



Prediction of Freezing of Gait in Parkinson's Disease Using a Random Forest Model Based on an Orthogonal Experimental Design: A Pilot Study

Zhanelue Chen¹, Gen Li², Chao Gao², Yuyan Tan², Jun Liu², Jin Zhao³, Yun Ling¹, Xiaoliu Yu¹, Kang Ren^{1*} and Shengdi Chen^{2*}

OPEN ACCESS

Edited by:

Guohui Lu,
The First Affiliated Hospital
of Nanchang University, China

Reviewed by:

Xiuyun Liu,
Johns Hopkins University,
United States
Changqin Quan,
Kobe University, Japan
Tetsuya Asakawa,
Fujian University of Traditional Chinese
Medicine, China

*Correspondence:

Kang Ren
renkang@gyenno.com
Shengdi Chen
chensd@rjh.com.cn

Specialty section:

This article was submitted to
Motor Neuroscience,
a section of the journal
Frontiers in Human Neuroscience

Received: 01 December 2020

Accepted: 02 February 2021

Published: 22 March 2021

Citation:

Chen Z, Li G, Gao C, Tan Y, Liu J,
Zhao J, Ling Y, Yu X, Ren K and
Chen SD (2021) Prediction
of Freezing of Gait in Parkinson's
Disease Using a Random Forest
Model Based on an Orthogonal
Experimental Design: A Pilot Study.
Front. Hum. Neurosci. 15:636414.
doi: 10.3389/fnhum.2021.636414

¹ Gyenno Science Co., Ltd., Shenzhen, China, ² Department of Neurology, Ruijin Hospital Affiliated to Shanghai Jiao Tong University School of Medicine and Center for Excellence in Brain Science and Intelligence Technology, Chinese Academy of Sciences, Shanghai, China, ³ School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan, China

Purpose: The purpose of this study was to introduce an orthogonal experimental design (OED) to improve the efficiency of building and optimizing models for freezing of gait (FOG) prediction.

Methods: A random forest (RF) model was developed to predict FOG by using acceleration signals and angular velocity signals to recognize possible precursor signs of FOG (preFOG). An OED was introduced to optimize the feature extraction parameters.

Results: The main effects and interaction among the feature extraction hyperparameters were analyzed. The false-positive rate, hit rate, and mean prediction time (MPT) were 27%, 68%, and 2.99 s, respectively.

Conclusion: The OED was an effective method for analyzing the main effects and interactions among the feature extraction parameters. It was also beneficial for optimizing the feature extraction parameters of the FOG prediction model.

Keywords: orthogonal experimental design, optimization, freezing of gait, Parkinson's disease, fog prediction

HIGHLIGHTS

- A novel method was developed to predict FOG in PD.
- An OED was first used to obtain the optimal feature extraction parameters.
- The main effects of the feature extraction parameters were analyzed first.
- Interactions among the feature extraction parameters were analyzed.

Abbreviations: ECG, electrocardiogram; FOG, freezing of gait; MPT, mean prediction time; MI, mutual information; ML, machine learning; OED, orthogonal experimental design; preFOG, prefreezing of gait; PD, Parkinson's disease; RF, random forest; SCR, skin-conductance response.

INTRODUCTION

Parkinson's disease (PD) is a neurodegenerative disease characterized by the degeneration of dopaminergic neurons of the substantia nigra resulting in bradykinesia, rigidity, tremor, and postural instability (Revuelta et al., 2015; Li et al., 2019). Over 50% of the PD patients who have lived with the disease for more than 10 years are affected by freezing of gait (FOG) (Okuma, 2006). In addition, FOG may occur in 25% of patients with early PD (Moore et al., 2008). FOG is defined as not being able to start or continue walking and feeling as if the feet have been "glued" or "magnetized" to the ground (Handojoseno et al., 2012). This condition causes both physical and psychological distress. Approximately 60% of PD patients experience falls each year, and they also have to endure multiple falls caused by complications and fall-related injuries (Wood et al., 2002; Moore et al., 2008; Allcock et al., 2009; Latt et al., 2009; Contreras and Grandas, 2012).

The response of FOG in PD to pharmaceutical treatment is limited (Okuma, 2006). However, previous works have shown that cueing-based training has specific effects on gait, freezing, and balance (Lim et al., 2005). Many works have aimed to develop small wearable devices that can detect FOG episodes (Han et al., 2003; Moore et al., 2008, 2013; Bachlin et al., 2009; Jovanov et al., 2009; Mazilu et al., 2012; Pepa et al., 2015; Zach et al., 2015). Gait recognition is a gait assessment tool that uses a machine learning (ML) algorithm. Researchers have attempted to develop a wearable system to improve the accuracy of identifying FOG events. To optimize the system, multiple aspects were investigated, including different sensors and their locations, a subset of the features extracted, different gait recognition models, and hyperparameters. Bächlin et al. (2009) used a threshold algorithm based on the freezing index and signal power to recognize FOG. Their model was more sensitive when sensors were worn on the hips. Mazilu et al. (2012) showed that the mean latency of FOG detection increased linearly with the window size, but the classification performance increased quickly and then stabilized. Using an ML algorithm such as random forest (RF), a single sensor was sufficient for FOG detection, and the placement of this sensor had little effect on the detection. Ensemble methods such as boosting and bagging are more appropriate than other methods (Mazilu et al., 2012). Mikos et al. (2017) found that features maximized their mutual information (MI) at different window lengths. The efficient parameter optimization method in gait recognition has not been fully studied. We found that when the dataset is very large, optimizing the hyperparameters is a very time-consuming process, so an efficient parameter optimization method needs to be discussed. Some works have studied the effects of feature extraction parameters, while the interaction among parameters has not received enough attention. We propose a statistical experimental design, i.e., an orthogonal experimental design (OED), to optimize the hyperparameters and analyzes the interaction.

Furthermore, by utilizing such a wearable system, timely rhythmic cues can be provided after FOG is detected. These systems have advantages in shortening the duration of FOG and, by providing immediate (at least hundreds of milliseconds)

rhythmic cues as a response to the FOG signal, will further improve these systems. Therefore, FOG prediction is necessary to solve the response problem and further avoid intervention failure. An unsupervised feature learning decision tree can be used to perform FOG prediction (Mazilu et al., 2013). Mazilu et al. (2015b) explored the association of gait with electrocardiogram (ECG) and skin-conductance response (SCR) features and then fit the data to a multivariate Gaussian distribution, which can be used for FOG detection. In previous studies, parameters such as the window size and the defined duration of precursor signs of FOG (preFOG; all the references to preFOG in this article mean preFOG in PD) are based on experience (Mazilu et al., 2013, 2015b). However, the principles of objectification, multipurpose use, and simplification (OMS) have been the trend in the development of a novel behavioral assessment for PD (Asakawa et al., 2016a,b, 2019).

In this study, an RF model evaluated with an episode-based strategy was developed to predict upcoming FOG. This model can be adopted by wearable devices to activate an early intervention to avoid some of the FOG episodes.

