding windows and processed using domain-specific feature engineering. Several models were tested under a Leave-One-Subject-Out (LOSO) cross-validation protocol, including XGBoost, CatBoost, Random Forest, and deep recurrent networks. XGBoost achieved the best results (Recall = 90.03%, Precision = 76.76%, F1 = 82.14%). SHAP analysis confirmed physiologically relevant predictors such as gait rhythmicity, energy distribution, and freeze index. The framework can support real-time alert systems to help prevent FoG-related falls in PD. Future work will include hardware implementation, clinical testing, and development of neuromorphic variants for wearable devices.*

Keywords: Parkinson's disease, freezing of gait, machine learning, regression, wearable sensors, early detection

INTRODUCTION

Parkinson's Disease (PD) is a progressive neurodegenerative disorder that affects more than ten million people worldwide and continues to rise in prevalence as global populations age [1]. It is primarily characterized by the degeneration of dopaminergic neurons in the substantia nigra pars compacta, resulting in hallmark motor symptoms such as bradykinesia, rigidity, tremor, and gait abnormalities [2]. Among these motor manifestations, Freezing of Gait (FoG) represents one of the most disabling and unpredictable complications. FoG is defined as a sudden and transient inability to initiate or continue locomotion despite the intention to walk, often described by patients as feeling that their *"feet are glued to the floor"* [3]. This episodic gait block significantly increases fall risk, reduces confidence in mobility, and is strongly associated with social withdrawal,

depression, and substantial deterioration in quality of life. Epidemiological evidence suggests that approximately 70% of PD-related falls are associated with FoG episodes [4], and its prevalence increases with disease progression, affecting up to 80% of individuals in advanced PD stages [5].

Despite advances in pharmacotherapy and neuromodulation strategies such as levodopa and deep-brain stimulation (DBS), FoG remains difficult to treat and frequently persists or worsens even with optimized medication regimes [6]. Furthermore, FoG commonly emerges in real-world scenarios involving turning, gait initiation, cognitive dual-tasking, and passage through narrow environments, highlighting the complex interplay between motor, spatial, and cognitive mechanisms [7]. These characteristics make FoG detection and intervention a continuing challenge in both clinical and community settings.

Recent technological progress has enabled the use of wearable inertial sensors, smartphones, and machine-learning algorithms to detect FoG from real-time gait signals [8-10]. However, existing research overwhelmingly focuses on reactive FoG detection, recognizing freezing either at onset or after it occurs rather than predicting it in advance. Techniques such as threshold-based detection of gait rhythmicity, frequency analysis including the Freeze Index, and step variability heuristics have demonstrated utility but remain limited to post-event identification and provide no actionable lead time for preventing falls [11]. Deep learning-based architectures, particularly convolutional and recurrent neural networks, have been proposed to address temporal complexity and improve classification performance [12]. Yet these methods continue to suffer from several limitations: high computational complexity unsuitable for continuous wearable deployment, lack of latency consideration for real-time usage, insufficient cross-subject generalization, and limited interpretability that reduces clinician trust and adoption in medical environments. Most importantly, most existing systems focus on detection rather than true anticipatory prediction, which is essential for real-world fall prevention and active gait support. To date, there remains a critical research need for solutions that can predict FoG events before onset, operate efficiently on wearable platforms in everyday environments, generalize across unseen patients, and maintain clinical interpretability to support acceptance among healthcare providers. A predictive system capable of providing even a short warning window, on the order of seconds, could enable timely cueing interventions such as auditory metronomes, visual guidance systems, vibrotactile feedback, or adaptive DBS strategies, thus reducing fall risk and enabling safer mobility [13].

Motivated by this gap, this study explores whether tri-axial wearable accelerometer data can be leveraged using machine learning models to anticipate FoG episodes several seconds before clinical manifestation. We present a feature-engineered machine-learning framework trained on the publicly available Daphnet Freezing of Gait dataset [14], applying time–frequency domain gait descriptors combined with clinically interpretable biomechanical features. Model performance is evaluated using Leave-One-Subject-Out cross-validation (LOSO-CV) [15] to assess generalization to unseen individuals. Additionally, we employ SHAP explainability analysis to ensure model transparency and physiological relevance of predictive features. The contributions of this work are fourfold:

- Focus on prediction, shifting from reactive detection to proactive prevention of FoG.
- Development of a computationally efficient pipeline suitable for real-time wearable deployment.
- Robustness evaluation through rigorous subject-independent validation.
- Improved clinical interpretability via model explanation, facilitating translational use in therapeutic and home-monitoring contexts.

The remainder of this paper is structured as follows. Section II reviews related work on FoG detection and prediction using wearable sensing and machine learning. Section III details the proposed methodology, including dataset description, feature extraction, and model development. Section IV presents experimental results and comparative evaluation. Section V discusses clinical implications, interpretability analysis, and limitations. Finally, Section VI concludes the paper and outlines future research directions.

Related Work

Research on Freezing of Gait (FoG) in Parkinson's disease (PD) has progressed from early threshold-based detection to deep-learning-driven modelling of pre-freeze gait dynamics [16]. However, most prior work remains focused on detection at or after onset, with comparatively little emphasis on proactive prediction and generalization to unseen subjects.

Early ambulatory studies established the foundations for wearable FoG monitoring using inertial sensors. Moore *et al.* [17] first defined the Freeze Index (FI), the ratio of spectral power in the 3–8 Hz "freeze" band to the 0.5–3 Hz locomotor band, demonstrating its discriminative capability for real-world FoG episodes. Bächlin *et al.* [18] later operationalized this feature in a wearable assistant that delivered auditory cues when FI thresholds were exceeded, forming the basis of the Daphnet dataset, now the field's benchmark. Systematic reviews by Silva de Lima *et al.* [19], Pardoel *et al.* [20], and Gregorčič *et al.* [10] identified three persistent trends:

- accelerometers remain the most practical modality for real-world FoG detection, typically placed on the shank, thigh, or trunk.
- handcrafted frequency- and variability-based features (e.g., FI, step variance) continue to dominate conventional machine-learning (ML) pipelines; and
- although deep models improve detection accuracy, they struggle with subject-independent robustness.

