This article has been accepted for publication in IEEE Transactions on Biomedical Engineering. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TBME.2024.3402677

A Double-Hurdle Quantification Model for Freezing of Gait of Parkinson's Patients

Ningcun Xu, Chen Wang, Liang Peng, Xiao-Hu Zhou, *Member, IEEE*, Jingyao Chen, Zhi Cheng, and Zeng-Guang Hou, *Fellow, IEEE*

***Abstract*— Freezing of gait (FOG) leads to an increased risk of falls and limited mobility in individuals with Parkinson's disease (PD). However, existing research ignores the fine-grained quantitative assessment of FOG severity. This paper provides a double-hurdle model that uses typical spatiotemporal gait features to quantify the FOG severity in patients with PD. Moreover, a novel multi-output random forest algorithm is used as one hurdle of the double-hurdle model, further enhancing the model's performance. We conduct six experiments on a public PD gait database. Results demonstrate that the designed random forest algorithm in the double-hurdle model–hyperparameter independence framework achieves outstanding performances with the highest correlation coefficient (CC) of 0.972 and the lowest root mean square error (RMSE) of 2.488. Furthermore, we study the effect of drug state on the gait patterns of PD patients with or without FOG. Results show that "OFF" state amplifies the visibility of FOG symptoms in PD patients. Therefore, this study holds significant implications for the management and treatment of PD.**

***Index Terms*— Parkinson's disease, freezing of gait, zero-inflated data distribution, double-hurdle model, random forest algorithm, mixed-task learning, MRMR**

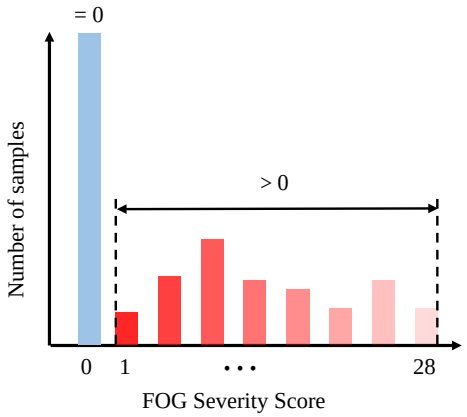


Fig. 1: The zero-inflated data distribution occurs when there are more PD patients with FOG severity scores of 0 than those with FOG severity scores greater than 0 in the collected PD gait databases. This imbalanced data distribution brings a significant challenge in accurately scoring the FOG severity of PD patients.

I. Introduction

Freezing of gait (FOG) refers to a sudden and brief episode where Parkinson's disease (PD) patients have difficulty initiating or continuing their normal walking pattern, generally occurring at the advanced stage of PD [1]. FOG can increase the risk of falls and reduce mobility in patients with PD [2]. Therefore, it is crucial to timely and accurately quantify FOG severity to help clinicians manage the FOG symptom more effectively and minimize its impacts on the quality of life [3].

In clinical practice, the New Freezing of Gait Questionnaire (NFOG-Q) has been commonly used for subjectively quantifying the FOG severity [4], [5]. However, the evaluation outcomes obtained through the NFOG-Q primarily depend on individual self-reporting and the experience of clinicians, which leads to a certain degree of subjectivity and uncertainty. With advancements in technology, researchers conduct a comprehensive and quantitative assessment of FOG severity in PD patients based on instrumented gait analysis [6], [7]. However, most studies focus on FOG severity in PD patients, overlooking the fine-grained of the assessment.

FOG generally appears in the later period of PD [9]. Thus, in PD gait databases, a significant portion of patients may not suffer from the FOG symptom. According to the scoring rules of NFOG-Q, the patients without FOG would receive a score of zero, which can lead to the zero-inflated data distribution in the collected databases and then bring a significant challenge in quantifying FOG severity, as shown in Fig.1.

This work was supported in part by the National Key Research and Development Program of China under Grant 2022YFC3601200, by the National Natural Science Foundation of China under Grant U1913601, Grant 62203441 and Grant U21A20479, and by the Beijing Natural Science Foundation under Grant L222013. (Corresponding author: Liang Peng, Zeng-Guang Hou.)

Ningcun Xu, Jingyao Chen and Zhi Cheng are with the Faculty of Innovation Engineering, Macau University of Science and Technology, Macao 999078, China, and with the State Key Laboratory of Multimodal Artificial Intelligence Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China (e-mail: 2202853nmi30001@student.must.edu.mo, 2009853pmi30003@student.must.edu.mo, 3220004221@student.must.edu.mo)

Chen Wang, Liang Peng and Xiao-Hu Zhou are with the State Key Laboratory of Multimodal Artificial Intelligence Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China (e-mail: wangchen2016@ia.ac.cn, liang.peng@ia.ac.cn, xiaohu.zhou@ia.ac.cn)

Zeng-Guang Hou is with the State Key Laboratory of Multimodal Artificial Intelligence Systems, Institute of Automation, CAS Center for Excellence in Brain Science and Intelligence Technology, Chinese Academy of Sciences, Beijing 100190, China, and with the School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing 100049, China (e-mail: zengguang.hou@ia.ac.cn)

Authorized licensed use limited to: INSTITUTE OF AUTOMATION CAS. Downloaded on May 21,2024 at 00:48:42 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Biomedical Engineering. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TBME.2024.3402677

Due to the zero-inflated data distribution, scoring FOG severity is regarded as a particular case of data imbalanced regression in this paper. Some solutions for the data imbalanced regression are helpful to address the zero-inflated data distribution during scoring FOG severity, such as resampling methods and reweighting methods [10], [11]. As one of the resampling methods, synthetic minority over-sampling technique (SMOTE) can create new PD patient samples for rare labels to balance the sample distribution in the PD gait databases [12], [13]. The limitation of SMOTE is that the PD patient samples with high-dimensional features are challenging to generate. The reweighting methods learn the distribution experience from the train sample labels of PD patients and apply it as prior knowledge to the loss function in the form of weight [14], [15]. However, the reweighting methods are suitable for scenarios with numerous PD patient samples.

The double-hurdle model is a widely employed solution to solve the zero-inflated data distribution in statistical analysis [16]. In this study, a novel double-hurdle model is designed to address the zero-inflated data distribution that arises in the FOG severity assessment. Instead of logistic regression or truncated regression models, machine learning algorithms are adopted as hurdles to make the model better for capturing nonlinear relationships in high-dimensional data. In the proposed double-hurdle model, the first hurdle can identify the patients with FOG and set up the FOG severity score of the patients without FOG as zeros, while the second hurdle can score the FOG severity of the patients with FOG. Additionally, the other novelty of the double-hurdle model is the provided two hyperparameter optimization paradigms: "hyperparameter sharing paradigm" and "hyperparameter independence paradigm". The former paradigm uses constraints between tasks to improve model robustness. In contrast, the latter paradigm provides more flexible hurdle combinations for the double-hurdle model.

Consequently, two tasks need to be accomplished in this study: distinguishing individuals with FOG from PD patients and assessing their FOG severity . Although these tasks fall into the realms of classification and regression separately, they are highly interrelated. This simultaneous learning paradigm can reduce the risk of overfitting and strengthen the model's ability to predict the FOG severity of unseen PD patients accurately [17].

In machine learning, multi-label algorithms can utilize the same PD patients gait database to accomplish multiple classification tasks, while multi-regression algorithms can utilize the same database to handle multiple regression tasks. However, the algorithm that can simultaneously complete the classification and regression tasks is the multi-output random forest algorithm (MRF) proposed by Linusson *et al.* [18], which is an extension of the random forest algorithm.

In this paper, a novel multi-output random forest algorithm based on mixed-task learning, MGWRF, is provided as one hurdle further to improve the performance of the proposed double-hurdle model [19]–[21]. Compared with MRF, the proposed MGWRF algorithm introduces a novel approach, called deviation-based differential information entropy, to quantify the gain of nodes for the regression task. The approach not only includes differential information entropy but also considers the predicted FOG scores, improving the algorithm's performance in scoring the FOG severity of PD patients. Moreover, in the proposed MGWRF algorithm, the weighted summation of information gains from two tasks is treated as the total information gain of a decision node. The optimal contribution level of each task towards the prediction results is determined to enhance the algorithm's performance in identifying patients with FOG and predicting the FOG severity of PD patients.

Finally, to investigate the effect of drug state on the gait patterns of PD patients with or without FOG, the maximum relevance minimum redundancy (MRMR) algorithm is employed to identify the ten top important spatiotemporal gait features that play a significant role in determining the performance of the double-hurdle model. The main contributions of our work can be summarised as follow:

1) To the best of our knowledge, this paper is the first to conduct a fine-grained assessment of the FOG severity in PD patients using the proposed double-hurdle model.
2) A novel multi-output random forest algorithm based on mixed-task learning, MGWRF, is proposed to improve further the double-hurdle model's performance in scoring FOG severity of PD patients.
3) The ten most significant spatiotemporal gait features are selected using the MRMR algorithm to investigate the impact of drug state on the gait patterns in PD patients with or without FOG.

