
Article:
Zhao, Y., Wang, X., Peng, X. et al. (8 more authors) (2024) Selecting and evaluating key MDS-UPDRS activities using wearable devices for Parkinson's disease self-assessment. IEEE Journal of Selected Areas in Sensors, 1. pp. 177-189. ISSN 2836-2071

https://doi.org/10.1109/JSAS.2024.3432714

Selecting and Evaluating Key MDS-UPDRS Activities Using Wearable Devices for Parkinson's Disease Self-Assessment

Yuting Zhao[1], Xulong Wang[2*], Xiyang Peng[2], Ziheng Li[1], Fengtao Nan[3], Menghui Zhou[2], Jun Qi[4], Yun Yang[1], Zhong Zhao[5], Lida Xu[6],*Fellow,IEEE*, Po Yang[2*],*Senior Member,IEEE*

***Abstract*— Parkinson's disease (PD) is a complex neurodegenerative disease in the elderly. This disease has no cure, but assessing these motor symptoms will help slow down that progression. Inertial sensing-based wearable devices (ISWDs) such as mobile phones and smartwatches have been widely employed to analyse the condition of PD patients. However, most studies purely focused on a single activity or symptom, which may ignore the correlation between activities and complementary characteristics. In this paper, a novel technical pipeline is proposed for fine-grained classification of PD severity grades, which identify the most representative activities. We also propose a multi-activities combination scheme based on MDS-UPDRS. Utilizing this scheme, symptom-related and complementary activities are captured. We collected 85 PD subjects of different severity grades using a single wrist sensor. Our best results demonstrate F1 scores of 95.75% for PD diagnosis and the fine-grained classification accuracy of PD disease grade is 82.41% when combing 4 activities which improved by 11.02% over a single activity. The experiments and theoretical analyses can serve as a useful foundation for future investigations into the effect of proposed solutions for PD diagnosis in uncontrolled environment setup, ultimately leading to self-PD assessment in the home environment.**

***Index Terms*— Feature selection, inertial sensing-based wearable devices, machine learning, multi-activities, parkinson's disease**

I. Introduction

PARKINSON'S disease (PD) is the second most prevalent neurodegenerative disorder globally, with a significant impact on over 6 million individuals [1], [2]. This condition is characterised by various motor impairments, including tremor, bradykinesia, muscle rigidity, and postural instability [3]. Nowadays, no disease-curing therapy for PD, but dopamine replacement therapies such as Levodopa or L-Dopa are still able to provide relief for controlling abnormal movements as they have been since its discover [4], [5]. PD patients' quality of life is directly enhanced by precise titration of medications by clinicians [6]. To precisely provide precise medication titration and assess the related motor impairments, the MDS-Unified Parkinson's Disease Rating Scale (MDS-UPDRS) [7] has been widely employed. However, MDS-UPDRS has some drawbacks, such as 1)inter-rater and intra-rater score variability, 2)subjectivity and 3)time-consuming. The scale only delivers a static snapshot view of the patient during the clinical visits. Therefore, it remains a challenge for clinicians to effectively track motor symptoms during the out-of-clinic time.

With the popularisation of various inertial sensing-based wearable devices(ISWDs) and the improvement of machine learning(ML) technologies, mobile sensing-based PD diagnosis approaches objectively estimate motor symptoms. Several studies [8]–[15] have demonstrated that ISWDs are able to support uncontrolled activity assessment of PD progression and symptoms effectively. Inertial sensor recording, ML methods' analysis and decision-making offer significant insight into achieving continual, precise PD motor assessment and treatment of disease progression and medication titration.

Despite the advancements of those techniques, there is limited understanding of the connection between the data from wearable sensors and the information on potential clinical motor symptoms of PD patients with different severity grades. Firstly, limitations of the application scenario. To ensure the consistency between experimental results and clinical findings, most studies of PD severity grade diagnosis focus on clinical or specific controlled laboratory scenarios. For instance, many studies [16], [17] scored patients on the MDS-UPDRS in clinical scenarios and required them to wear inertial sensors, which were employed to record signals of tremor, bradykinesia and other motor symptoms. This work demonstrates the high correlation of data from inertial sensors with clinical motor symptoms. If the above studies focus on PD motor detection in the controlled scenario, then in the uncontrolled environment the focus slides towards PD daily monitoring. Many studies [9], [18] identified the advantages of smartwatches for remote patient monitoring and found mobile and wearable technolo-

This research was supported by the National Natural Science Foundation of China (No. 62061050, No. 62301452) and the 13th Postgraduate Research Innovation Project of Yunnan University 2021Y397.