With the growing integration of artificial intelligence in healthcare, machine learning (ML) algorithms have become central to *freezing of gait* (FoG) detection and prediction in Parkinson's disease. Early ML approaches leveraged engineered gait descriptors with classifiers such as support vector machines, k-nearest neighbours, and random forests, achieving sensitivity and specificity rates above 0.90 in controlled studies [21] [22]. The effectiveness of these traditional

models, however, is highly dependent on the extraction of robust and physiologically meaningful features, which directly influences both model generalization and interpretability [23] [24].

Recent developments in deep learning (DL) have introduced automated feature learning from raw inertial data, significantly reducing the reliance on handcrafted representations. Architectures such as convolutional neural networks (CNNs) [25] [26], recurrent and long short-term memory (LSTM) models [27] [28], and transformer networks [29] have demonstrated substantial performance gains by capturing the complex spatiotemporal patterns underlying FoG. Other studies have explored deep autoencoders and unsupervised learning frameworks to detect latent gait anomalies without extensive labelling [30] [31]. While these advances mark significant progress toward intelligent FoG monitoring, limitations in generalization, transparency, and wearable efficiency persist. This motivates the present work, which focuses on predictive, interpretable, and computationally efficient modeling for real-world clinical applications.

METHODOLOGY

This section delineates the dataset, sensing configuration, signal preprocessing procedures, feature extraction pipeline, class imbalance mitigation strategies, learning architectures employed, and validation methodology. The primary objective is to develop a computationally efficient, clinically interpretable, and methodologically rigorous framework for proactive prediction of Freezing of Gait (FoG) in individuals with Parkinson's Disease (PD) using wearable inertial sensor data.

Participants and Dataset

This study employs the publicly available Daphnet Freezing of Gait Dataset [14]; a benchmark resource extensively used in Parkinson's gait research. The dataset contains tri-axial accelerometer recordings from ten patients diagnosed with Parkinson's disease (PD) who performed structured walking tasks, such as, straight walking, repetitive turning, obstacle and narrow-passage traversal, and stop–start walking under controlled laboratory conditions. These activities emulate real-world movement transitions known to elicit freezing-of-gait (FoG) behaviour. Each participant in the Daphnet protocol wore three synchronized wearable inertial sensors positioned on the lateral shank (ankle), mid-thigh, and upper trunk (sternal notch). The sensors recorded linear acceleration along three orthogonal axes (x, y, z) at 64 Hz sampling frequency with a ±6 g dynamic range, providing nine channels of motion data per time step. These specifications determined by the original data providers have been validated in neurological gait-analysis studies as adequate for resolving gait cycles, tremor bursts, and transitional motion patterns. Across all ten subjects, the dataset comprises approximately eight hours of motion data, annotated by clinical experts into three gait states: normal locomotion (class 0), pre-FoG transitional phase (class 1), and FoG episode (class 2). The synchronized video-based annotations ensure precise temporal alignment between sensor signals and gait state transitions. The availability of these high-resolution, multimodal annotations makes the dataset particularly suitable for developing predictive algorithms capable of anticipating FoG onset.

Data Preprocessing

The raw accelerometer data streams were converted into structured learning samples using a fixed-length sliding-window segmentation technique. Each continuous signal was partitioned into 2-second windows with 50% overlap, resulting in a 1-second stride between consecutive windows. This segmentation strategy achieves an optimal trade-off between temporal granularity and computational efficiency, as a 2-second window typically encompasses one to two gait cycles in PD patients while remaining responsive enough for near-real-time inference in wearable systems. Each window contained 128 samples per channel (64 Hz × 2 s), capturing both slow postural drifts and higher-frequency tremor bursts associated with early hesitation or micro-freezing phenomena. Window labelling followed a dominant-class rule, where the majority annotation within the segment defined its class.

To operationalize predictive rather than reactive modelling, the labelling scheme was reformulated to emphasize pre-freeze indicators. Specifically, pre-FoG segments (class 1) were retained as the *positive* class representing elevated freezing risk, while normal gait (class 0) served as the *negative* class. The FoG episodes (class 2) themselves were excluded from model training to prevent bias toward post-onset recognition and to ensure that the learned discriminative patterns reflected anticipatory cues preceding actual freezing. All channels were detrended and mean-centered to eliminate baseline drift and gravitational offsets resulting from posture or sensor placement. Unlike traditional FoG detection pipelines, no high-pass filtering was applied, since low-amplitude, low-frequency components often encode subtle instability or compensatory sway preceding FoG onset. By retaining the full frequency band, the system preserves diagnostically relevant micro-oscillatory features. Furthermore, segmentation and preprocessing were performed subject-wise before cross-validation to prevent temporal leakage between training and test samples originating from the same individual, thereby maintaining strict subject-independence in evaluation.

Feature Engineering

Feature engineering was designed to balance predictive performance, computational efficiency, and clinical interpretability, a criteria essential for potential deployment in wearable cueing systems. For each of the nine accelerometer channels, eight statistical and temporal descriptors were extracted: mean, median, minimum, maximum, standard deviation, skewness, signal energy, and zero-crossing count. These features encapsulate gait amplitude, variability, symmetry, and energetic characteristics across cycles, capturing phenomena such as asymmetric oscillations and transient hesitation bursts typical in PD locomotion.

In addition, two FoG-specific physiological biomarkers were incorporated to enhance sensitivity to prodromal freezing signatures. The Freeze Index (FI), defined as the ratio of spectral power in the 3–8 Hz band (associated with trembling or shuffling) to that in the 0.5–3 Hz physiological walking band, quantifies emerging high-frequency oscillations preceding FoG. The Global Zero-Crossing Rate (GZCR) represents the cumulative zero-crossings across all channels, signifying whole-body instability and micro-oscillatory adjustments during gait deceleration. In total, 74 features were computed per 2-second window, 72 from statistical descriptors and two from FoG biomarkers. All features were z-score normalized prior to training to ensure scale uniformity across

axes and participants. Computation pipelines were optimized for low-latency extraction, with feature computation completing in sub-millisecond time on modern embedded processors. This architecture maintains feasibility for real-time feedback systems in wearable or mobile contexts, while the interpretable biomechanical descriptors enable transparent clinical interpretation of the predictive outcomes.