The rest of this paper is organized as follows. Section II provides the related works. Section III describes the public PD gait database and the data preprocessing methods. Section IV introduces the double-hurdle model. Section V shows the experiments and result analysis. Section VI concludes this paper.

II. Related Work

In recent years, numerous studies have concentrated on quantitatively assessing FOG severity in PD patients through instrumented gait analysis (IGA) [6], [7], [22].

Firstly, some studies employ the quantitative assessment of FOG severity by detecting and analyzing FOG episodes during the walking of PD patients. Sigcha *et al*. introduced a Transformer-based model, significantly enhancing the accuracy of FOG episode detection [23]. Borzì *et al*. further advanced real-time FOG episode detection [24]. These advancements have considerably improved the practical application of FOG episode detection algorithms. Notably, studies have uncovered a correlation between FOG episodes and the motor function of PD patients [25]. Zhang *et al*. discovered a strong positive association between the duration of detected FOG episodes and the patient's Hoehn and Yahr (H&Y) stage [26].

Secondly, other studies quantify FOG severity in PD patients by collecting and analyzing gait data during walking processes where FOG episodes do not occur. These studies eliminate the need to induce FOG episodes during walking experiments, reducing data collection challenges. Aich *et al*.

Authorized licensed use limited to: INSTITUTE OF AUTOMATION CAS. Downloaded on May 21,2024 at 00:48:42 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Biomedical Engineering. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TBME.2024.3402677

TABLE I: Extracted gait features from the raw spatiotemporal gait feature set X^O

Groups	Gait Features(unit)
Rhythm gait features	Swing Phase Time(s); Stance Phase Time(s); Single Limb Support Time(s); Double Limbs Support Time(s); Cadence(steps/minutes); Walk Ratio
Phase gait features	Stance Phase (%); Swing Phase (%); Single Limb Support Phase (%); Double Limbs Support Phase (%)
Pace gait features	Step Length (m); Stride Length (m); Speed (m/s)
Asymmetry gait features	Stance Time Asymmetry; Swing Time Asymmetry; Step Length Asymmetry; Stride Length Asymmetry; Step Width Asymmetry
Variability gait features	Stance Time Variability; Swing Time Variability; Stride Length Variability; Step Length Variability; Single Limb Phase Variability; Double Limbs Phase Variability; Step Length Variability

[1] $walk\ ratio = \frac{step\quad length(m)}{cadence(steps/minute)}$

collected five gait parameters during walking and applied SVM to identify PD patients with FOG, achieving an 89.14% accuracy [27]. Park *et al*. extracted kinematic features during turning and used Random Forest to distinguish PD patients with FOG, obtaining a precision of 81.4% [28]. Kwon *et al*. categorized FOG severity into four levels using the Movement Disorder Society-Sponsored Revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS) and developed a multi-task deep learning model, notably enhancing the F1-score for FOG severity classification to 96.7% [29]. However, these studies lack a fine-grained assessment of FOG severity.

This study uses the method, like the second kind of studies, to evaluate FOG severity in PD patients. To refine the assessment granularity, we utilize the NFOG-Q clinical scale to score the FOG severity in PD patients. The range of FOG scores spans from 0 to 28. Moreover, we develop a double-hurdle model to address the zero-inflation problem, and the MGWRF algorithm is proposed to improve the double-hurdle model's performance on quantitative FOG severity in PD patients.

III. Materials

A. PD Gait Database

The PD gait database utilized in this study was obtained from the Laboratory of Biomechanics and Motor Control at the Federal University of ABC [30]. It comprised 13 patients diagnosed with FOG (9 males and 4 females, age: 62±10 years, height: 165.6±7.4 cm, weight: 70.7±13.9 kg, NFOG-Q score: 18.5±5.1) and 13 patients without FOG (11 males and 2 females, age: 66.5±8.6 years, height: 166.4±6.7 cm, weight: 72.1±11.4 kg, NFOG-Q score: 0±0). In addition, the PD gait database provided 19 spatiotemporal gait features, including Stance Time (s), Swing Time (s), Step Length (m), Cadence (steps/s), Stride Length (m), Speed (m/s), Stride Width (m), Double Limbs Support Time (s), Gait Cycle Time (s), and so on. In the PD gait database, each patient perform walking experiment in the "ON" and "OFF" states, and there are 3 PD patients with missing experimental data in the "OFF" state.

B. Spatiotemporal Gait Feature Preprocessing

In this paper, the spatiotemporal gait features from the PD gait database are utilized to evaluate FOG severity of 23 PD patients who are in the "OFF" state. We denote the original gait database as $X^O \in \mathbb{R}^{m\times n\times g}$, where $m = 23$, $n = 19$, $g = 20$ represent the number of PD patients, the number of gait features, and the number of gait cycles, respectively. After expanding, nondimensionalizing, and reducing dimensionality of the original dataset X^O, we obtain the dataset $X^D \in \mathbb{R}^{m\times p\times q}$, where $m = 23$, $p = 25$, $q = 1$ represent the number of PD patients, the number of gait features, and the number of gait cycles, respectively [31]- [33]. Each gait feature value for each participant in X^D is the average of the corresponding gait features in X^O across all gait cycles. These 25 gait features also can be categorized into five common groups [34], including rhythm gait features, phase gait features, pace gait features, asymmetry gait features, and variability gait features, as illustrated in Table I. The higher variability and asymmetry in gait features indicate weaker walking ability. A higher walk ratio represents a more efficient walking pattern. Faster walking speed indicates stronger walking ability. A longer stance phase indicates potential balance issues. Pathological gait may lead to shorter step and stride lengths to maintain a certain speed.

IV. Double-Hurdle Model

A. Overall Framework

The double-hurdle model is proposed to quantify the FOG severity of PD patients, as illustrated in Fig. 2. The first step of this framework is the preprocessing of the original spatiotemporal gait features X^O and we obtain the dimensionless feature set X^D. This preprocessing includes feature extracting, nondimensionalizing, reducing dimensionality and grouping.

In the second step, X^D is standardized to mitigate the impact of amplitude differences of gait features on the model's performance. Thus, we obtain the standardized gait feature set $X^N \in \mathbb{R}^{m\times p\times q}$. Subsequently, utilizing the MRMR algorithm, we select the k most important gait features to construct the gait feature set $X^S \in \mathbb{R}^{s\times k\times q}$, where s, k represent the number of PD patients and the number of selected gait features. X^S serves as the input for the double-hurdle model. In this study, we exclusively employ the MRMR algorithm within the classification task to complete gait feature selection.

In the third step, a double-hurdle model is developed to address the zero-inflated data distribution when assessing the FOG severity in PD patients. The first hurdle of the model aims to classify the samples into two groups: the zero sample ($\hat{y}_c = 0$) representing patients without FOG, with a corresponding

Authorized licensed use limited to: INSTITUTE OF AUTOMATION CAS. Downloaded on May 21,2024 at 00:48:42 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Biomedical Engineering. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TBME.2024.3402677

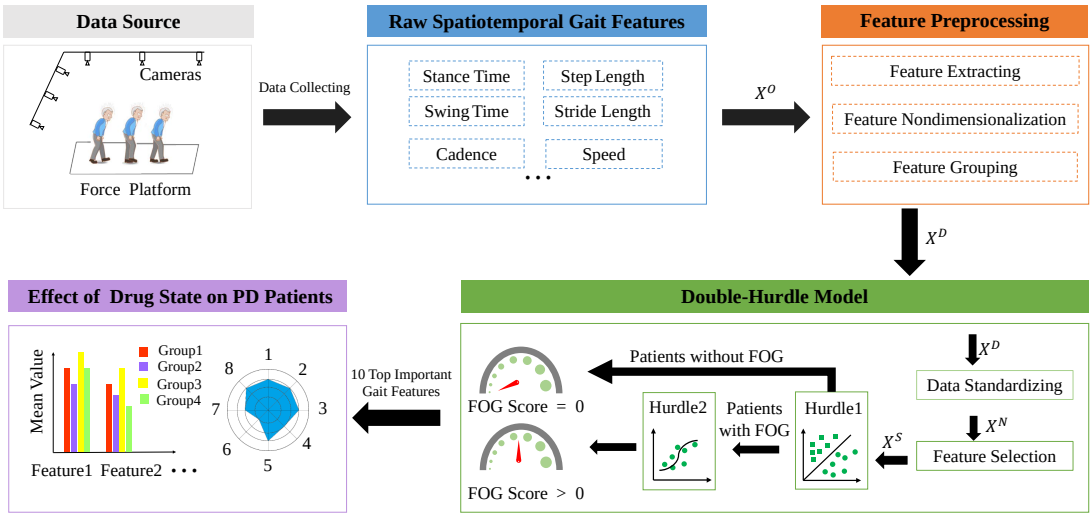


Fig. 2: The overall framework of double-hurdle model for quantifying FOG severity of PD patients.The detailed description is present in section IV-A

FOG severity score of $\hat{y}_r = 0$, and the non-zero sample ($\hat{y}_c = 1$) indicating patients with FOG. The primary purpose of the second hurdle is to predict the FOG severity score ($\hat{y}_r$) for the patient with FOG. The FOG severity scores of PD patients with FOG are constant values greater than 0.