MATERIALS AND METHODS

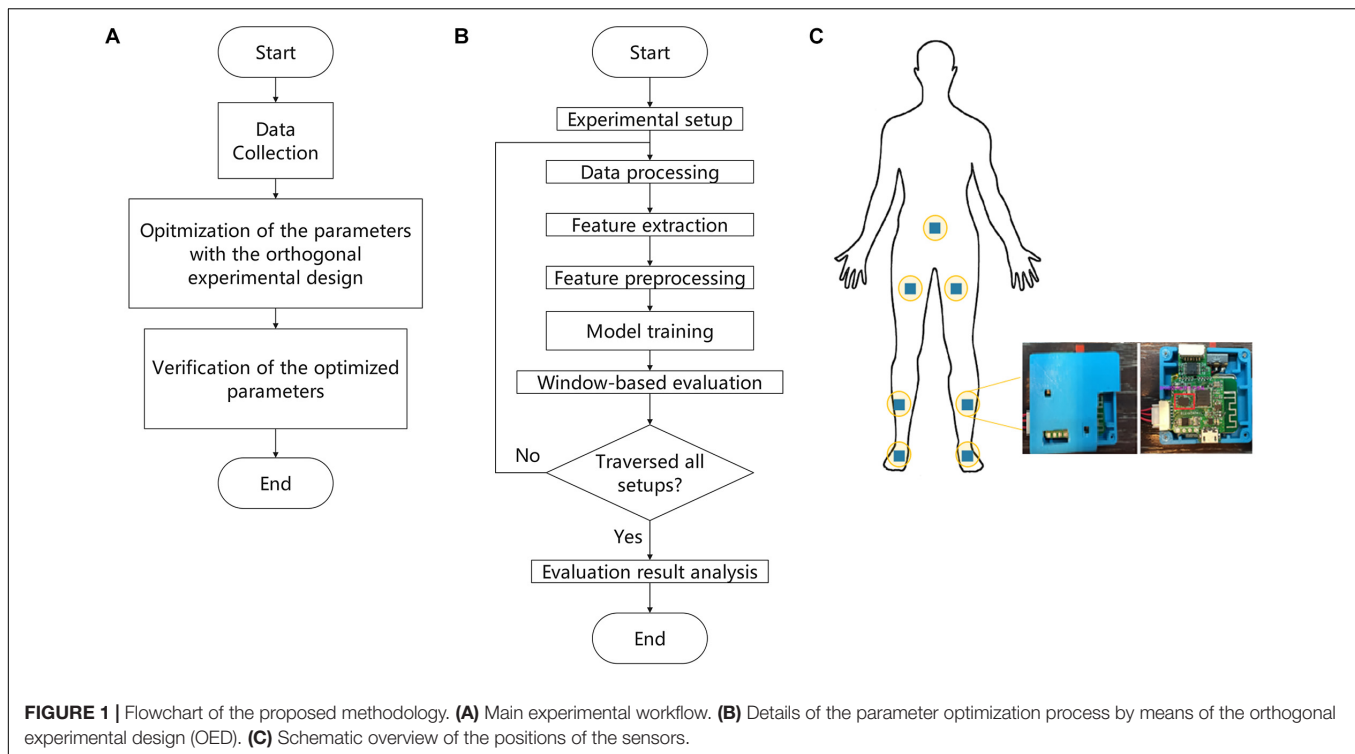
Methodology Overview

The proposed methodology for building a model to predict FOG based on OED consists of three parts, as shown in **Figure 1A**: (a) data collection, (b) optimization of the parameters with the OED, and (c) verification of the optimized parameters. This article was focused on optimizing feature extraction parameters with the OED, and the details included seven steps, as shown in **Figure 1B**: (a) experimental setup, (b) data processing, (c) feature extraction, (d) feature preprocessing, (e) model training, (f) window-based evaluation, and (g) evaluation results analysis.

Data Collection

The study participants were diagnosed with PD by movement disorder specialists according to the Movement Disorder Society (MDS) diagnostic criteria (Postuma et al., 2015). The eligibility criteria for the participants were a Hoehn-Yahr stage between 2 and 3 in FOG OFF state, not having cognitive dysfunction according to the Mini-Mental State Examination (MMSE), not having serious vision or hearing impairment, and not having any disease affecting walking ability. Participants were excluded if they had secondary PD causes, such as inflammatory, drug-induced, vascular, and toxin-induced Parkinsonism. Participants with other neurodegenerative diseases, such as progressive supranuclear palsy and multiple system atrophy, were also excluded. All the participants were familiar with the process and signed consent forms. Patients with severe PD (Hoehn-Yahr stage > 3.0) were not included due to safety issues. The characteristics of the patients are presented in **Table 1**. This study was approved by the ethics committee of the Ruijin Hospital, affiliated with the Shanghai Jiao Tong University School of Medicine.

Data were collected from 24 participants in a laboratory setting designed to provoke FOG with a walking protocol,



which included gait initiation, walking with 360- and 180-degree turns, walking in straight lines, passing narrow corridors, and walking through the crowded hospital halls between December 2016 and April 2017 (Mazilu et al., 2015a). Two nurses accompanied the participants during all the test procedures to prevent falls. The inertial measurement unit (IMU; BMI160, Bosch, Germany) generated nine signals sampled at 100 Hz as output. The nine signals represented the measurements of triaxial sensors: an accelerometer with sensitivity of 4,096 least significant bit (LSB)/g, a gyroscope with sensitivity of 16.4 LSB/deg/s, and a magnetometer. The data collection system contained seven wearable IMUs attached to different parts of the body (**Figure 1C**). Moreover, the participants underwent FOG evaluation with the sensors in all procedures. The entire experiment was recorded on video with an iPhone 6s Plus

(A1699, Apple Inc., CA, United States), which was aligned with the signals on the timeline.

Currently, there is no detailed diagnostic criteria of FOG. To make it accurate, videos describing FOG in terms of the MDS-Sponsored Revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS) score were adopted. The FOG episodes were labeled offline by two independent gait experts who were blinded with respect to group allocation. If the labels of videos were inconsistent, the raters discussed labeling the gait as either FOG or not FOG. In total, 88 non-FOG episodes and 89 FOG episodes were captured. A FOG gait sequence in a patient video is shown as an example in **Figure 2**. The clinicians also labeled the start of other walking events, i.e., gait initiation, turns, and stops.

Twenty-four patients with FOG volunteered for this study. However, videos were excluded if the patient was blocked during recording and if the sensors fell off. Finally, 14 videos were included. The average inter-rater reliability was 0.928.

TABLE 1 | Characteristics of the included patients.