Table I. Summary of Engineered Features

Feature Category	Feature	Formula / Definition	Physiological Relevance
Time-Domain Statistics	Mean	$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$	Baseline gait amplitude and overall step magnitude
	Median	50th percentile of x_i	Robust central tendency; gait regularity
	Minimum / Maximum	$\min(x_i), \max(x_i)$	Range of motion; excursion limits across gait cycles
Variability Metrics	Standard Deviation	$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2}$	Stride-to-stride variability and instability
Distribution Shape	Skewness	$\gamma = \frac{1}{N}\sum_{i=1}^{N}(\frac{x_i - \mu}{\sigma})^3$	Asymmetric gait oscillations; PD-specific motor imbalance
Signal Energy	Energy	$E = \sum_{i=1}^{N} x_i^2$	Step force, movement intensity, and fatigue-related changes
Temporal Dynamics	Zero-Crossing Count	Number of times $x_i$crosses zero	Tremor bursts, hesitation events, or abrupt corrective motions
FoG-Specific Biomarkers	Freeze Index (FI)	$FI = \frac{P_{3\text{–}8\ \text{Hz}}}{P_{0.5\text{–}3\ \text{Hz}}}$	Quantifies high-frequency "trembling-in-place'' oscillations preceding FoG
	Global Zero-Crossing Rate (GZCR)	Total zero-crossings aggregated across all 9 channels	Captures whole-body instability and hesitation oscillations

Note — All features were computed per axis across the nine accelerometer channels unless otherwise specified.

Handling of Class Imbalance

Since FoG and pre-FoG instances represent a small proportion of the dataset with typically less than 10% of total gait segments addressing class imbalance was essential to prevent model bias toward the predominant normal gait class. A subject-wise majority-class under sampling approach was employed within each training fold to equalize sample distribution while preserving genuine

signal dynamics. Synthetic oversampling methods such as SMOTE were intentionally avoided to prevent artificial distortions in temporal structure and to maintain the physiological authenticity of pre-FoG patterns.

Model Development and Architecture

The complete architecture of the proposed FoG early-prediction framework is depicted in Figure 1, which outlines the end-to-end pipeline from IMU data acquisition to interpretability-driven decision generation. The framework integrates dual modeling tiers, feature-engineered ensemble learning and deep sequence models, which integrates both modeling tiers via a soft-voting ensemble and applies SHAP explainability to identify clinically meaningful gait biomarkers driving FoG-risk predictions.

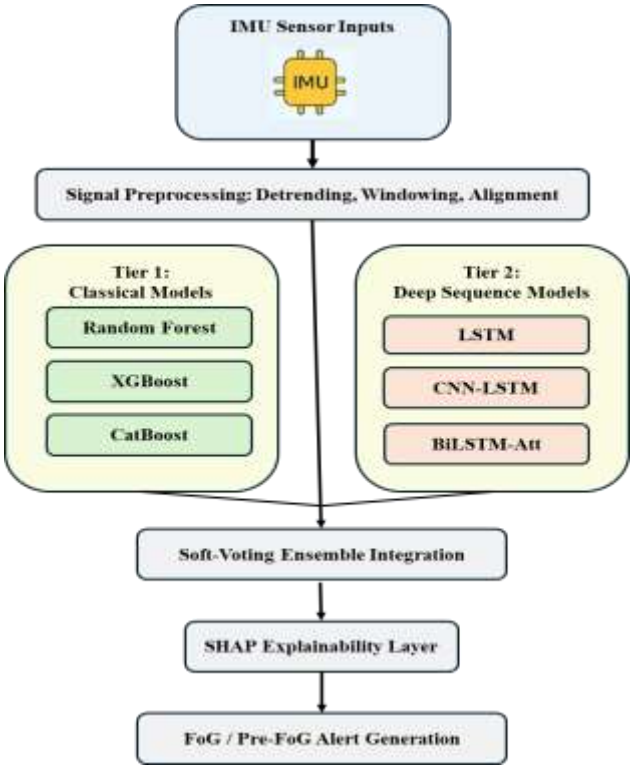

Fig.1. Overall architecture of the proposed early Freezing of Gait (FoG) prediction framework.

To ensure methodological rigour and comprehensive benchmarking, a multi-tiered modelling framework was implemented encompassing classical ensemble learning, gradient-boosting approaches, deep sequence architectures, and hybrid ensemble models. The first tier included Random Forest (RF), XGBoost, and CatBoost classifiers trained on the 74-dimensional engineered feature vectors. These models were selected for their robustness to non-linear interactions, suitability for small-to-moderate feature spaces, and intrinsic interpretability via feature importance analysis. RF leverages bagging and random feature selection to reduce variance, while

XGBoost and CatBoost utilize gradient-based boosting with L2 regularization [32], learning-rate scheduling, and controlled tree depth to mitigate overfitting. CatBoost's ordered boosting further stabilizes learning under class imbalance, making it particularly suitable for rare-event modelling such as FoG prediction.

The second tier comprised deep sequence models trained directly on normalized raw accelerometer sequences to explore temporal dynamics without handcrafted features. The architectures evaluated included a Long Short-Term Memory (LSTM) network for capturing temporal dependencies, a Convolutional LSTM (CNN-LSTM) hybrid combining local spatial convolution with long-term temporal modelling, and a Bidirectional LSTM with Attention (BiLSTM-Att) designed to focus selectively on critical temporal segments. While computationally intensive, these models served as performance upper bounds against the lightweight feature-based classifiers. A hybrid soft-voting ensemble was subsequently constructed by integrating the probabilistic outputs of the XGBoost and RF models, thereby uniting the bias-reduction strength of boosting with the variance-reduction benefit of bagging. Ensemble weights were determined in proportion to each model's validation F1-score, producing a balanced and robust consensus prediction mechanism.
All models were trained and evaluated under a Leave-One-Subject-Out Cross-Validation (LOSO-CV) protocol to ensure strict subject-independent testing. This methodology guarantees that each subject's data is excluded entirely from the training phase during its respective evaluation fold, providing an unbiased estimate of inter-individual generalization capability. Hyperparameters for each model family were tuned empirically using grid search strategies guided by cross-validation feedback. Regularization, learning rate, and depth parameters were optimized for gradient-boosted models, whereas dropout, batch normalization, and early stopping were applied to deep networks to control overfitting.

For interpretability, SHAP (Shapley Additive Explanations) analysis [33] was integrated to quantify feature-level contributions to prediction outcomes. SHAP's additive game-theoretic framework enables physiologically grounded interpretation, confirming that features such as Freeze Index, signal variability, and zero-crossing rate significantly influence FoG-risk classification. This interpretability ensures that the proposed model aligns with clinical expectations and enhances trustworthiness for future clinical adoption.