The fourth step in the framework involves analyzing the ten most important spatiotemporal gait features obtained from the third step. This analysis aims to examine the effect of drug state on walking ability and gait patterns of PD patients with or without FOG. The effect analysis is presented in both bar table and radar chart forms. Detailed discussions on the analysis results will be presented in Section V-H.

B. MGWRF Algorithm

In this section, we improve the construction process of decision tree, which involves the node splitting criterion and the prediction mechanism of leaf node. Building on these improvements, a novel multi-output random forest algorithm, MGWRF, is introduced to leverage the improved decision tree to handle both classification and regression tasks simultaneously. In this paper, the classification task involves distinguishing PD patients with or without FOG, while the regression task focuses on quantifying the FOG severity of PD patients. This subsection provides a detailed explanation of the improved construction process of decision tree and the construction process of the MGWRF algorithm.

1) Construction Process of Improved Decision Tree: When making predictions using a decision tree, the process starts at the root node t_r, as shown in Fig.3. The root node t_r is then split into child nodes according to a specific feature x_i^s. This process continues until reaching a leaf node t_f. Then, the predicted category value $\hat{y}_c$ or regression value $\hat{y}_r$ is obtained according to the prediction mechanism of leaf node. During the construction process of decision trees, we refer to the root node and child nodes as decision nodes t_d. Therefore, the construction process of the decision tree needs to solve two key problems. The first is to choose which feature x_i^s for the splitting of the decision node t_d; the second is when to stop the node's splitting and how to predict the value of the classification result $\hat{y}_c$ or the value of the regression result $\hat{y}_r$ to the leaf nodes t_f.

In this paper, we utilize the information gain criterion to select the gait feature x_i^s for splitting the decision node. Compared with using other gait features, the information gain of the decision node is the largest after using the selected gait feature x_i^s to divide the decision node t_d. Drawing inspiration from Linusson *et al.* [18], we design a function to calculate the information gain of decision nodes, which simultaneously considers both classification and regression tasks. The function involves the following four steps. In the first step, for the classification task C, the information entropy $H(\cdot)$ of the decision node t_d is computed as the Shannon entropy, as shown in the equation (1).

$$H_C(t_d) = -\sum_{c=1}^{2} p_c \log p_c, \tag{1}$$

where p_c represents the probability of the class label c in the decision node t_d. In our study, there are two types of class labels: $c = 1$ corresponds to patients without FOG, and $c = 2$ corresponds to patients with FOG. The entropy value $H_C(t_d)$ of the decision node t_d is normalized to compare information entropy values across different tasks, as shown in equation (2).

$$H_{norm,C}(t_d) = \frac{H_C(t_d)}{H_C(t_r)}, \tag{2}$$

Then, the feature x_i^s is used to part the current decision node t_d into M decision nodes $t_{d+1,m}, m = 1, 2, ..., M$. After split based on the gait feature x_i^s, the information entropy $H_{norm,C}(t_d \mid x_i^s)$ of the node t_d is calculated using the following equation (3).

$$H_{norm,C}(t_d \mid x_i^s) = \sum_{m=1}^{M} \frac{|S_m|}{|S|} H_{norm,C}(t_{d+1,m}), \tag{3}$$

Authorized licensed use limited to: INSTITUTE OF AUTOMATION CAS. Downloaded on May 21,2024 at 00:48:42 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Biomedical Engineering. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TBME.2024.3402677

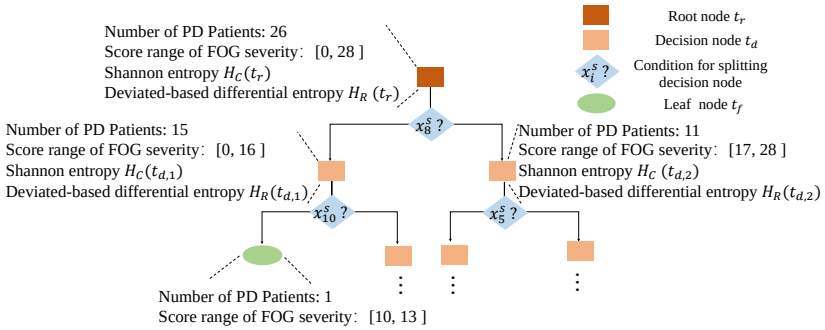


Fig. 3: Construction process of the improved decision tree. At the beginning, the root node t_r is split into decision nodes t_d according to the selected gait feature x_8^s. This process continues until reaching a leaf node t_f. The values of the leaf node t_f are set up to the class with the highest probability among the corresponding PD patient samples for the classification task C and the mean regression label value of the corresponding PD patient samples for the regression task R.

where $|S_m|$ and $|S|$ represent the number of samples in the decision nodes t_d and $t_{d+1,m}$ respectively.

In the second step, the deviated-based differential entropy $H_R(\cdot)$ of the decision node t_d is proposed for the regression task, as shown in the following equation (4).

$$H_R(t_d) = -\int_{\Delta Y_R} f(\Delta y_r) \log f(\Delta y_r) d\Delta y_r, \tag{4}$$

where $f(\cdot)$ represents the probability density function of regression label set ΔY_R. In this set, $\Delta y_{r,j}$ is computed as the difference between the true regression label $y_{r,j}$ and the mean value $\bar{y}_r$ of the true regression labels in the node t_d. Similar to the first step, for the regression task R, the deviated-based differential entropy $H_{norm,R}(t_d)$ of the decision node t_d is normalized using the following equation (5). After divided by using the feature x_i^s, the entropy value $H_{norm,R}(t_d \mid x_i^s)$ of the decision node t_d is calculated using the equation (6).

$$H_{norm,R}(t_d) = \frac{H_R(t_d)}{H_R(t_r)}, \tag{5}$$

$$H_{norm,R}(t_d \mid x_i^s) = \sum_{m=1}^{M} \frac{|S_m|}{|S|} H_{norm,R}(t_{d+1,m}), \tag{6}$$

In the third step, each decision node's information gain $IG_T(t_d \mid x_i^s)$ is calculated using the equation (7).

$$IG_T(t_d \mid x_i^s) = H_{norm,T}(t_d) - H_{norm,T}(t_d \mid x_i^n), \tag{7}$$

where T represents the classification task C or the regression task R.

In the fourth step, the total information gain $IG(t_d \mid x_i^s)$ of the decision node t_d is calculated as the weighted summation of the information gains for all tasks, as shown in the following equation (8).

$$IG(t_d \mid x_i^s) = \omega * IG_C(t_d \mid x_i^s) + (1-\omega) * IG_R(t_d \mid x_i^s), \tag{8}$$

where the range of the weight ω is $[0, 1]$.

The prediction mechanism of leaf node is designed to address the second problem. Firstly, if a decision node t_d contains less than two samples, the splitting process is immediately stopped and the decision node t_d is designated as the leaf node t_f. Secondly, for the classification task C, the value of the leaf node t_f is set up to the class with the highest probability among the corresponding training samples. Lastly, for the regression task R, the value of the leaf node t_f is set up to the mean regression label value of the corresponding training samples.

2) Construction Process of MGWRF Algorithm: Based on the construction process of the improved decision tree, this paper introduces the MGWRF algorithm's construction process to distinguish PD patients with or without FOG and quantify the FOG severity of PD patients. In the first step, we randomly extract P sample subsets from N PD patient gait data samples. Subsequently, we utilize these P sample subsets to construct P decision trees according to the construction process of the improved decision tree. Finally, we combine the outputs from all the decision trees to obtain the output result of the MGWRF algorithm. In the classification task, each decision tree is employed to determine whether the test PD patient suffers or does not suffer FOG. The predictions from all decision trees are aggregated, and the category with the highest number of votes becomes the final outcome of the MGWRF algorithm. In the regression task, the output produced by the MGWRF algorithm is the mean of the outputs generated by all decision trees.

V. Experiment and Result Analysis

A. Experiment Setup

1) Training and Testing Strategies: In this study, we repeat five times five-fold cross-validation on the whole PD gait dataset to evaluate the performance of all algorithms. In each time of five-fold cross-validation, the whole PD gait dataset is split into five folds at the patient level but not the gait cycle level. The feature selection is performed on the folds used for training.