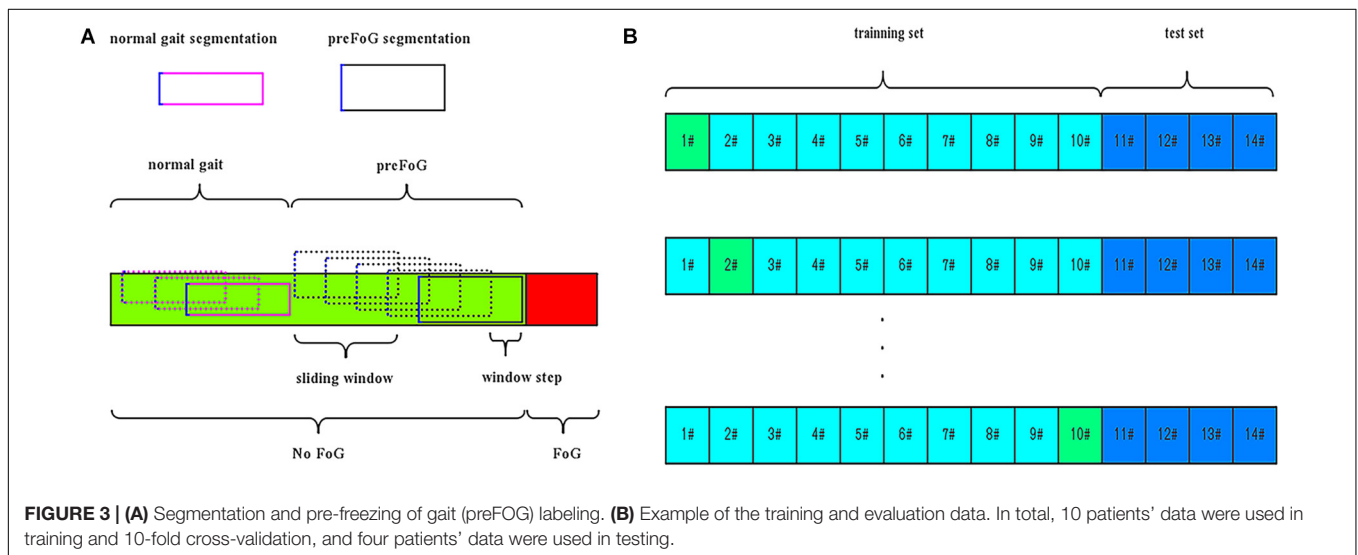
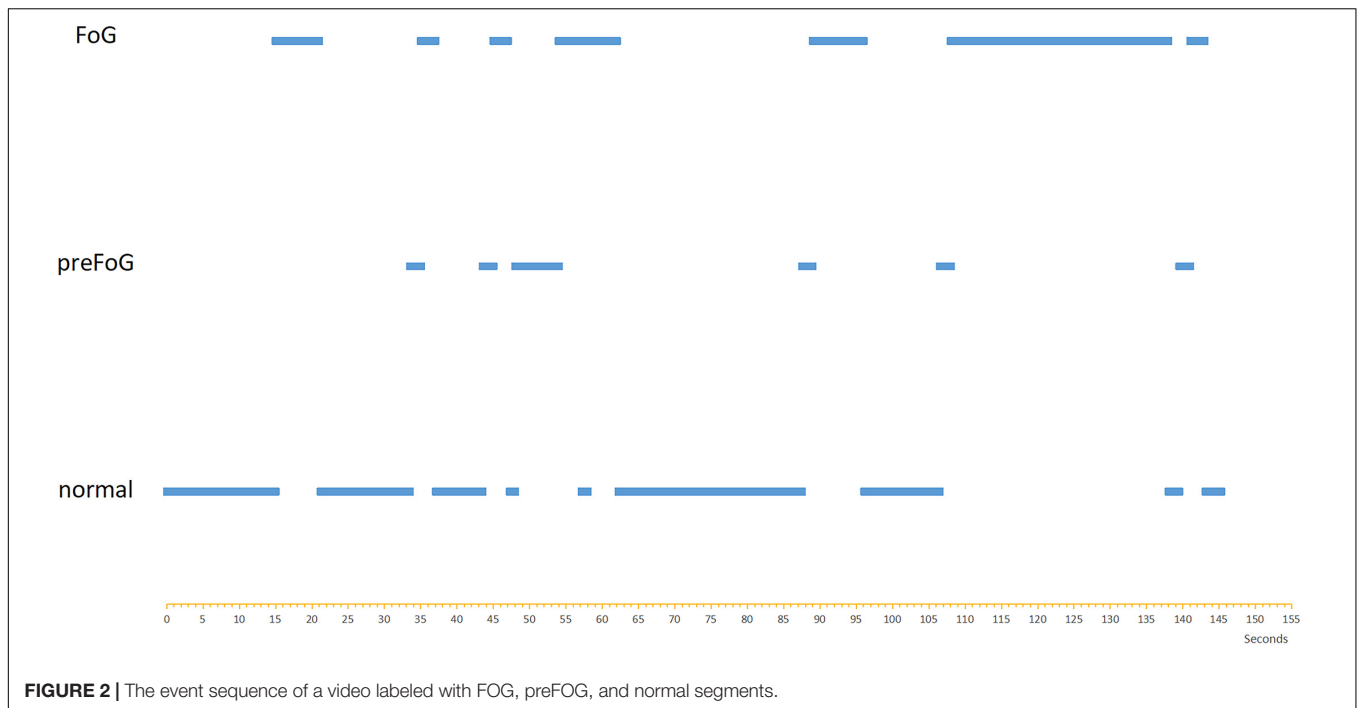
	Male	Female	<i>p</i> -value
Number (%)	8 (57.14)	6 (42.86)	–
Age, years, mean (SD)	71.83 (11.67)	69.20 (5.89)	0.642
Age of Onset, years, mean (SD)	5.00 (2.61)	6.40 (4.34)	0.549
Hoehn–Yahr Stage, <i>N</i> (%)			1.000
	3 (37.50)	2 (33.33)	
	3 (37.50)	2 (33.33)	
	2 (25.00)	2 (33.33)	
MDS–UPDRS score, mean (SD)	51.17 (5.56)	52.00 (7.81)	0.847
FOGQ score, mean (SD)	8.83 (1.17)	8.80 (0.84)	0.957

FOGQ, Freezing of Gait Questionnaire; MDS, movement disorders society; SD, standard deviation; UPDRS, Unified Parkinson's Disease Rating Scale.

Data Preparation

In addition to FOG and normal locomotion, the walking periods before FOG episodes were considered a third class called preFOG. It was hypothesized that there was a detectable deterioration of gait in this phase that precedes FOG (Mazilu et al., 2013). Different durations of the preFOG episodes were assumed. PreFOG episodes can be retrieved only through data mining from segments of data preceding FOG episodes, as shown in **Figure 3A**.

The accelerometer signals used in this study had burst outliers either larger than the 97.5th percentile or smaller than the 2.5th percentile. These outliers were replaced with the median value



of the whole time series (Xia et al., 2018). The signals were also detrended by a high-pass filter.