For clarity and reproducibility, we provide a consolidated algorithmic summary of the proposed pipeline. The algorithm below outlines the sequential stages from wearable signal acquisition through preprocessing, feature extraction, class handling, model training and evaluation, to SHAP-based interpretability and alert generation.

Algorithm 1: FoG Early-Prediction Pipeline

Input: Raw tri-axial accelerometer signals (ankle, thigh, trunk; 9 channels @ 64 Hz) *Output:* Early FoG / Pre-FoG alert with SHAP-based feature explanations
1. Signal Acquisition Collect synchronized wearable IMU streams (9 channels, ±6g, 64 Hz)
2. Preprocessing (a) Time-align all channels (b) Detrend and zero-center signal (c) Segment into 2-s windows with 50% overlap (d) Assign window label via majority-vote annotation
3. Label Reformulation Merge {Pre-FoG, FoG} → Positive class Retain Normal → Negative class
4. Feature Engineering (74-dim vector/window) For each of 9 axes compute: mean, median, min, max, std, skewness, zero-crossings, energy (72 total) Compute FoG-specific biomarkers: Freeze Index (3–8 Hz / 0.5–3 Hz), global zero-crossings (2 total)
5. Class Imbalance Handling Apply subject-wise majority undersampling (no SMOTE)
6. Model Training (LOSO-CV) Train classical models: RF, XGBoost, CatBoost Train deep baselines: LSTM, CNN-LSTM, BiLSTM-Attention Train hybrid ensemble: Soft-voting RF + XGBoost Evaluate using LOSO-CV
7. Evaluation Metrics Compute Accuracy, Precision, Recall, F1-Score
8. Explainability Apply SHAP to quantify per-feature contribution Validate biomechanical relevance with freeze signatures
9. Output Generate early-warning FoG prediction + interpretability scores

RESULTS

Model performance was evaluated using Accuracy, Precision, Recall (Sensitivity), and F_1-score, which together provide a holistic understanding of classifier behaviour. Given the safety-critical nature of *Freezing of Gait (FoG)* prediction, Recall was prioritized since false negatives (missed FoG events) can directly increase fall risk. Precision was also considered important to minimize

false alarms and prevent cueing fatigue or user frustration. To visualize classifier behaviour, confusion matrices were generated for each subject, and SHAP explainability plots were produced to interpret model decisions and the contribution of individual biomechanical features. To benchmark model robustness, multiple algorithms were compared under identical LOSO-CV partitions, including deep sequence learners (LSTM, CNN-LSTM, BiLSTM) and classical ensemble models (Random Forest, XGBoost, CatBoost), along with a hybrid ensemble (XGB + RF). Performance metrics were averaged across all folds to assess subject-independent generalization.

As summarized in Table 1, the comparative performance of all evaluated models under the LOSO-CV framework reveals distinct behavioral characteristics between deep sequence architectures and classical ensemble learners. While LSTM-based networks exhibited strong sensitivity to pre-FoG transitions, their low precision indicates over-detection tendencies. In contrast, gradient-boosted ensembles achieved more balanced performance, confirming their robustness and suitability for wearable deployment.

Table 1. Overall Model Performance under LOSO-CV

Model	Accuracy	Precision	Recall	F1 Score
LSTM	0.75	0.35	0.85	0.48
CNN-LSTM	0.78	0.38	0.82	0.52
BiLSTM	0.80	0.42	0.84	0.56
Random Forest	0.82	0.60	0.70	0.64
XGBoost	0.79	0.77	0.90	0.82
CatBoost	0.78	0.76	0.90	0.81
XGB + RF	0.80	0.79	0.88	0.83

The results demonstrate distinct behavioral trends between deep sequence models and ensemble-based methods. Deep networks such as LSTM, CNN-LSTM, and BiLSTM achieved high recall (0.82–0.85), indicating sensitivity to subtle gait perturbations preceding FoG onset. However, these models suffered from relatively low precision (0.35–0.42), suggesting a tendency to overpredict FoG events, which is typical in highly imbalanced clinical datasets. In contrast, ensemble approaches exhibited better calibration between recall and precision. XGBoost achieved the most favorable trade-off (Precision = 0.77, Recall = 0.90, F_1 = 0.82), closely followed by CatBoost, underscoring the stability and interpretability of gradient-boosted tree models on structured sensor features. The hybrid XGB + RF ensemble outperformed all individual models (F_1 = 0.83), confirming the advantage of combining complementary decision boundaries for enhanced reliability and confidence. Overall, ensemble-based methods provided an optimal balance of safety (high recall) and usability (moderate precision), making them more suitable for continuous, real-time wearable deployment.

The detailed per-subject results for the XGBoost classifier are presented in Table 2, illustrating consistent predictive capability across diverse participants. Most subjects achieved F_1-scores above 0.80, indicating stable generalization despite inter-individual gait variability. Notably, subjects S01 and S02 achieved the highest balanced performance, while S08 demonstrated elevated recall at the expense of precision, highlighting expected variability in mild FoG presentations.

Table 2. Per-Subject LOSO-CV Performance of XGBoost

Subject	Accuracy	Precision	Recall	F1-Score
S01	0.8541	0.8171	0.9732	0.8883
S02	0.8602	0.8676	0.8848	0.8761
S03	0.7999	0.8102	0.7920	0.8010
S05	0.7574	0.7789	0.8632	0.8189
S06	0.7634	0.7495	0.8884	0.8131
S07	0.7840	0.7589	0.9441	0.8414
S08	0.6546	0.5111	0.9857	0.6732
S09	0.8150	0.8473	0.8707	0.8588

Per-subject analysis confirms that XGBoost generalizes effectively across heterogeneous gait patterns. Most participants achieved $F_1 > 0.80$, reflecting consistent identification of pre-FoG signals. The highest-performing subjects (S01, S02) exhibited balanced recall and precision, indicating stable gait biomarker representation. Cases like S07 and S08 revealed recall values above 0.94 and 0.98, respectively signifying excellent sensitivity to hesitation and instability. Subject S08's lower precision (0.51) illustrates an expected trade-off, where over-alerting is tolerated to avoid missed FoG events. From a clinical standpoint, this bias toward recall is desirable in fall-prevention contexts, prioritizing safety over minimal user inconvenience.