This paper introduces a double-hurdle model for evaluating the severity of FOG in each patient with PD. In this model, the first hurdle identifies PD patients with FOG and assigns a severity score of 0 to those without FOG. The second hurdle quantifies the FOG severity in PD patients who experience it. The training and testing procedures of the double-hurdle model are elaborated in detail in Fig. 4. During the training phase, the following steps are performed. First, the PD gait

Authorized licensed use limited to: INSTITUTE OF AUTOMATION CAS. Downloaded on May 21,2024 at 00:48:42 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Biomedical Engineering. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TBME.2024.3402677

TABLE II: Results in single model framework

Experiment Types	Algorithms	Classification Metrics ACC	Pre	Rec	F1 score	Regression Metrics CC	RMSE
Comparison Experiment	SVM	0.843±0.135	1.000±0.000	0.667±0.279	0.767±0.200	0.830±0.140	9.152±1.755
	MLP	0.817±0.186	0.833±0.211	0.833±0.211	0.820±0.183	0.533±0.471	8.273±1.423
	RF	0.917±0.105	0.950±0.100	0.900±0.200	0.905±0.131	0.855±0.121	6.483±3.542
Ablation Experiment	MRF	0.917±0.105	1.000±0.000	0.833±0.211	0.893±0.137	0.778±0.110	9.199±2.601
	MGRF	0.917±0.105	1.000±0.000	0.833±0.211	0.893±0.137	0.882±0.093	9.024±1.919
	MWRF	0.950±0.100	1.000±0.000	0.900±0.200	0.933±0.133	0.888±0.096	9.124±1.435
	MGWRF(Ours)	0.933±0.133	1.000±0.000	0.867± 0.267	0.900±0.200	0.924±0.075	8.940±2.335

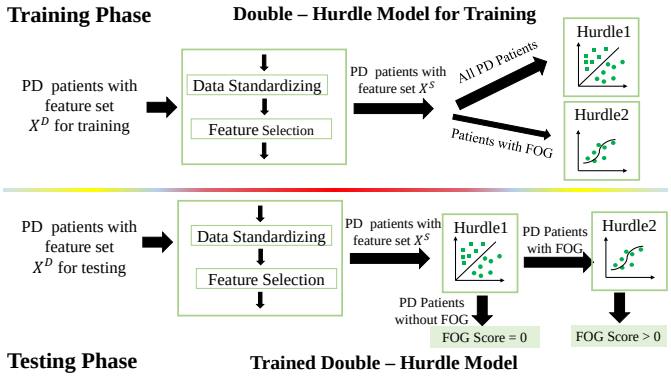


Fig. 4: Training and testing strategies for the double-hurdle model.

database for training is standardized, and then the MRMR algorithm is employed to select the k most important gait features from the training database. Next, using the training database with the selected gait features, the first hurdle is trained to classify patients into the two groups, *i.e.*, patients with FOG and patients without FOG. Finally, the second hurdle is trained using the selected gait features from the training database, which exclusively includes patients with FOG. During the testing phase, the same preprocessing step is applied to the test samples as we do during the training phase. Next, the first hurdle is utilized to determine whether a sample should be classified as 0 or 1. A classification of 0 indicates that the sample represents a PD patient without FOG symptom, resulting in a FOG severity score of 0 for the patient. Conversely, a classification of 1 indicates that the sample represents a PD patient with FOG. In this case, the second hurdle is employed to predict the patient's FOG severity score.

2) Algorithms for Comparison: In this paper, different algorithms are used as hurdles to study the performance of the double-hurdle model. Several advanced algorithms are evaluated in our study, including Support Vector Machine (SVM), Multilayer Perceptron (MLP), RF, and MRF. Compared with the MGWRF, MRF uses differential entropy to calculate the information gain of the decision node in the regression task and selects either the mean value, maximum value, or a random value of the decision node's information gains in different tasks as the information gain of the decision node. In addition, to verify the performance of MGWRF designed in this paper, its two variants are proposed by this paper, including MGRF and MWRF. Compared with MRF, MGRF uses the deviation-based differential entropy to calculate the information gain of the decision node in the regression task and MWRF uses the weighted sum of the information gains of the decision node in different tasks as the information gain of the decision node.

The paper explores the influence of different hyperparameters of various algorithms on the performance of double-hurdle model. Specifically, the number of gait features k is varied within the range of 8 to 20. For SVM, the hyperparameters 'c' and 'gamma' are constrained within the range [0, 1]. MLP utilizes varying hidden layer sizes, specifically 16, 32, or 64, and offered multiple activation functions: 'relu', 'identity', 'logistic', or 'tanh'. In the case of RF, MRF, MGRF, MWRF, and MGWRF, they all have 8 to 21 estimators. For MWRF and MGWRF, the weights w of the information gains for different tasks are limited to the range [0, 1]. The grid search method is used to find the optimal hyperparameters of each algorithm. All machine learning algorithms are built using scikit-learn.

3) Evaluation Metrics: Several common metrics are used to evaluate the performance of the first hurdle in identifying PD patients with or without FOG, including accuracy (Acc), precision (Pre), recall (Rec), F1 score (F1), and area under receiver operating characteristic (ROC) curve (AUC). The evaluation metrics used for testing the performance of the double-hurdle model in quantifying FOG severity are root mean square error (RMSE) and correlation coefficient (CC). We calculate the mean and standard deviation of each metric in all folds to indicate the performance of the double-hurdle model.

B. Results in Single Model Framework

1) Result Analysis in Comparison Experiment: In the comparison experiment, the RF algorithm achieves the best performance in identifying PD patients with or without FOG. Compared with RF, the MGWRF algorithm demonstrates some improvements in various aspects. It achieves a 0.016 increase in Acc, a notable 0.069 increase in CC, and a 2.457 increase in RMSE. Additionally, the F1 score of the MGWRF algorithm surpasses that of SVM and MLP by 0.133 and 0.08, respectively. Moreover, the RMSE of MGWRF is 0.212 lower than that of SVM. These results validate the effectiveness of the mixed-task learning paradigm, which can improve the performance of the RF algorithm on each task. Additionally,

Authorized licensed use limited to: INSTITUTE OF AUTOMATION CAS. Downloaded on May 21,2024 at 00:48:42 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Biomedical Engineering. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TBME.2024.3402677

TABLE III: Results in double-hurdle model – hyperparameter sharing framework

Experiment Types	Algorithms	Classification Metrics				Regression Metrics	
		Acc	Pre	Rec	F1 score	CC	RMSE
Comparison Experiment	SVM-SVM	0.843±0.135	1.000±0.000	0.667±0.279	0.767±0.200	0.862±0.114	6.072±2.146
	MLP-MLP	0.793±0.112	0.933±0.133	0.667±0.279	0.727±0.167	0.819±0.089	6.880±2.220
	RF-RF	0.900±0.122	0.933±0.133	0.900±0.200	0.893±0.137	0.874±0.111	6.215±2.710
Ablation Experiment	MRF-MRF	0.883±0.145	1.000±0.000	0.767±0.291	0.893±0.211	0.911±0.103	4.269±2.632
	MGRF-MGRF	0.917±0.105	1.000±0.000	0.833±0.211	0.893±0.137	0.924±0.948	3.967±2.770
	MWRF-MWRF	0.917±0.105	0.933±0.133	0.933±0.133	0.920±0.098	0.920±0.105	3.948±2.551
	MGWRF-MGWRF(Ours)	0.967±0.067	1.000±0.000	0.933±0.133	0.960±0.080	0.939±0.109	3.767±1.991

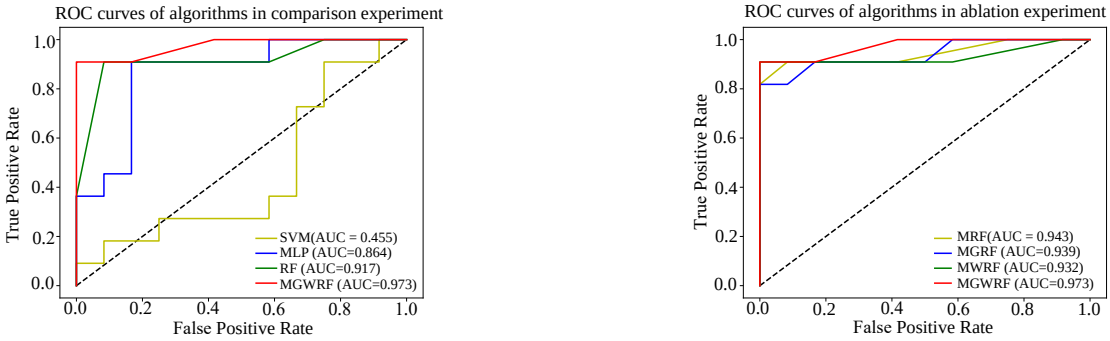


Fig. 5: ROC curves of algorithms in single model framework. (a) ROC curves of various algorithms in comparison experiment, including that SVM, MLP, RF, MGWRF. (b) ROC curves of algorithms in ablation experiment, including MRF, MGRF, MWRF, MGWRF.

Fig.5(a) illustrates the ROC curve of each algorithm in the comparison experiment. The MGWRF algorithm achieves the highest AUC of 0.973, which is increased by 0.056, 0.109, 0.518 than RF, MLP, and SVM, respectively.