The video and inertial signals were synchronized. The timestamps of the sensors were shown on the screen and were recorded in the first frame of the video. The signal data labeled with FOG were removed. To prepare the data instances for feature extraction, a sliding window was used to segment the whole time series into many overlapping data slices. The window size, which determines the length of the data segment, was fixed in advance in previous works. For the step size of the sliding window, it was evident that a smaller value of this parameter can generate more data instances. The signal data were segmented into slices by the means, as shown in **Figure 3A**.

According to statistical and digital signal processing knowledge and previous research (Zhang, 2017), the information on the raw signal features is listed in **Table 2**; in total, 924 (seven IMU sensors \times 3D accelerometer and 3D gyroscope \times 22 features) features were obtained (Pham et al., 2017). These features were then used to train a classification model. Due to the large difference in the scales of the data, the extracted feature data were standardized for model training.

Experimental Setup

The window size, sliding step length, and preFOG duration involved in the extraction of features can be optimized to enhance the classifier's performance. The window size, sliding step length,

TABLE 2 | Information on the features extracted from the signal.

Expression	Remarks	Expression	Remarks
$F_1 = \max(x)$	Maximum value of the signal	$F_{12} = \frac{x_{max}}{\sqrt{\frac{1}{N_s} \sum_{i=1}^{N_s} x(i)^2}}$	Crest factor
$F_2 = \min(x)$	Minimum value of the signal	$F_{13} = \frac{\bar{x}}{\sqrt{\frac{1}{N_s} \sum_{i=1}^{N_s} (x(i) - \bar{x})^2}}$	Reciprocal coefficient of variation
$F_3 = \frac{1}{N_s} \sum_{i=1}^{N_s} x(i) $	Mean of the absolute value of the signal	$F_{14} = \frac{1}{N_s} \sum_{j=1}^{N_s} \left[\frac{x(j)}{\sqrt{\frac{1}{N_s} \sum_{i=1}^{N_s} x(i) - \bar{x} ^2}} \right]^3$	Skewness coefficient
$F_4 = \max(x) - \min(x)$	Signal range	$F_{15} = \frac{1}{N_s} \sum_{j=1}^{N_s} \left[\frac{x(j)}{\sqrt{\frac{1}{N_s} \sum_{i=1}^{N_s} x(i) - \bar{x} ^2}} \right]^4$	Kurtosis coefficient
$F_5 = \sqrt{\frac{1}{N_s} \sum_{i=1}^{N_s} x(i)^2}$	Root mean square	$F_{16} = \frac{x_{max}}{\frac{1}{N_s} \sum_{i=1}^{N_s} x(i)^2}$	Clearance factor
$F_6 = \frac{1}{N_s} \sum_{i=1}^{N_s} x(i)$	Mean of the signal	$F_{17} = \frac{x_{max}}{\frac{1}{N_s} \sum_{i=1}^{N_s} x(i) }$	Impulse factor
$F_7 = \sqrt{\frac{1}{N_s} \sum_{i=1}^{N_s} (x(i) - \bar{x})^2}$	Standard deviation	$F_{18} = \frac{\frac{1}{N_s} \sum_{i=1}^{N_s} (\Delta x(i) - \Delta \bar{x})^4}{[\frac{1}{N_s} \sum_{i=1}^{N_s} (\Delta x(i) - \Delta \bar{x})^2]^2}$	Energy operator $\Delta \bar{x} \frac{1}{N_s} \sum_{i=1}^{N_s} \Delta x(i)$ if the data point is not an endpoint, $\Delta x(i) x(i)^2 - x(i+1) \times x(i-1)$ else, $\Delta x(i1) x(1)^2 - x(2) \times x(N_s)$ $\Delta x(N_s) x(N_s)^2 - x(1) \times x(N_s - 1)$
$F_8 = \frac{\frac{1}{N_s} \sum_{i=1}^{N_s} (x(i) - \bar{x})^3}{[\frac{1}{N_s} \sum_{i=1}^{N_s} (x(i) - \bar{x})^2]^3}$	Skewness depicts the symmetry of the signal distribution	$F_{19} = \frac{1}{N_{mf}} \sum_{j=1}^{N_{mf}} X(j)$	Mean frequency $X(j)$: amplitude at the corresponding frequency
$F_9 = \frac{\frac{1}{N_s} \sum_{i=1}^{N_s} (x(i) - \bar{x})^4}{[\frac{1}{N_s} \sum_{i=1}^{N_s} (x(i) - \bar{x})^2]^4}$	Kurtosis depicts the steepness of the signal distribution	$F_{20} = \frac{\sum_{j=1}^{N_{mf}} f(j) \times X(j)}{\sum_{j=1}^{N_{mf}} X(j)}$	Center frequency $f(j)$: frequency $X(j)$: amplitude at the corresponding frequency
$F_{10} = \frac{1}{N_s} \sum_{i=1}^{N_s} (x(i) - \bar{x})^2$	Variance of the signal	$F_{21} = \sqrt{\frac{\sum_{j=1}^{N_{mf}} f(j)^2 \times X(j)}{\sum_{j=1}^{N_{mf}} X(j)}}$	Root mean square of the frequency $f(j)$: frequency $X(j)$: amplitude at the corresponding frequency
$F_{11} = \frac{\sqrt{\frac{1}{N_s} \sum_{i=1}^{N_s} x(i) ^2}}{\frac{1}{N_s} \sum_{i=1}^{N_s} x(i) }$	Waveform factor	$F_{22} = \sqrt{\frac{\sum_{j=1}^{N_{mf}} (f(j) - f_{FC})^2 \times X(j)}{\sum_{j=1}^{N_{mf}} X(j)}}$	$f(j)$: frequency $X(j)$: amplitude at the corresponding frequency $f_{FC}: F_{20}$

N_s , length of the signal; \bar{x} , mean of the signal; x_{max} , maximum signal; N_{mf} , length of the spectrum signal.

and preFOG duration were based on the sensor’s sampling rate. Each unit measurement means one sampling point, and the duration is 10 ms. The feature selection hyperparameters and the RF classifier were not the focus of this study. The Taguchi OED and RF were introduced into this study to build the models.

The OED is a type of general fractional factorial design (Cavazzuti, 2012). It is based on a design matrix proposed by Genichi Taguchi and allows the consideration of a selected subset of combinations of multiple factors at multiple levels. Taguchi orthogonal arrays (OAs) are balanced to ensure that all levels of all factors are considered equally. For this reason, the factors can be evaluated independently of each other despite the fractionality of the design. In the Taguchi OA design, only the main effects and two-factor interactions are considered, and higher-order interactions are assumed to be non-existent. In addition, designers are asked to identify (based on their knowledge of the subject matter) which interactions might be significant before conducting the experiment. The full factorial design of the three factors (window size, step length, and preFOG duration) with four levels consisted of 64 runs, while an $L_{16}(4^3)$ OED scheme was chosen based on the OED. The different levels of window size, step length, and preFOG duration were chosen according to previous works. The details of the $L_{16}(4^3)$ scheme are shown in **Table 3** and visualized in **Figure 4**.

RF is an ensemble ML algorithm that combines a number of classification or regression trees and is based on the bagging technique. The RF algorithm is a powerful model

TABLE 3 | Detailed orthogonal experimental design (OED) for optimizing the feature extraction parameters.