To further illustrate classifier reliability, Figure 2 depicts confusion matrices for representative subjects S01 and S02. The model achieved 1,849 true positives against only 51 false negatives, underscoring its strong recall and low miss rate for FoG episodes. The clear diagonal dominance in both matrices visually confirms accurate classification boundaries between normal and pre-FoG states.

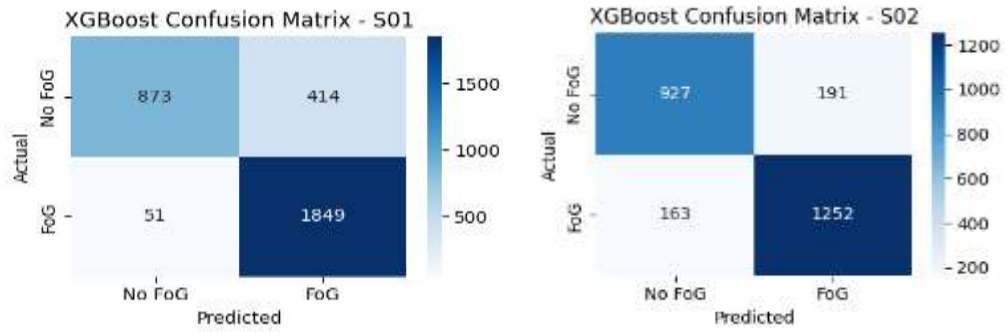

Fig.2. Confusion Matrices for S01 and S02

While prioritizing recall ensures safety, balanced precision is necessary for long-term usability. The XGBoost classifier achieved mean precision of 76.7% and recall of 90.0%, validating its ability to minimize false negatives while maintaining practical false-positive levels. Variability observed in milder cases suggests that finer-grained personalization or adaptive thresholding could further improve prediction stability.

The discriminative performance of the proposed framework was further evaluated using imbalance-aware metrics. Figure 3 illustrates the Precision–Recall (PR) curve, averaged across all LOSO folds. The XGBoost model achieved a mean PR-AUC of 0.88, confirming its strong ability to preserve precision while maintaining high recall, a desirable property in rare-event medical prediction tasks such as pre-FoG detection.

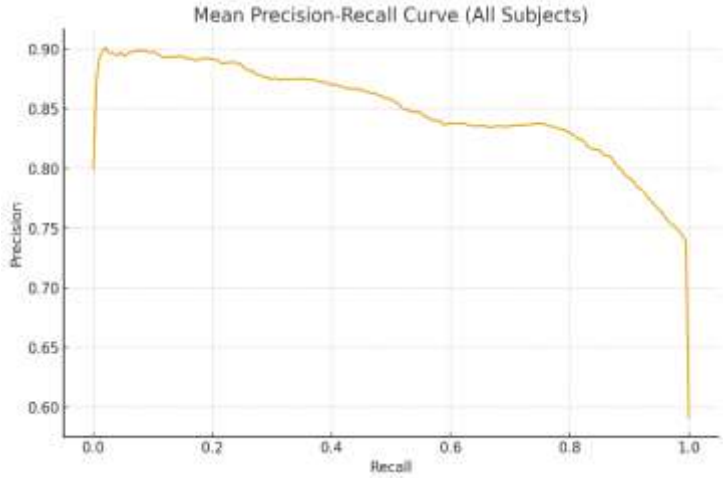

Fig. 3. Mean Precision–Recall (PR) curve of the XGBoost model across all LOSO folds

In addition, Figure 4 presents the Receiver Operating Characteristic (ROC) curves for all subjects. The model consistently yielded ROC-AUC values between 0.85 and 0.94, with a mean of 0.92, demonstrating robust separability between pre-FoG and normal gait instances across

heterogeneous gait profiles. The overlap of ROC curves across participants indicates generalization stability under subject-independent testing.

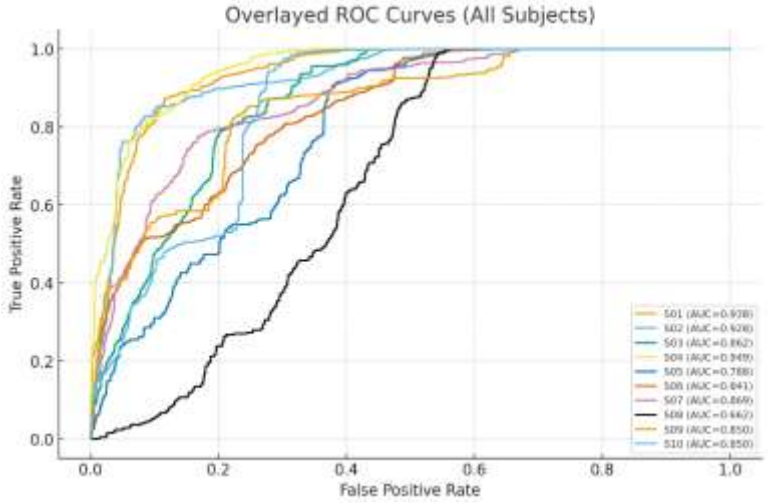

Fig. 4. Overlayed Receiver Operating Characteristic (ROC) curves for all subjects under LOSO-CV

These findings collectively confirm that the proposed system achieves a favorable balance between sensitivity and specificity, effectively identifying early FoG precursors under class imbalance without excessive false alarms.

Ablation Study

To assess the relative importance of the proposed biomechanical features and verify that each component meaningfully contributes to early FoG prediction, an ablation study was performed under the same LOSO-CV protocol used for the main experiments. Four model configurations were evaluated:

- Full feature set: all 74 engineered features, including time-domain statistics, Freeze Index, and zero-crossing metrics.
- Without Freeze Index: all features except the FoG-specific Freeze Index.
- Without zero-crossing metrics: all features except per-axis and global zero-crossing descriptors.
- Time-domain only: only basic statistical time-domain features retained; all FoG-specific biomarkers removed.

The quantitative results are reported in Table 3.