2) Result Analysis in Ablation Experiment: In ablation experiment, the MGWRF algorithm demonstrates superior performance in identifying patients with or without FOG, as shown in Table II. It achieves an Acc of 0.933, a Pre of 1.000, a Rec of 0.867, and an F1 score of 0.900. Compared to MWRF, the MGWRF algorithm experiences a slight decrease of 0.017 in Acc and 0.033 in Rec. However, the MGWRF algorithm excels with a CC of 0.924 and an RMSE of 8.940 in scoring the FOG severity. Additionally, Fig.5(b) illustrates the ROC curve of each algorithms in the ablation experiment. The MGWRF algorithm achieves the highest AUC of 0.973, which is increased by 0.030, 0.034, 0.041 than MRF, MGRF, and MWRF, respectively. The oscillations in the ROC curves of SVM and MLP indicate their lower effectiveness than RF-based algorithms in detecting the PD patients with FOG symptoms. These findings indicate that the MGWRF algorithm proposed in this paper effectively utilizes shared knowledge from different tasks to enhance its performance in assessing the FOG severity of PD patients.

C. Results in Double-Hurdle Model – Hyperparameter Sharing Framework

1) Result Analysis in Comparison Experiment: In comparison experiment, the RF-RF model achieves exceptional results with the highest Acc of 0.900, and the highest F1 score of 0.893, as shown in Table III. Compared with the RF-RF model, the MGWRF-MGWRF model outperforms the RF-RF model in several metrics. It improves Acc by 0.067, F1 score by 0.067, CC by 0.065, and reduces RMSE by 2.448. These results demonstrate the mixed-task learning paradigm can help the RF algorithm use the shared knowledge between tasks to improve its performance in quantifying the FOG severity of PD patients.

2) Result Analysis in Ablation Experiment: In the ablation experiment, the MGWRF-MGWRF model achieves remarkable performance in quantifying the FOG severity of PD patients. It obtains the highest Acc of 0.967, surpassing the MRF-MRF model by 0.084 and outperforming MGRF-MGRF and MWRF-MWRF models by 0.05, as shown in Table III. These results highlight the ability of the MGWRF-MGWRF model to leverage the shared information between different tasks, improving the performance of the double-hurdle model in quantifying the FOG severity of PD patients.

3) Result Analysis in Frameworks Comparison: The performance of models in the double-hurdle model–hyperparameter sharing framework does not show significant differences in distinguishing PD patients with or without FOG than the corresponding models in the single model framework. This phenomenon occurs because these models in the double-hurdle framework employ the same hyperparameters and use grid search to find optimal hyperparameters for the classification task, just like their counterparts in the single model framework. For the regression task, the models in the double-hurdle model–hyperparameter sharing framework outperform

Authorized licensed use limited to: INSTITUTE OF AUTOMATION CAS. Downloaded on May 21,2024 at 00:48:42 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Biomedical Engineering. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TBME.2024.3402677

TABLE IV: Results in double-hurdle model – hyperparameter independence framework

Experiment Types	Algorithms	Classification Metrics				Regression Metrics	
		Acc	Pre	Rec	F1 score	CC	RMSE
Comparison Experiment	MGWRF-SVM	0.950±0.100	1.000±0.000	0.900±0.200	0.933±0.133	0.924±0.094	5.176±2.589
	MGWRF-MLP	0.967±0.067	1.000±0.000	0.933±0.133	0.950±0.080	0.919±0.121	5.925±2.564
	MGWRF-RF	0.950±0.100	1.000±0.000	0.900±0.200	0.933±0.133	0.923±0.088	4.854±2.531
Ablation Experiment	MGWRF-MRF	0.950±0.100	1.000±0.000	0.900±0.200	0.933±0.133	0.928±0.093	4.117±2.874
	MGWRF-MGRF	0.933±0.133	1.000±0.000	0.867±0.267	0.900±0.200	0.955±0.072	3.300±1.857
	MGWRF-MWRF	0.967±0.067	1.000±0.000	0.933±0.133	0.960±0.080	0.972±0.047	3.024±1.672
	MGWRF-MGWRF(Ours)	0.967±0.067	1.000±0.000	0.933±0.133	0.960±0.080	0.972±0.050	2.488±1.869

TABLE V: Results in comparison experiment with deep learning algorithms

Experiment Types	Model	Classification Task				Regression Task	
		ACC	Pre	Rec	F1 score	CC	RMSE
Single-task Learning	CNN-LSTM	0.733±0.226	0.667±0.279	0.800±0.245	0.700±0.245	0.780±0.129	9.234±1.224
	ResNet	0.700±0.292	0.600±0.490	0.533±0.452	0.560±0.463	0.755±0.128	4.690±0.063
	TCNet	0.800±0.187	0.767±0.291	0.833±0.211	0.773±0.225	0.638±0.112	11.238±1.866
	FT-Transformer	0.883±0.145	0.767±0.291	0.933±0.133	0.793±0.245	0.767±0.350	5.349±2.867
Multi-task Learning	CNN-LSTM	0.513±0.086	0.400±0.490	0.180±0.223	0.248±0.305	0.596±0.292	11.629±2.526
	ResNet	0.620±0.194	0.500±0.447	0.400±0.374	0.433±0.389	0.518±0.135	5.221±0.301
	TCNet	0.470±0.117	0.200±0.400	0.100±0.200	0.133±0.267	0.549±0.330	4.067±1.822
	FT-Transformer	0.533±0.323	0.667±0.422	0.600±0.374	0.567±0.327	0.791±0.261	2.394±2.092
Double-hurdle Model	CNN-LSTM*	0.487±0.086	0.600±0.490	0.267±0.226	0.367±0.306	0.779±0.087	10.332±2.317
	ResNet*	0.470±0.117	0.700±0.400	0.367±0.194	0.480±0.261	0.553±0.277	6.106±3.057
	TCNet*	0.563±0.127	0.700±0.400	0.480±0.319	0.514±0.260	0.590±0.238	12.839±1.789
	FT-Transformer*	0.630±0.218	0.700±0.400	0.580±0.382	0.581±0.325	0.514±0.512	4.923±2.418
	MGWRF-MGWRF*	0.967±0.067	1.000±0.000	0.933±0.133	0.960±0.080	0.939±0.109	3.767±1.991
	MGWRF-MGWRF**	0.967±0.067	1.000±0.000	0.933±0.133	0.960±0.080	0.972±0.050	2.488±1.869

* indicates that the algorithm serves as the hurdle in the hyperparameter sharing framework of the double-hurdle model.
** indicates that the algorithm serves as the hurdle in the hyperparameter independence framework of the double-hurdle model.

the counterparts in the single model framework. For instance, compared to the MGWRF model, the MGWRF-MGWRF model increases the CC by 0.031 and decreases the RMSE by 5.315. As a result, the double-hurdle model–hyperparameter sharing framework offers an effective solution to the zero-inflated data distribution.

D. Results in Double-Hurdle model – HyperParemeter Independence Framework

1) Result Analysis in Comparison Experiment: Table IV presents the results of all models in the double-hurdle model–hyperparameter independence framework. Whether in comparison experiment or in ablation experiment, all models exhibit the same ability to identify PD patients with FOG. This consistency is achieved by employing MGWRF as the first hurdle of the double-hurdle model, ensuring optimal performance of the double-hurdle model on the classification task. Compared to the MGWRF-MLP model, the CC of the MGWRF-MGWRF model improves by 0.053. Compared to the MGWRF-RF model, the RMSE of the MGWRF-MGWRF model decreases by 2.366, as shown in Table IV. These findings further validate the mixed-task paradigm can help the RF algorithm leverage shared information between different tasks to enhance the double-hurdle model's performance in assessing the FOG severity of PD patients.

2) Result Analysis in Ablation Experiment: In ablation experiment, the MGWRF-MGWRF model achieves outstanding results with the highest CC of 0.972 and the lowest RMSE of 2.488, as shown in Table IV. The RMSE of the MGWRF-MGWRF model has a 95% confidence interval of (1.716, 3.259), calculated using the t-test. Furthermore, the MGWRF-MGWRF model demonstrates a strong correlation (correlation coefficient = 0.972, p-value = 3.567e-12 < 0.05) with the NFOG-Q clinical scale. These findings demonstrate that the MGWRF algorithm can capture the shared knowledge between classification and regression tasks to improve its ability to address the zero-inflated data distribution arising in assessing the FOG severity of PD patients.

3) Result Analysis in Frameworks Comparison: Compared to the previous two frameworks, the double-hurdle model–hyperparameter independence framework offers a more flexible combination of algorithms, improving the performance of the double-hurdle model in assessing FOG severity of PD patients, as shown in Table II, III, IV. Therefore, the double-hurdle model with the "hyperparameter independence paradigm" is also a more effective solution to the zero-inflated data distribution than the single model framework and the double-hurdle model–hyperparameter sharing framework.