DOE name	ID	Window size	Step	PreFOG duration
T01	1	128	5	150
T02	2	128	10	250
T03	3	128	20	500
T04	4	128	30	600
T05	5	256	5	250
T06	6	256	10	150
T07	7	256	20	600
T08	8	256	30	500
T09	9	400	5	500
T10	10	400	10	600
T11	11	400	20	150
T12	12	400	30	250
T13	13	500	5	600
T14	14	500	10	500
T15	15	500	20	250
T16	16	500	30	150

DOE, Design of experiment; preFOG, pre-freezing of gait.

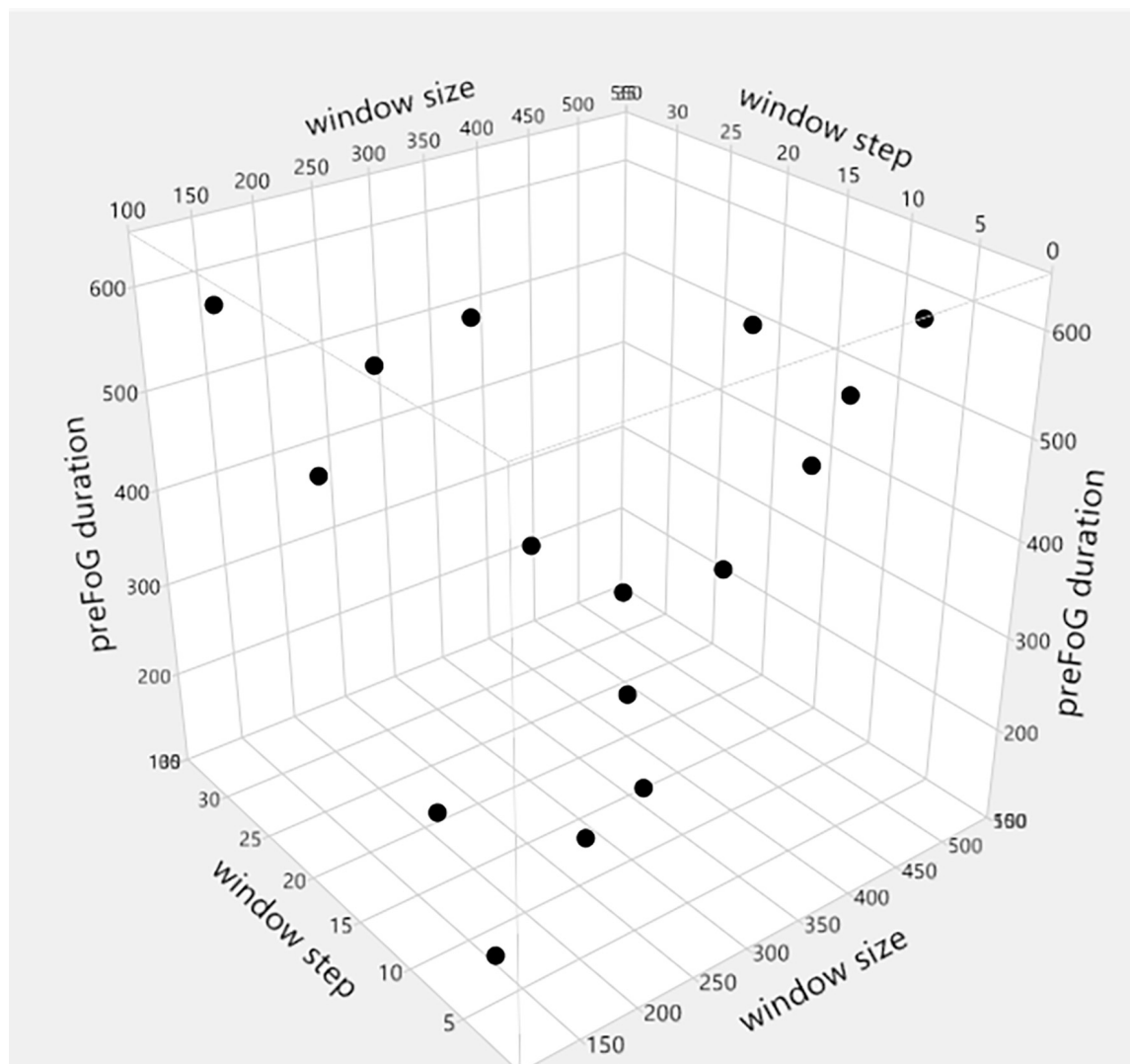


FIGURE 4 | $L_{16}(4^3)$ orthogonal experimental design (OED).

and is used in many studies for the classification of FOG and other activity-related problems (Orphanidou et al., 2018). Common hyperparameters determine the accuracy of the RF classifier, such as the number of estimators (`n_estimators`), the maximum number or ratio of features (`max_features`), the maximum depth (`max_depth`), the minimum number of samples (`min_samples_split`), and the minimum number of samples needed by each leaf node (`min_samples_leaf`).

A set of parameters for feature extraction was fetched line by line from **Table 3**, and then the data were prepared according to the procedures shown in **Figure 1B**. The RF model was trained with the feature data and evaluated with the window-based strategy, and the evaluation results were stored in a table. The feature extraction parameters were optimized, and their main effects and interactions were analyzed.

After analysis, the optimized parameters were verified. The feature data were prepared according to the procedure described

in the “Data Preparation” section with the optimized feature extraction parameters. A subset of features was selected to train the RF model according to the work by Zhang and Sawchuk (2011). The top-K method generated feature subsets, and the best subset was selected out according to the F1 score and kappa value. The prepared data were used to train the RF model, which was then evaluated with the episode-based strategy.

Evaluation and Metrics

The preFOG prediction model was evaluated using leave-one-patient-out cross-validation, which meant a patient-independent evaluation. The data were split into a training set and a test set. The RF classifier was trained on feature data selected from $N - 1$ patients in the training set and evaluated with the data from the remaining patient, and some patients’ data were saved as the test set to avoid overfitting, as shown in **Figure 3B**. A window-based strategy was used for parameter optimization,

and an episode-based strategy was used for verification of the optimized parameters.

The F1 score was adopted to evaluate the performance of the ML model with the setting configuration in the experiment. True positives were defined as gaits correctly classified by our method. False positives were defined as gaits that were identified as a certain class type but were found to be false from the video record. False negatives were gaits that were not identified as a certain class type, but the video record agreed with that class type. True negatives were those gaits for which both the applied method and the video agreed on the classification of not a certain class type (Djurić-Jovičić et al., 2013).

$$F_1 - score = \frac{2True\ Positives}{2TruePositives + FalseNegatives + FalsePositives}$$

Statistics Analysis

The Python (version 3.6.4) package scikit-learn (version 0.19.2) was used for the ML step. A trial version of JMP (version 13.2.0) and Minitab (version 18) were used to design the experimental scheme and to analyze the collected data.