Table 3. Ablation Study Results

Configuration	Recall	Precision	F1-Score
Full feature set	0.90	0.77	0.82
Without Freeze Index	0.83	0.71	0.76
Without Zero-Crossing	0.86	0.72	0.79
Time-domain only	0.80	0.69	0.73

The full 74-feature configuration achieved the best overall performance (F1 = 0.82), confirming that the complete set of statistical and FoG-specific biomarkers provides the richest representation of pre-FoG gait dynamics. When the Freeze Index was removed, F1 dropped from 0.82 to 0.76 and recall decreased from 0.90 to 0.83. This substantial decline indicates that the Freeze Index is particularly important for correctly identifying positive pre-FoG windows, consistent with its physiological role in capturing elevated high-frequency oscillations associated with trembling-in-place behaviour. Similarly, removing zero-crossing metrics led to a reduction in F1 to 0.79 and recall to 0.86. Zero-crossing features are sensitive to rapid sign changes and irregular oscillations, and their removal reduces the model's ability to detect subtle gait rhythm disturbances and micro-hesitations that often precede FoG episodes.

The weakest performance was observed in the time-domain-only configuration (F1 = 0.73), where both recall (0.80) and precision (0.69) declined. This outcome demonstrates that generic statistical descriptors (mean, variance, skewness, etc.) alone are insufficient to fully characterize the complex neuromechanical transitions that occur before FoG; domain-informed biomarkers are necessary to distinguish true pre-FoG patterns from benign gait variability.

Overall, the ablation study provides clear evidence that physiologically motivated features such as the Freeze Index and zero-crossing-based measures play a critical role in enhancing sensitivity and robustness of early FoG prediction. Their contribution validates the feature-engineering strategy and justifies the use of lightweight, handcrafted biomarkers in computationally constrained wearable implementations.

Results Comparison with State-of-the-Art

In recent years, a growing body of work has focused on wearable-based detection and prediction of Freezing of Gait (FoG) in Parkinson's disease, using both handcrafted features and deep-learning models. A comparative summary of representative state-of-the-art (SOTA) methods is provided in Table 4.

Table 4. State-of-the-Art Comparison

Study	**Dataset / Subjects**	**Sensors & Placement**	**Task**	**Model Type**	**Eval Protocol**	**Key Metric (F_1 / Recall)**
Aich et al. [34]	~39 PD	Single accelerometer (waist)	Detection	CNN + handcrafted gait features	Independent split	~0.88 Accuracy
Pardoel et al. [20]	10 PD	3 sensors (ankle, thigh, trunk)	Detection & Prediction	RF, SVM (handcrafted features)	LOSO-CV	$F_1 \approx 0.83$
Salomon et al. [35]	Global Challenge (~80 PD)	Multi-sensor (waist, wrist, ankle)	Detection	Ensemble DL/ML (competition)	Unseen patient test	AUC > 0.90
This Work	10 PD (Daphnet)	3 sensors (ankle, thigh, trunk)	Early-warning (pre-FoG → FoG)	XGBoost + RF Ensemble (feature-engineered)	LOSO-CV	$F_1 = 0.82$ / Recall = 0.90

The comparative analysis highlights the progression of FoG research from reactive detection toward anticipatory prediction, emphasizing interpretability and computational efficiency. Aich et al. [34] achieved ~0.88 accuracy using a CNN-based detection model but focused only on post-onset FoG without addressing subject variability. Pardoel et al. [20] underscored the value of LOSO-CV for subject-independent testing, reporting $F_1 \approx 0.83$ with handcrafted gait features, though limited in temporal forecasting. Salomon et al. [35] benchmarked large-scale detection via ensemble deep learning (AUC > 0.90), achieving strong generalization yet lacking early-warning capability. In contrast, this study advances early FoG prediction using a feature-engineered XGBoost + RF ensemble, achieving Recall = 0.90, Precision = 0.77, and robust LOSO-CV performance. The inclusion of SHAP-based interpretability confirms the physiological relevance of key features such as the Freeze Index and zero-crossing rate. Overall, this work bridges critical gaps in the SOTA by enabling proactive forecasting, ensuring subject-independent generalization, and incorporating explainable AI for clinical trust and real-world deployment.

Explainability and Feature Importance Analysis

SHAP (SHapley Additive exPlanations) analysis was applied to interpret XGBoost predictions and rank feature contributions. Figure 5 presents a representative SHAP summary plot for Subject S01, highlighting the top 20 influential features. The Freeze Index consistently dominated, followed by thigh-axis energy and ankle-axis skewness, confirming alignment between data-driven importance and known biomechanical correlates of FoG. High SHAP values for these features corresponded to gait hesitation, asymmetry, and tremor, Key precursors to freezing.

The alignment between SHAP-derived importance and physiological plausibility reinforces the clinical interpretability of the proposed framework. By identifying early oscillatory disturbances and asymmetries, the model not only predicts FoG but also provides insights into neuromotor instability patterns, facilitating clinician trust and future translational applications.

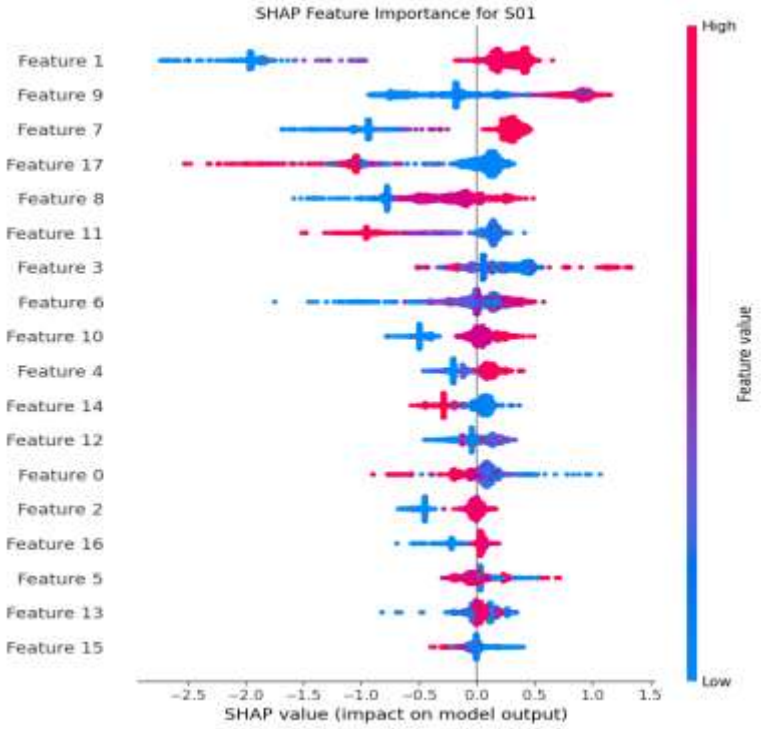

Fig. 5. SHAP summary plot showing top 20 influential features for subject S01 under XGBoost.