E. Comparison Study with Deep Learning Algorithms

We compare the proposed MGWRF-MGWRF model with state-of-the-art (SOTA) deep learning algorithms, including Convolutional Neural Network - Long Short-Term Memory (CNN-LSTM) [36], Residual Networks [37], Temporal

Authorized licensed use limited to: INSTITUTE OF AUTOMATION CAS. Downloaded on May 21,2024 at 00:48:42 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Biomedical Engineering. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TBME.2024.3402677

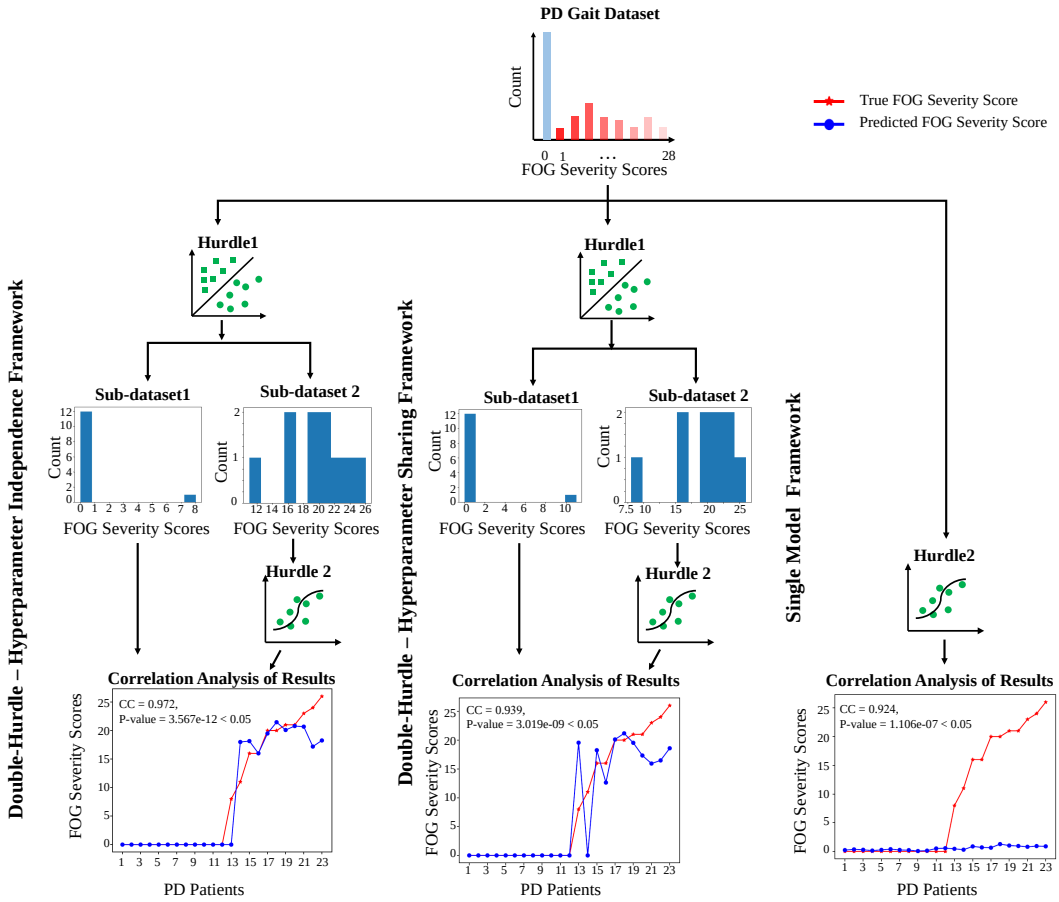


Fig. 6: Results from interpretability study for the double-hurdle model. The top part visualizes the zero-inflated problem in the PD gait dataset. The first hurdle divides the PD gait dataset with imbalanced samples into two sub-datasets with a relatively balanced sample distribution. The middle part illustrates the distribution of samples in the two sub-datasets. The bottom part, from left to right, shows the performance of the double-hurdle model–hyperparameter independence framework, the double-hurdle model–hyperparameter sharing framework, and the single model framework in quantitatively assessing the FOG severity of PD patients.

TABLE VI: The structure of deep learning algorithms

Algorithms	Module1	Module2	Module3	Module4
CNN-LSTM[1]	ConvBlock1 + ConvBlock2	2×LSTM	Pool&Flatten + FC1	FC2 FC3
ResNet[2]	ConvBlock1	2×ResBlocks	Pool&Flatten	
TCNet[3]	ConvBlock1	2×TCBlocks		
FT-Transformer[4]	FT-Transformer Module			

Convolutional Networks (TCN) [38], and Feature Tokenizer - Transformer (FT-transformer) [39]. These algorithms are widely used in the field of IGA [32]. The structures of the deep learning algorithms based on multi-task learning are illustrated in Table VI. ConvBlock comprises 1D convolutional layer with a kernel size of 3, followed by 1D BatchNorm, ReLU, and Max-Pooling with a kernel size of 2. The input and output sizes of the 1D convolutional layers are as follows: 1×16 for ConvBlock1, 16×32 for ConvBlock2, and 32×16 for ConvBlock3. The LSTM layer has an input size of 6 and a hidden size of 16. The ResBlock replicates the configurations of ConvBlock2 and ConvBlock3. In TCBlock, there are two dilated causal convolutions with the input and output sizes of 16×32 and 32×16, the kernel size of 3 and 3, the dilation sizes of 1 and 2, and the padding sizes of 1 and 2, respectively. The FT-Transformer is implemented using the rtdl package, with 25 input features and an output dimension of 16. Pool&Flatten consists of an 1D AdaptiveAvgPool layer with an output dimension of 1 and a Flatten layer. Table VI illustrates three fully connected layers: FC1, FC2, and FC3, with input and output sizes of 32×16, 16×2, and 32×1, respectively. FC2 is utilized for the classification task, while FC3 is employed for the regression task.

We optimize deep learning algorithms using the Adam optimizer with a default learning rate of 1e-3. Cross-entropy loss is used for the classification task, while Mean Squared Error (MSE) loss is employed for the regression task. For multi-task learning, we combine both loss functions. Performance evaluation is conducted via five times five-fold cross-validation. All deep learning algorithms are conducted on a workstation equipped with an Intel Core i7-8700 CPU with 6 cores and a base frequency of 3.20 GHz. The training of deep learning algorithms is performed using PyTorch version 1.13.1 and Python version 3.8.16.

Authorized licensed use limited to: INSTITUTE OF AUTOMATION CAS. Downloaded on May 21,2024 at 00:48:42 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Biomedical Engineering. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TBME.2024.3402677

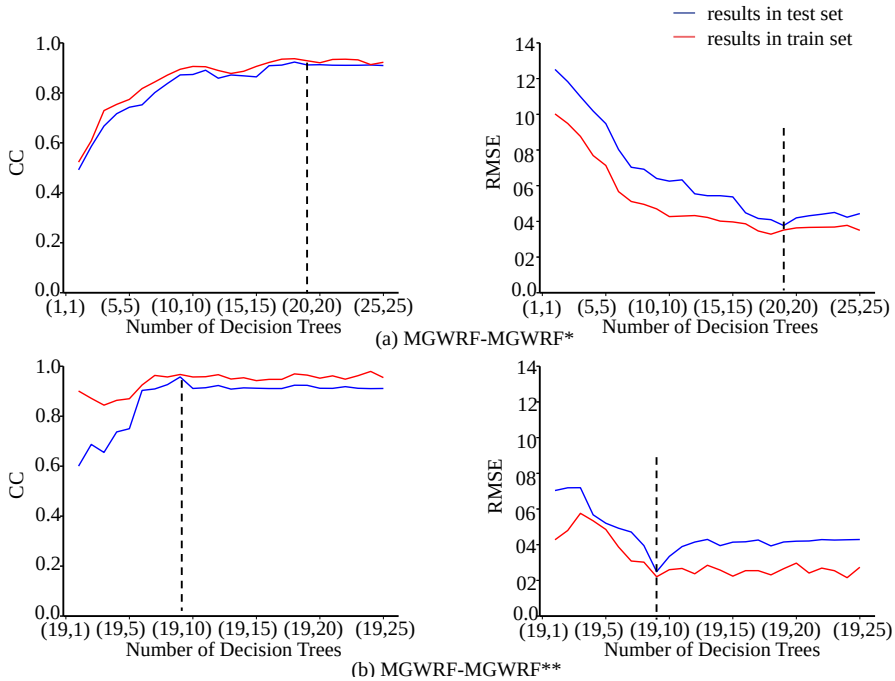


Fig. 7: Results from trade-off analysis of double-hurdle model. There are two subplots (a, b) showcasing the validation curves for MGWRF-MGWRF in the hyperparameter sharing framework and MGWRF-MGWRF in the hyperparameter independence framework. The x-axis of a and b is composed of pairs of data, where the first position of the data represents the number of decision trees in the first hurdle of the double-hurdle model, and the second position represents the number of decision trees in the second hurdle. When the number of decision trees reaches the dotted line, the model achieves its optimal performance.