The associated p -values of large effects are often very small. Visualizing these small values graphically can be challenging. When transformed to the LogWorth [$-\log_{10}(p\text{-value})$] scale, highly significant p -values have large LogWorth values and non-significant p -values have small LogWorth values. A LogWorth of zero corresponds to a non-significant p -value of 1. Any LogWorth above two corresponds to a p -value less than 0.01.

Interaction effects occur when the effect of one variable depends on the value of another variable. Interaction effects are common in regression analysis, ANOVA, and designed experiments. In this paper, “window size*preFOG duration” represents the interaction between the window size and preFOG duration.

RESULTS

The parameters related to feature extraction included the window size, sliding step length, and preFOG duration. The preFOG duration (LogWorth: 2.364, p -value < 0.01) had the largest impact on the kappa value, followed by the window size (Figure 5). The window size (LogWorth: 3.216, p -value < 0.01) had the largest impact on the F1 score followed by the preFOG

duration (Figure 6). Thus, there was a certain interaction between the window size and the preFOG duration defined in the experiment because of the statistically significant effect from the source item “window size*preFOG duration” (Figure 6).

Generally, the increase in the window size had an obvious positive effect on the F1 score, the increase in preFOG duration had an obvious negative effect on the F1 score, and the sliding step length did not affect our experiments (Figure 7). After analysis in the Minitab, the best combination of parameters in feature extraction was obtained and represented by a tuple (window size, step, and preFOG duration): (500, 20, and 250).

As the feature dimension is relatively high, this work also selects subsets of features and obtains the most important set of features {“F03,” “F05,” “F07,” “F09,” “F10,” “F11,” “F12,” “F13,” “F14,” “F15,” “F16,” and “F17”}. The OED was used to optimize the hyperparameters of the RF classifier. The optimal combination of parameters was presented as a tuple (n_estimators, max_features, max_depth, min_samples_split, and min_samples_leaf): (800, 0.1, 4, 4, and 4). In this tuple, max_features is the maximum ratio of features that can be used in a single subtree of the RF; “0.1” means a maximum of $924 \times 0.1 = 92$ features can be used.

With the optimized parameters, the RF model was trained and evaluated with an episode-based strategy. Under application conditions, FOG can be recorded as an episode, but normal gait cannot be evaluated as an episode. Therefore, we chose the false-positive rate, hit rate, and mean prediction time (MPT) for model evaluation. The false-positive rate, hit rate, and MPT were used as the indexes for evaluation of the prediction system. The false-positive rate increased with the hit rate. The false-positive rate, hit rate, and MPT were 27%, 68%, and 2.99 s, respectively.

DISCUSSION

Our key assumption was that FOG is not a sudden episode. In the period before FOG occurs, the patient’s kinematic features change. Based on this assumption, a prediction model was constructed, and the OED was used to optimize the hyperparameters to improve efficiency. To the best of our knowledge, this is the first time that the OED has been used for the ML-based prediction of FOG in PD. The interactions among the feature extraction parameters and the

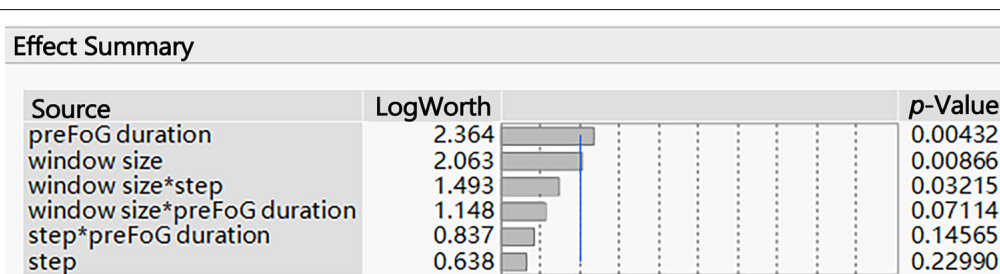


FIGURE 5 | Summary of the effects of the feature extraction parameters on the kappa value. The p -values show the statistical significance of the association between the parameters and the kappa value. *Multiplication.

Effect Summary			
Source	LogWorth		p-Value
window size	3.216		0.00061
preFoG duration	3.125		0.00075
window size*preFoG duration	1.796		0.01598
step*preFoG duration	1.457		0.03489
window size*step	1.198		0.06343
step	1.021		0.09521

FIGURE 6 | Summary of the effects of the feature extraction parameters on the F1 score. The p -values show the statistical significance of the association between the parameters and the F1 score. *Multiplication.

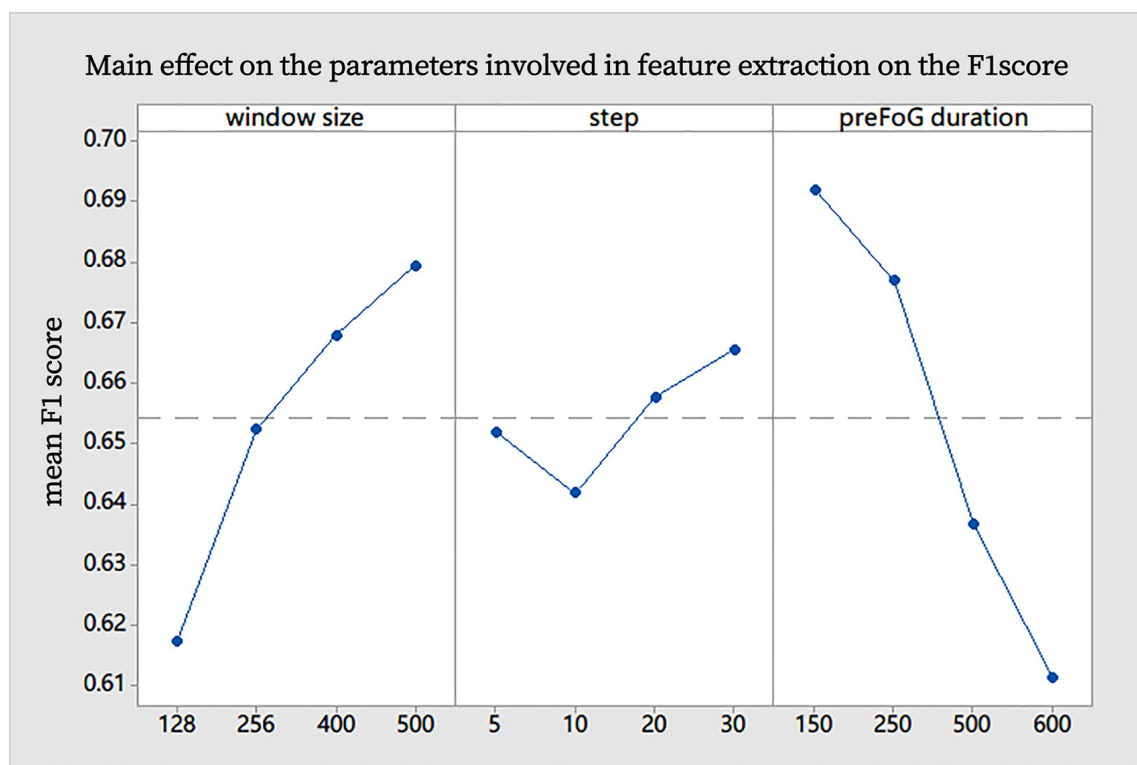


FIGURE 7 | Main effect on the parameters involved in feature extraction on the F1 score.

performance were first analyzed. The OED makes analyzing the effects of hyperparameters and finding the optimal ones more cost effective.