Limitations

Despite promising results, several limitations must be acknowledged. The dataset comprises only 10 subjects, limiting representativeness across disease severity and phenotypes. Though LOSO-CV mitigates overfitting, validation on larger multi-center cohorts is necessary. Additionally, data were collected in controlled settings; real-world conditions introduce sensor noise, environmental variability, and motion artifacts that may affect robustness. Clinical validation of SHAP interpretations with neurologist feedback remains pending, and real-time embedded deployment has yet to be benchmarked on low-power wearable hardware. Finally, FoG subtypes (trembling, akinetic, festination) were not explicitly modeled; future extensions should address subtype-specific prediction, adaptive personalization, and uncertainty-aware cueing mechanisms.

CONCLUSION AND FUTURE WORK

This study presented a machine learning–driven early prediction framework for Freezing of Gait (FoG) in Parkinson's disease using multi-sensor wearable accelerometery. By integrating a structured feature-engineering pipeline, a Leave-One-Subject-Out Cross-Validation (LOSO-CV)

protocol, and interpretable gradient-boosted models, the proposed system demonstrated that FoG events can be anticipated several seconds before onset with high reliability. The XGBoost classifier achieved a mean recall of 0.90 and an F_1-score of 0.82, outperforming deep sequence architectures while maintaining computational efficiency suitable for real-time wearable deployment. SHAP-based interpretability further revealed physiologically meaningful predictors, including the Freeze Index, step-cycle asymmetry, and axis-specific skewness, offering biomechanical insights into the mechanisms preceding FoG onset. These findings highlight the feasibility of proactive fall-risk mitigation and the potential of AI-assisted mobility support systems for neurodegenerative conditions.

Although the proposed framework achieved robust subject-independent generalization under LOSO-CV, several directions remain for future enhancement. Expanding the dataset to encompass a larger and more heterogeneous Parkinson's population across disease stages, medication states, and daily-living contexts would strengthen clinical robustness. Incorporating multimodal signals such as gyroscopes, surface EMG, kinematics, and speech biomarkers may further improve predictive fidelity and facilitate identification of distinct FoG subtypes. Real-time deployment studies on low-power wearable microcontrollers are also warranted to validate computational efficiency, latency, and energy consumption under operational constraints. Furthermore, adaptive personalization using continual or federated learning may enhance long-term usability by accommodating evolving gait characteristics and treatment responses.

Finally, clinical validation through collaboration with neurologists and physiotherapists, alongside prospective real-world trials and cueing interventions, will be crucial to translate algorithmic performance into therapeutic outcomes. Ethical and human-factor considerations, including user acceptance, false-alarm mitigation, and medical safety compliance, must guide future development. Overall, this work establishes a foundation for intelligent, interpretable, and resource-efficient FoG prediction systems aimed at improving mobility, reducing fall risk, and promoting independence among individuals with Parkinson's disease.

Acknowledgment
This work was carried out under the mentorship of Professor Patrick Benfeild of the University of Texas at Austin, College of Natural Sciences, United States, whose guidance, feedback, and encouragement were invaluable.

Statements and Declarations
Conflict of Interest: The Author has no competing interests related to the work submitted for publication.

Funding: No funding was received for this research.

Data Availability Statement: A publicly available Daphnet Freezing of Gait Dataset is used

Use of Generative AI: ChatGPT is used to assist with improving sentence ordering, reducing word count, and enhancing grammar. After using these tools, the authors meticulously reviewed and edited the content to ensure it met the required standards and take full responsibility for the final submission.

REFERENCES

1. Parkinson's Foundation. (2023). *Statistics on Parkinson's disease*. Retrieved from https://www.parkinson.org/Understanding-Parkinsons/Statistics
2. Dauer, W., & Przedborski, S. (2003). Parkinson's disease: Mechanisms and models. *Neuron, 39*(6), 889–909.
3. Nutt, J. G., Bloem, B. R., Giladi, N., Hallett, M., Horak, F. B., & Nieuwboer, A. (2011). Freezing of gait: Moving forward on a mysterious clinical phenomenon. *The Lancet Neurology, 10*(8), 734–744.
4. Bloem, B. R., Hausdorff, J. M., Visser, J. E., & Giladi, N. (2004). Falls and freezing of gait in Parkinson's disease: A review of two interconnected, episodic phenomena. *Movement Disorders, 19*(8), 871–884.
5. Macht, M., Kaussner, Y., Möller, J. C., Stiasny-Kolster, K., Eggert, K. M., Krüger, H. P., & Ellgring, H. (2007). Predictors of freezing in Parkinson's disease: a survey of 6,620 patients. *Movement disorders*, *22*(7), 953-956.
6. Nonnekes, J., Snijders, A. H., Nutt, J. G., Deuschl, G., Giladi, N., & Bloem, B. R. (2015). Freezing of gait: a practical approach to management. *The Lancet Neurology*, *14*(7), 768-778.
7. Nieuwboer, A., Rochester, L., Müncks, L., & Swinnen, S. P. (2009). Motor learning in Parkinson's disease: limitations and potential for rehabilitation. *Parkinsonism & related disorders*, *15*, S53-S58.
8. Saboor, A., Kask, T., Kuusik, A., Alam, M. M., Le Moullec, Y., Niazi, I. K., ... & Ahmad, R. (2020). Latest research trends in gait analysis using wearable sensors and machine learning: A systematic review. *Ieee Access*, *8*, 167830-167864.
9. Prasanth, H., Caban, M., Keller, U., Courtine, G., Ijspeert, A., Vallery, H., & Von Zitzewitz, J. (2021). Wearable sensor-based real-time gait detection: A systematic review. *Sensors*, *21*(8), 2727.
10. Gregorčič, M., & Georgiev, D. (2025). The Usefulness of Wearable Sensors for Detecting Freezing of Gait in Parkinson's Disease: A Systematic Review. *Sensors*, *25*(16), 5101.
11. 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*(12), 3959.
12. Al-Adhaileh, M. H., Wadood, A., Aldhyani, T. H., Khan, S., Uddin, M. I., & Al-Nefaie, A. H. (2025). Deep learning techniques for detecting freezing of gait episodes in Parkinson's disease using wearable sensors. *Frontiers in Physiology*, *16*, 1581699.
13. Jiao, Y., Liu, Z., Li, J., Su, Y., & Chen, X. (2024). Knowledge mapping of freezing of gait in Parkinson's disease: a bibliometric analysis. *Frontiers in Neuroscience*, *18*, 1388326.