Table V summarizes the experiment results for all algorithms across different experiment types. Our MGWRF-MGWRF model outperforms deep learning algorithms in classification and achieves lower RMSE values. This suggests that deep learning algorithms may not fully utilize their feature extraction capabilities for FOG severity assessment. Furthermore, deep learning algorithms based on double-hurdle models exhibit less effective performance on both tasks, due to their higher model complexity. In contrast, the proposed MGWRF-MGWRF model excels at learning features from small datasets, resulting in superior performance in identifying PD patients with FOG and quantifying the FOG severity of all PD patients.

F. Interpretability Study of Double-Hurdle Model

To illustrate the advantages of the double-hurdle model with an additional classification hurdle, we introduce an interpretable study for the model, as shown in Fig.6. From the middle part of Fig.6, whether in the hyperparameter sharing framework or the hyperparameter independence framework, the first hurdle can effectively identify PD patients with FOG symptoms and divide the PD gait dataset into two sub-datasets with a relatively balanced sample distribution: the sub-dataset1 for Parkinson's patients without FOG and the sub-dataset2 for those with FOG. In the sub-dataset1, the severity scores of FOG symptoms are concentrated at 0 points. In the sub-dataset2, the distribution of severity scores for FOG symptoms is relatively balanced than the PD gait dataset. Therefore, the double-hurdle model can significantly improve the quality of the dataset, alleviating the impact of the zero-inflated problem on quantitatively scoring FOG severity in PD patients.

From the bottom of Fig.6, it can be observed that, in both the hyperparameter sharing framework and the hyperparameter independence framework, the double-hurdle model outperforms the single hurdle model in assessing FOG severity of PD patients. This further indicates that the additional classification hurdle can help the double-hurdle model effectively address the impact of zero-inflated problem on the quantitative assessment of FOG severity.

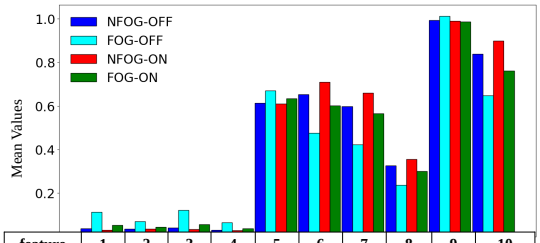
G. Trade-off Analysis of Double-Hurdle Model

The validation curve is utilized to explore the trade-off between the complexity and performance of the double-hurdle model. Complexity is measured by the number of decision trees in the model, while performance is assessed using RMSE and CC metrics. We perform 5 iterations of 5-fold cross-validation and compute the mean of these results to represent the performance of the double-hurdle model. The trade-off analysis results for two models, MGWRF-MGWRF in the hyperparameter independence framework and MGWRF-MGWRF in the hyperparameter sharing framework, are depicted in Fig.7.

It is evident that under the hyperparameter sharing framework, MGWRF-MGWRF achieves optimal performance when the number of decision trees is set to (19, 19), resulting in an RMSE of 3.767 and a CC of 0.939, as shown in Fig.7. Therefore, when performing the trade-off analysis of the double-hurdle model under the hyperparameter independence

Authorized licensed use limited to: INSTITUTE OF AUTOMATION CAS. Downloaded on May 21,2024 at 00:48:42 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Biomedical Engineering. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TBME.2024.3402677



feature	1	2	3	4	5	6	7	8	9	10
NFOG-OFF	0.037	0.035	0.041	0.030	0.613	0.653	0.597	0.326	0.992	0.838
FOG-OFF	0.113	0.070	0.122	0.064	0.669	0.475	0.421	0.237	1.012	0.647
NFOG-ON	0.003	0.036	0.033	0.029	0.610	0.709	0.660	0.354	0.990	0.898
FOG-ON	0.053	0.045	0.056	0.038	0.634	0.601	0.564	0.300	0.986	0.761

Fig. 8: Bar plot illustrates the effects of drug state on gait features of PD patients with or without FOG. The horizontal axis of the bar plot represents ten gait features. The features denoted by numbers 1 to 10 correspond to the following gait spatiotemporal features: Stride Length Variability, Swing Time Variability, Step Length Variability, Single Limb Support Time Variability, Stance Phase, Stride Length, Speed, Step Length, Stance Time Asymmetry, and Walk Ratio, respectively. The vertical axis represents the mean value of each gait features in the corresponding group.

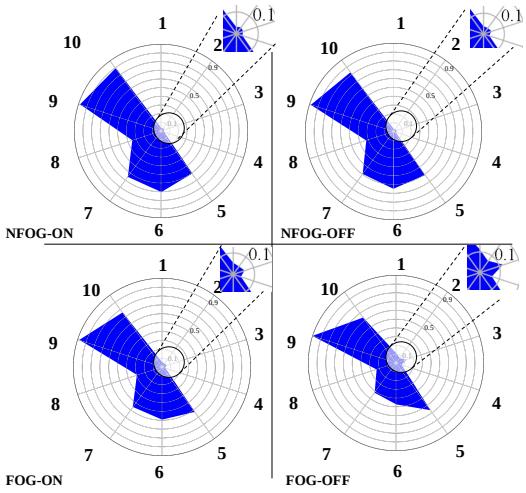


Fig. 9: Radar plots illustrate the effects of drug state on the gait patterns of PD patients with or without FOG. The features denoted by numbers 1 to 10 are the same as those from the Fig.8.

framework, we adjust only the number of decision trees in the second hurdle while keeping the number of decision trees in the first hurdle constant at 19. In the hyperparameter independence framework, the double-hurdle model achieves optimal performance when the number of decision trees reaches (19, 9), obtaining a CC of 0.967 and an RMSE of 2.488. Due to its more flexible hyperparameter combination schemes, the independence framework demonstrates superior performance in FOG severity assessment compared to the hyperparameter sharing framework. In conclusion, the proposed double-hurdle model with its unique structure effectively solves the zero-inflated problem, achieving the most optimal performance in the FOG severity assessment of PD patients.

H. Effect of Drug State on Gait Patterns of PD Patients

Based on the above experiment results, we further study the effect of drug state on PD patients with or without FOG, which can provide more valuable perspectives for treating and managing PD. In each iteration of five-fold cross validation, we obtain five gait feature sets using MRMR algorithm. Each gait feature set has k gait features and their importance scores. Here, we aggregate the gait features from the all gait feature sets and select the 10 gait features with the top importance scores. These selected features encompass Stride Length Variability, Swing Time Variability, Step Length Variability, Single Limb Support Time Variability, Stance Phase, Stride Length, Speed, Step Length, Stance Time Asymmetry, and Walk Ratio. Notably, the selected features are consistent with the results in the previous literature [31]. Based on the medication states, the patients from the public PD gait database are categorized into four groups: "NFOG-ON", "FOG-ON", "NFOG-OFF", and "FOG-OFF", where "NFOG-ON" refers to PD patients without FOG in the "ON" drug state. Fig. 8 illustrates the mean values of various gait features across different patient groups. Compared to the "FOG-ON" group, the "NFOG-ON" group exhibits a significant increase of 17.97% in Stride Length, 17.02% in Speed, 18% in Step Length, and 18% in Walk Ratio. Conversely, Stride Length Variability, Swing Time Variability, Step Length Variability, and Single Limb Support Variability decrease by 94.34%, 20%, 41.07%, and 23.68%, respectively. These findings suggest that FOG symptom leads to decreased pace gait features and increased gait variability of PD patients, resulting in reduced walking consistency and increased risk of falls. This observation aligns with prior research findings documented in the existing literature [31]. Furthermore, "OFF" drug state makes FOG symptoms more visible in PD patients.

This paper utilizes the radar chart to visually depict the effects of drug state on PD patients with or without FOG, shown in Fig. 9. Additionally, we apply the Procrustes algorithm to calculate the similarity between gait patterns of different groups under various medication states [35]. A lower similarity score indicates a higher resemblance between two gait shapes. In the "OFF" state, the gait shape similarity between the "NFOG-OFF" and "FOG-OFF" groups is 0.034. Conversely, in the "ON" state, the gait shape similarity between the "NFOG-ON" and "FOG-ON" groups is 0.008, which is less than 0.034. These findings suggest that in the "OFF" state, the gait patterns of PD patients with FOG significantly differ from those without FOG, more prominently than in the "ON" state. Therefore, the "OFF" state accentuates the abnormality in gait shape, making FOG symptoms more discernible. In conclusion, our study provides evidence that the "OFF" state amplifies the visibility of FOG symptoms in PD patients, highlighting the importance of considering medication states when assessing FOG severity in PD.

VI. Discussion and Conclusion

This paper proposes a double-hurdle model for assessing FOG severity of PD patients. Firstly, our study quantifies the

Authorized licensed use limited to: INSTITUTE OF AUTOMATION CAS. Downloaded on May 21,2024 at 00:48:42 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Biomedical Engineering. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TBME.2024.3402677

FOG severity using gait data from PD patients during periods without FOG episodes. This approach simplifies experimental setups and reduces the risk of falls to PD patients compared to previous studies [23]–[26]. Secondly, unlike prior works that classify FOG severity into binary or five-class categories, our study offers a finer granularity assessment on a score scale from 0 to 28. Finally, using histogram and radar chart techniques, we demonstrate that FOG correlates with reduced gait pace and increased gait variability, consistent with findings in prior literature [31]. Additionally, the proposed MGWRF-MGWRF outperforms other machine learning algorithms and deep learning algorithms in identifying PD patients with FOG and quantifying the FOG severity of all PD patients [36]–[39], as detailed in Section V.