The OED was introduced to acquire optimized parameters involved in feature extraction. Having implemented the experimental design, the interactions were analyzed, the optimized parameters were obtained, and the efficiency of the test was improved. The existence of interactions hinted that it was necessary to tune the parameters simultaneously using combinatorial experiments and not a one-factor experimental design. Mikos et al. (2017) elucidated how the window size affects the MI between the feature and the FOG classification, which is a measure of the correlation between variables, and their works showed optimal window lengths for FOG classification vary

across feature types. Mazilu et al. (2013) illustrated the effects of preFOG duration on the prediction and illuminated gait parameter changes prior to FOG (Ferster et al., 2015), but they still used the empirical and a *priori* preFOG duration. Sixteen runs from the OED were used to analyze the effects of parameters such as the window size, preFOG duration, and sliding window step length on the F1 score for preFOG identification, while the full factorial design of three factors with four levels required 64 runs. The workload was low, and the optimal parameters were obtained. The optimized preFOG duration and window size also implied the length of the preFOG period. With the improvement in the model training efficiency, more models can be trained with the same data scale and computing resources. Therefore, training personalized FOG prediction models for different

individuals becomes possible. This is a possible direction to solve the problem of the poor generalizability of the model to different patients in real-world applications.

The classifier was evaluated by using a leave-one-patient-out, episode-based strategy and obtained an MPT of 2.99 s for FOG prediction with a false-positive rate of 27% and a hit rate of 68%. The performance of the model was not excellent in the patient-independent evaluation but was still comparable to previous works. As in the previous work by Mazilu et al. (2013) the FOG prediction performance was highly patient-dependent. The optimal parameters were not suitable for all patients, although our work solved these problems to some extent. We also developed a model with a patient-dependent method and obtained high performance with an F1 score of 0.89. In many ML models, the training and test data came from different patients, the identically distributed assumption is often violated, leading to poor performance. When the data from a patient were split into training and test data, the two datasets were often highly interdependent, leading to a good performance, but the model was overfit.

There were some limitations to our research. First, we did not include patients with a Hoehn–Yahr score greater than three since most of these patients have difficulty moving and are at risk of falling. The results might be affected if severe patients were included. Second, our method should be evaluated with the data of PD patients without FOG who were not included in our study. Third, the number of patients included in our study is small, weakening the experimental efficacy and supporting evidence. Future studies will be dedicated to improving the performance of the FOG prediction system by eliminating these defects.

Overall, a prediction classifier beneficial for early interventions for FOG was obtained. The OED was helpful in optimizing the hyperparameters. Larger-scale studies are needed.

REFERENCES

- Allcock, L., Rowan, E., Steen, I., Wesnes, K., Kenny, R., and Burn, D. (2009). Impaired attention predicts falling in Parkinson's disease. *Parkinsonism Relat. Disord.* 15, 110–115. doi: 10.1016/j.parkreldis.2008.03.010
- Asakawa, T., Fang, H., Sugiyama, K., Nozaki, T., Hong, Z., Yang, Y., et al. (2016a). Animal behavioral assessments in current research of Parkinson's disease. *Neurosci. Biobehav. Rev.* 65, 63–94. doi: 10.1016/j.neubiorev.2016.03.016
- Asakawa, T., Fang, H., Sugiyama, K., Nozaki, T., Kobayashi, S., Hong, Z., et al. (2016b). Human behavioral assessments in current research of Parkinson's disease. *Neurosci. Biobehav. Rev.* 68, 741–772. doi: 10.1016/j.neubiorev.2016.06.036
- Asakawa, T., Sugiyama, K., Nozaki, T., Sameshima, T., Kobayashi, S., Wang, L., et al. (2019). Can the latest computerized technologies revolutionize conventional assessment tools and therapies for a neurological disease? the example of Parkinson's disease. *Neurol. Med. Chir. (Tokyo)* 59, 69–78. doi: 10.2176/nmc.ra.2018-0045
- Bächlin, M., Hausdorff, J. M., Roggen, D., Giladi, N., Plotnik, M., and Tröster, G. (2009). "Online detection of freezing of gait in parkinson's disease patients: a performance characterization," in *Proceedings of the 4th International Conference on Body Area Networks*, Los Angeles, CA, 1–8. doi: 10.32457/ejhr.v7i1.1395
- Bachlin, M., Plotnik, M., Roggen, D., Maidan, I., Hausdorff, J. M., Giladi, N., et al. (2009). Wearable assistant for Parkinson's disease patients with the freezing

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

This study was approved by Ruijin Hospital affiliated to Shanghai Jiao Tong University School of Medicine. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

KR and SC designed and supervised the study, and double-checked the statistical analysis. ZC and GL collected the sample, performed the statistical analysis, and drafted the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by grants from the National Natural Science Foundation of China (grant nos. 81430022, 91332107, and 81371407), the Shanghai Jiao Tong University School of Medicine (DLY201603), and the Shanghai Municipal Science and Technology Major Project (grant no. 2018SHZDZX05).

ACKNOWLEDGMENTS

We thank all the patients who participated in this study.

- of gait symptom. *IEEE Trans. Inform. Technol. Biomed.* 14, 436–446. doi: 10.1109/titb.2009.2036165
- Cavazzuti, M. (2012). *Optimization Methods: From Theory to Design Scientific and Technological Aspects in Mechanics*. Berlin: Springer Science & Business Media.
- Contreras, A., and Grandas, F. (2012). Risk of falls in Parkinson's disease: a cross-sectional study of 160 patients. *Parkinsons Dis.* 2012:362572.
- Djurić-Jovičić, M. D., Jovičić, N. S., Radovanović, S. M., Stanković, I. D., Popović, M. B., and Kostić, V. S. (2013). Automatic identification and classification of freezing of gait episodes in Parkinson's disease patients. *IEEE Trans. Neural Syst. Rehabil. Eng.* 22, 685–694. doi: 10.1109/tnsre.2013.2287241
- Ferster, M. L., Mazilu, S., and Tröster, G. (2015). "Gait parameters change prior to freezing in Parkinson's disease: a data-driven study with wearable inertial units," in *Proceedings of the 10th EAI International Conference on Body Area Networks*, Brussels, 159–166.
- Han, J. H., Lee, W. J., Ahn, T. B., Jeon, B. S., and Park, K. S. (2003). "Gait analysis for freezing detection in patients with movement disorder using three dimensional acceleration system," in *Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEEE Cat. No. 03CH37439)* (Piscataway, NJ: IEEE), 1863–1865.
- Handojooseno, A. A., Shine, J. M., Nguyen, T. N., Tran, Y., Lewis, S. J., and Nguyen, H. T. (2012). "The detection of Freezing of Gait in Parkinson's disease patients using EEG signals based on Wavelet decomposition," in *Proceedings of the 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (Piscataway, NJ: IEEE), 69–72.