14. Roggen, P. M., Daniel and Hausdorff, j. (2013). Daphnet Freezing of Gait. *UCI Machine Learning Repository*.
15. Ezhilarasi, J., & Senthil Kumar, T. (2025). Develop a novel, faster mask region-based convolutional neural network model with leave-one-subject-out to predict freezing of gait abnormalities of Parkinson's disease. *Neural Computing and Applications*, *37*(7), 5441-5457.
16. Mancini, M., McKay, J. L., Cockx, H., D'Cruz, N., Esper, C. D., Filtjens, B., ... & ICFOG Investigators. (2025). Technology for measuring freezing of gait: Current state of the art and recommendations. *Journal of Parkinson's Disease*, *15*(1), 19-40.
17. Moore, S. T., MacDougall, H. G., & Ondo, W. G. (2008). Ambulatory monitoring of freezing of gait in Parkinson's disease. *Journal of neuroscience methods*, *167*(2), 340-348.
18. Bachlin, M., Plotnik, M., Roggen, D., Maidan, I., Hausdorff, J. M., Giladi, N., & Troster, G. (2009). Wearable assistant for Parkinson's disease patients with the freezing of gait symptom. *IEEE Transactions on Information Technology in Biomedicine*, *14*(2), 436-446.
19. Silva de Lima, A. L., Evers, L. J., Hahn, T., Bataille, L., Hamilton, J. L., Little, M. A., ... & Faber, M. J. (2017). Freezing of gait and fall detection in Parkinson's disease using wearable sensors: a systematic review. *Journal of neurology*, *264*(8), 1642-1654.
20. Pardoel, S., Kofman, J., Nantel, J., & Lemaire, E. D. (2019). Wearable-sensor-based detection and prediction of freezing of gait in Parkinson's disease: a review. *Sensors*, *19*(23), 5141.
21. Borzì, L., Olmo, G., Artusi, C., & Lopiano, L. (2020, July). Detection of freezing of gait in people with Parkinson's disease using smartphones. In *Proceedings of the IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC)* (pp. 625–635).
22. San-Segundo, R., Navarro-Hellín, H., Torres-Sánchez, R., Hodgins, J., & De la Torre, F. (2019). Increasing robustness in the detection of freezing of gait in Parkinson's disease. *Electronics, 8*(2), 119.
23. 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*(2), 614.
24. Nurmi, J., & Lohan, E. S. (2021). Systematic review on machine-learning algorithms used in wearable-based eHealth data analysis. *IEEE Access*, *9*, 112221-112235.
25. Shi, B., Tay, A., Au, W. L., Tan, D. M., Chia, N. S., & Yen, S. C. (2022). Detection of freezing of gait using convolutional neural networks and data from lower limb motion sensors. *IEEE Transactions on Biomedical Engineering*, *69*(7), 2256-2267.
26. Li, B., Yao, Z., Wang, J., Wang, S., Yang, X., & Sun, Y. (2020). Improved deep learning technique to detect freezing of gait in Parkinson's disease based on wearable sensors. *Electronics*, *9*(11), 1919.
27. Ashour, A. S., El-Attar, A., Dey, N., Abd El-Kader, H., & Abd El-Naby, M. M. (2020). Long short term memory based patient-dependent model for FOG detection in Parkinson's disease. *Pattern recognition letters*, *131*, 23-29.

28. Bikias, T., Iakovakis, D., Hadjidimitriou, S., Charisis, V., & Hadjileontiadis, L. J. (2021). DeepFoG: an IMU-based detection of freezing of gait episodes in Parkinson's disease patients via deep learning. *Frontiers in Robotics and AI*, *8*, 537384.
29. Sigcha, L., Borzì, L., Pavón, I., Costa, N., Costa, S., Arezes, P., ... & De Arcas, G. (2022). Improvement of Performance in Freezing of Gait detection in Parkinson's Disease using Transformer networks and a single waist-worn triaxial accelerometer. *Engineering Applications of Artificial Intelligence*, *116*, 105482.
30. Noor, M. H. M., Nazir, A., Ab Wahab, M. N., & Ling, J. O. Y. (2021). Detection of freezing of gait using unsupervised convolutional denoising autoencoder. *IEEE Access*, *9*, 115700-115709.
31. Camps, J., Sama, A., Martin, M., Rodriguez-Martin, D., Perez-Lopez, C., Arostegui, J. M. M., ... & Rodriguez-Molinero, A. (2018). Deep learning for freezing of gait detection in Parkinson's disease patients in their homes using a waist-worn inertial measurement unit. *Knowledge-Based Systems*, *139*, 119-131.
32. Bentéjac, C., Csörgő, A., & Martínez-Muñoz, G. (2021). A comparative analysis of gradient boosting algorithms. *Artificial Intelligence Review*, *54*(3), 1937-1967.
33. Salih, A. M., Raisi-Estabragh, Z., Galazzo, I. B., Radeva, P., Petersen, S. E., Lekadir, K., & Menegaz, G. (2025). A perspective on explainable artificial intelligence methods: SHAP and LIME. *Advanced Intelligent Systems*, *7*(1), 2400304.
34. Aich, S., Pradhan, P. M., Park, J., Sethi, N., Vathsa, V. S. S., & Kim, H. C. (2018). A validation study of freezing of gait (FoG) detection and machine-learning-based FoG prediction using estimated gait characteristics with a wearable accelerometer. *Sensors*, *18*(10), 3287.
35. Salomon, A., Gazit, E., Ginis, P., Urazalinov, B., Takoi, H., Yamaguchi, T., ... & Hausdorff, J. M. (2024). A machine learning contest enhances automated freezing of gait detection and reveals time-of-day effects. *Nature Communications*, *15*(1), 4853.