However, the relatively smaller number of patients may limit the generalizability of our study. In the future, more gait data from PD patients will be collected to verify further the double-hurdle model's effectiveness. Furthermore, we will incorporate data from additional modalities to enrich the evaluation of FOG in PD patients, offering a more profound understanding of FOG symptoms.

REFERENCES

[1] M. Gilat *et al.*, "A systematic review on exercise and training-based interventions for freezing of gait in Parkinson's disease", *NPJ. Parkinsons Dis.*, vol. 7, no. 1, pp. 1-18, 2021.

[2] Y. Yang *et al.*, "Artificial intelligence-enabled detection and assessment of Parkinson's disease using nocturnal breathing signals", *Nature Medicine*, vol. 28, no. 10, pp. 2207-2215, 2022.

[3] M. J. Armstrong and M. S. Okun, "Diagnosis and treatment of Parkinson disease: A review", *Jama*, vol. 323, no. 6, pp. 548-560, 2020.

[4] L. Pepa *et al.*, "A fuzzy logic system for the home assessment of freezing of gait in subjects with Parkinsons disease", *Exp. Syst. Appl.*, vol. 147, 2020, Art. no. 113197.

[5] R. Ou *et al.*, "Freezing of gait in Parkinson's disease with glucocerebrosidase mutations: prevalence, clinical correlates and effect on quality of life", *Front. Neurosci.*, vol. 17, 2023, Art. no. 1288631.

[6] J. Mei *et al.*, "Machine learning for the diagnosis of parkinson's disease: A review of literature", *Front. Aging. Neurosci.*, vol. 13, 2021, Art. no. 633752.

[7] V. Skaramagkas *et al.*, "Multi-modal deep learning diagnosis of Parkinson's disease: A systematic review", *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, no. 1, pp. 2399-2423, 2023.

[8] C. Wang *et al.*, "A hierarchical architecture for multi-symptom assessment of early Parkinsons disease via wearable sensors", *IEEE Trans. Cognit. Develop. Syst.*, vol. 14, no. 4, pp. 1553-1563, 2021.

[9] C. Gao *et al.*, "Freezing of gait in Parkinsons disease: Pathophysiology risk factors and treatments", *Transl. Neurodegener.*, vol. 9, 2020, Art. no. 12.

[10] X. Gu *et al.*, "Beyond Supervised Learning for Pervasive Healthcare", *IEEE Rev. Biomed. Eng.*, vol. 17, no. 1, pp. 42-62, 2024.

[11] Y. Zhang *et al.*, "Deep Long-Tailed Learning: A Survey", *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 9, pp. 10795-10816, 2023.

[12] C. Zhang *et al.*, "Short-Term Electricity Price Forecast Using Frequency Analysis and Price Spikes Oversampling", *IEEE Trans. Power Syst.,* vol. 38, no. 5, pp. 4739-4751, 2023.

[13] E. Troullinou *et al.*, "A Generative Neighborhood-Based Deep Autoencoder for Robust Imbalanced Classification", *IEEE Trans. Artif. Intell.*, vol. 5, no. 1, pp. 80-91, 2024.

[14] Y. Yang *et al.*, "Artificial intelligence-enabled detection and assessment of Parkinson's disease using nocturnal breathing signals", *Nat. Med.*, vol. 28, no. 1, pp. 2207-2215, 2022.

[15] M. Steininger *et al.*, "Density-based weighting for imbalanced regression", *Mach. Learn.*, vol. 110, no. 8, pp. 2187-2211, 2021.

[16] J. Liang *et al.*, "A Poisson-Based Distribution Learning Framework for Short-Term Prediction of Food Delivery Demand Ranges", *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 12, pp. 14556-14569, 2023.

[17] Y. Zhang and Q. Yang, "A survey on multi-task learning", *IEEE Trans. Knowledge Data Eng.*, vol. 34, no. 12, pp. 5586-5609, 2021.

[18] Z. Chen *et al.*, "Technique of Feature Extraction Based on Interpretation Analysis for Multilabel Learning in Nonintrusive Load Monitoring With Multiappliance Circumstances", *IEEE Trans. Ind. Informat.,* doi: 10.1109/TII.2023.3341244.

[19] H. Peng *et al.*, "Feature selection based on mutual information: Criteria of max-dependency max-relevance and min-redundancy", *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 8, pp. 1226-1238, 2005.

[20] Y. Wang *et al.*, "Feature Selection With Maximal Relevance and Minimal Supervised Redundancy", *IEEE Trans. Cybern.*, vol. 53, no. 2, pp. 707-717, 2023.

[21] S. Carta *et al.*, "Statistical arbitrage powered by explainable artificial intelligence ", *Exp. Syst. Appl.*, vol. 206, 2022, Art. no. 117763.

[22] Y. Guo *et al.*, "Detection and assessment of Parkinson's disease based on gait analysis: A survey", *Front. Aging Neurosci.*, vol. 14, 2022, Art. no. 837.

[23] L. Sigcha *et al.*, "Improvement of Performance in Freezing of Gait detection in Parkinson's Disease using Transformer networks and a single waist-worn triaxial accelerometer", *Eng. Appl. Artif. Intell.*, vol. 116, 2022, Art. no. 105482.

[24] L. Borzì *et al.*, "Real-time detection of freezing of gait in Parkinson's disease using multi-head convolutional neural networks and a single inertial sensor", *Artif. Intell. Med.*, vol. 135, 2023, Art. no. 102459.

[25] D. M. Tan *et al.*, "Freezing of gait and activity limitations in people with Parkinson's disease", *Arch. Phys. Med. Rehabil.*, vol. 92, no. 7, pp. 1159-1165, 2011.

[26] Y. Zhang *et al.*, "Prediction of freezing of gait in patients with Parkinson's disease by identifying impaired gait patterns", *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 3, pp. 591-600, 2020.

[27] S. Aich *et al.*, "A validation study of freezing of gait (FoG) detection and machine-learning-based FoG prediction using estimated gait characteristics with a wearable accelerometer", *Sensors*, vol. 18, no. 10, 2018, Art. no. 3287.

[28] H. Park *et al.*, "Classification of parkinson's disease with freezing of gait based on 360° turning analysis using 36 kinematic features", *J. Neuroeng. Rehabil.*, vol. 18, no. 1, 2021, Art. no. 177.

[29] H. Kwon *et al.*, "An explainable Spatial-Temporal graphical convolutional network to score freezing of gait in parkinsonian patients", *Sensors*, vol. 23, no. 4, 2023, Art. no. 1766.

[30] T. Shida *et al.*, "A public data set of overground walking full-body kinematics and kinetics in individuals with Parkinson's disease", *Front. Neurosci.*, vol. 17, 2023, Art. no. 992585.

[31] B. Intzandt *et al.*, "The effects of exercise on cognition and gait in Parkinson's Disease: a scoping review", *Neurosci Biobehav Rev.*, vol. 95, no. 1, pp. 136-169, 2018.

[32] R. Kaur et al., "Predicting Multiple Sclerosis From Gait Dynamics Using an Instrumented Treadmill: A Machine Learning Approach," *IEEE Trans. Biomed. Eng.*, vol. 68, no. 9, pp. 2666-2677, Sept. 2021.

[33] M.I.A. Ferreira *et al.*, "Machine learning models for Parkinson's disease detection and stage classification based on spatial-temporal gait parameters," *Gait Posture*, vol. 98, pp. 49-55, 2022.

[34] A. Seifert *et al.* "Doppler Radar for the Extraction of Biomechanical Parameters in Gait Analysis", *IEEE J. Biomed. Health Informat.*, vol. 25, no. 2, pp. 547-558, 2021.

[35] D. Sethi *et al.*, "A comprehensive survey on gait analysis: History parameters approaches pose estimation and future work", *Artif. Intell. Med.*, vol. 129, 2022, Art. no. 102314.

[36] T. Sainath *et al.*, "Convolutional, Long Short-Term Memory, fully connected Deep Neural Networks," in *Proceedings of the 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, South Brisbane, Australia, pp. 4580-4584, 2015.

[37] K. He *et al.*, "Deep Residual Learning for Image Recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, USA, pp. 770-778, 2016.

[38] S. Bai *et al.*, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling." *arXiv:1803.01271*, 2018.

[39] Y. Gorishniy *et al.*, "Revisiting deeplearning models for tabular data," in *Proceedings of the 35-th conference on advances in neural information processing systems (NeurIPS)*, Curran Associates, Inc., 2021.

Authorized licensed use limited to: INSTITUTE OF AUTOMATION CAS. Downloaded on May 21,2024 at 00:48:42 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.