- Jovanov, E., Wang, E., Verhagen, L., Fredrickson, M., and Fratangelo, R. (2009). “deFOG—A real time system for detection and unfreezing of gait of Parkinson’s patients,” in *Proceedings of the 2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (Piscataway, NJ: IEEE), 5151–5154.
- Latt, M. D., Lord, S. R., Morris, J. G., and Fung, V. S. (2009). Clinical and physiological assessments for elucidating falls risk in Parkinson’s disease. *Mov. Disord.* 24, 1280–1289. doi: 10.1002/mds.22561
- Li, G., Ma, J., Cui, S., He, Y., Xiao, Q., Liu, J., et al. (2019). Parkinson’s disease in China: a forty-year growing track of bedside work. *Transl. Neurodegener.* 8:22.
- Lim, I., van Wegen, E., de Goede, C., Deutekom, M., Nieuwboer, A., Willems, A., et al. (2005). Effects of external rhythmical cueing on gait in patients with Parkinson’s disease: a systematic review. *Clin. Rehabil.* 19, 695–713. doi: 10.1191/0269215505scr906oa
- Mazilu, S., Blanke, U., and Tröster, G. (2015a). “Gait, wrist, and sensors: detecting freezing of gait in Parkinson’s disease from wrist movement,” in *Proceedings of the 2015 IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops)* (Piscataway, NJ: IEEE), 579–584.
- Mazilu, S., Calatroni, A., Gazit, E., Mirelman, A., Hausdorff, J. M., and Tröster, G. (2015b). Prediction of freezing of gait in Parkinson’s from physiological wearables: an exploratory study. *IEEE J. Biomed. Health Inform.* 19, 1843–1854. doi: 10.1109/jbhi.2015.2465134
- Mazilu, S., Calatroni, A., Gazit, E., Roggen, D., Hausdorff, J. M., and Tröster, G. (2013). “Feature learning for detection and prediction of freezing of gait in Parkinson’s disease,” in *Proceedings of the International Workshop on Machine Learning and Data Mining in Pattern Recognition*, ed. P. Perner (Berlin: Springer), 144–158. doi: 10.1007/978-3-642-39712-7_11
- Mazilu, S., Hardegger, M., Zhu, Z., Roggen, D., Tröster, G., Plotnik, M., et al. (2012). “Online detection of freezing of gait with smartphones and machine learning techniques,” in *Proceedings of the 2012 6th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth) and Workshops* (Piscataway, NJ: IEEE), 123–130.
- Mikos, V., Heng, C.-H., Tay, A., Chia, N. S. Y., Koh, K. M. L., Tan, D. M. L., et al. (2017). “Optimal window lengths, features and subsets thereof for freezing of gait classification,” in *Proceedings of the 2017 International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS)* (Piscataway, NJ: IEEE), 1–8.
- Moore, S. T., MacDougall, H. G., and Ondo, W. G. (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
- Moore, S. T., Yungher, D. A., Morris, T. R., Dilda, V., MacDougall, H. G., Shine, J. M., et al. (2013). Autonomous identification of freezing of gait in Parkinson’s disease from lower-body segmental accelerometry. *J. Neuroeng. Rehabil.* 10:19. doi: 10.1186/1743-0003-10-19
- Okuma, Y. (2006). Freezing of gait in Parkinson’s disease. *J. Neurol.* 253(Suppl. 7), VII27–VII32.
- Orphanidou, N. K., Hussain, A., Keight, R., Lishoa, P., Hind, J., and Al-Askar, H. (2018). “Predicting freezing of gait in Parkinsons disease patients using machine learning,” in *Proceedings of the 2018 IEEE Congress on Evolutionary Computation (CEC)* (Piscataway, NJ: IEEE), 1–8.
- Pepa, L., Verdini, F., Capecci, M., and Ceravolo, M. G. (2015). “Smartphone based freezing of gait detection for parkinsonian patients,” in *Proceedings of the 2015 IEEE International Conference on Consumer Electronics (ICCE)* (Piscataway, NJ: IEEE), 212–215.
- Pham, T. T., Moore, S. T., Lewis, S. J. G., Nguyen, D. N., Dutkiewicz, E., Fuglevand, A. J., et al. (2017). Freezing of gait detection in parkinson’s disease: a subject-independent detector using anomaly scores. *IEEE Trans. Biomed. Eng.* 64, 2719–2728. doi: 10.1109/tbme.2017.2665438
- Postuma, R. B., Berg, D., Stern, M., Poewe, W., Olanow, C. W., Oertel, W., et al. (2015). MDS clinical diagnostic criteria for Parkinson’s disease. *Mov. Disord.* 30, 1591–1601. doi: 10.1002/mds.26424
- Revuelta, G. J., Embry, A., Elm, J. J., Gregory, C., Delambo, A., Kautz, S., et al. (2015). Pilot study of atomoxetine in patients with Parkinson’s disease and dopa-unresponsive Freezing of Gait. *Transl. Neurodegener.* 4:24.
- Wood, B., Bilclough, J., Bowron, A., and Walker, R. (2002). Incidence and prediction of falls in Parkinson’s disease: a prospective multidisciplinary study. *J. Neurol. Neurosurg. Psychiatry* 72, 721–725. doi: 10.1136/jnnp.72.6.721
- Xia, Y., Zhang, J., Ye, Q., Cheng, N., Lu, Y., and Zhang, D. (2018). Evaluation of deep convolutional neural networks for detection of freezing of gait in Parkinson’s disease patients. *Biomed. Signal Process. Control* 46, 221–230. doi: 10.1016/j.bspc.2018.07.015
- Zach, H., Janssen, A. M., Snijders, A. H., Delval, A., Ferraye, M. U., Auff, E., et al. (2015). Identifying freezing of gait in Parkinson’s disease during freezing provoking tasks using waist-mounted accelerometry. *Parkinsonism Relat. Disord.* 21, 1362–1366. doi: 10.1016/j.parkreldis.2015.09.051
- Zhang, M., and Sawchuk, A. A. (2011). “A feature selection-based framework for human activity recognition using wearable multimodal sensors,” in *Proceedings of the 6th International Conference on Body Area Networks: BodyNets*, Brussels, 92–98.
- Zhang, Y. (2017). Can a smartphone diagnose parkinson disease? a deep neural network method and telediagnosis system implementation. *Parkinsons Dis.* 2017:6209703.

Conflict of Interest: ZC, YL, XY, and KR were employed by the company Gyenno Science Co., Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Chen, Li, Gao, Tan, Liu, Zhao, Ling, Yu, Ren and Chen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.