[1]National Pilot School of Software, Yunnan University, China, {zhaoyuting,liziheng9050}@mail.ynu.edu.cn,yangyun@ynu.edu.cn

[2]Department of Computer Science,University of Sheffield,UK, {xl.wang,xpeng24,menghui.zhou,po.yang}@sheffield.ac.uk

[3]College of information engineering, Northwest A&F University, China ,fengtao.nan@nwsuaf.edu.cn

[4]Department of Computing, Xi'an JiaoTong-Liverpool University, China , jun.qi@xjtlu.edu.cn

[5]Yunnan First People's Hospital, China , wasx-1128new@163.com

[6]Old Dominion University, US, lxu@odu.edu

*Corresponding authors xl.wang@sheffield.ac.uk; po.yang@sheffield.ac.uk

gies are more than capable of tracking motor fluctuations in PD. In contrast to PD daily monitoring, there have been few studies that have assessed the role of timely self-assessment of PD motor symptoms using wearable devices. In contrast to PD daily monitoring, there have been few studies that have assessed the role of timely self-assessment of PD motor symptoms using wearable devices.

Secondly, limitations to research on a single symptom or activity. Many studies hold a preference for one motor symptom. For instance, Rigas et al. [19] accessed the tremor type (resting/action postural) and symptom severity. Ullrich et al. [10] proposed an unsupervised standardized detection for gait symptoms at home. To impose a minimal constraint on patients, Kim et al. [20] quantified the bradykinesia via the clinical motor of finger taps using a gyrosensor sensor. Giubert et al. [21] analysed sit-to-stand task on the MDS-UPDRS scale and emphasised leg agility task play a key role in the PD diagnosis. However, specialising in a single symptom to assess PD may suffer from inadequate evaluation results. Even when clinicians assess PD severity using the MDS-UPDRS scale, they also need to consider multiple activities in combination. In the MDS-UPDRS scale for *constancy of rest tremor*(item 3.18), for example, the raters are required to examine all movements before scoring this item, which implies that there exists a potential clinical connection between symptoms and motors and that motors assessment alone is not recommended. Additionally, more sensors and performing more activities would enable get more accurate PD severity diagnosis, but this is not practical in the uncontrolled scenario. Rigas et al. [19] utilised 6 sensors to assess tremor and Ricci et al. [22] employed 12 sensors throughout the body to estimate the effectiveness of Levodopa treatment. Sensors all over the body and too many compulsory activities impose a heavy burden on the patient and reinforce the erasure of wearable applications in uncontrolled environments.

Targeting at above-mentioned issues, this paper focuses on finding the representative MDS-UPDRS activities with a single wearable sensor in uncontrolled environments. More precisely, we firstly collected 85 PD subjects of different severity grades and 70 healthy controls, each subject performed the 14 activities within the part-III of MDS-UPDRS scale and was scored by the movement disorders neurologist. Meanwhile, only one single device where on the wrist was utilised to record the activities of subjects. We then explored the potential correlations of all the activities and thereby identified the most representative ones. These activities provide the highest amount of motor symptoms' clues of all activities. Afterwards, we selected the most significant features most relevant to the disease and conducted experiments using our proposed activity combination method. Utilizing these, we were able to accurately classify PD severity grades on a fine-grained scale. Furthermore, as illustrated in Fig.1, we proposed a novel technical pipeline for identifying the most representative activities in uncontrolled environments. Selecting the fewest activities to reduce the burden on patients. Notably, instead of designing complex fusion algorithms based on multiple activities, we focused on capturing the most critical features across all activities, which will serve as a useful insight for the development of lightweight wearable devices in an uncontrolled environment. Overall, we make the following contributions:

- A novel technical pipeline is proposed for classifying PD severity grades in a fine-grained manner and a simple yet effective multi-activity combination scheme is designed for analysing complementary and redundant activities in MDS-UPDRS.
- The most representative activities from the MDS-UPDRS scale and their symptom-related features are provided to enable effective PD self-assessment in the uncontrolled environment.
- A large-scale data collection is implemented which includes 14 activity data from 85 PD and 70 healthy individuals which have been labelled by neurologists.

The rest of this paper is arranged as follows: Section II summarizes the related work in this research field. Section III describes participants and data acquisition. Section IV describes the methods used in this work. We discuss the results of our research in section V. Finally, section VII summarizes this paper and puts forward the future prospects.

TABLE I
UPDRS PARADIGM ACTIVITIES

Num	Activity name	Abbr.
1	Finger taps	FT
2	Clench and open alternately	COA
3	Rapid alternating movements of hands	ALTER
4	Hand rotation-right	HR-R
5	Hand rotation-left	HR-L
6	Finger to nose-left	FN-L
7	Finger to nose-right	FN-R
8	Standing with arms hold	STANDH
9	Walk back and forth	WA
10	Arising from chair	AC
11	Drinking water	DRINK
12	Picking things	PICK
13	Sitting	SIT
14	Standing	STAND

II. RELATED WORK

A. Single Activity/Symptom-based Diagnosis

There are many studies on PD, but most of them are based on a single activity or symptom and single symptom, unfortunately, the provided information does not give rise to a severity assessment of the condition. Guo et al. [23] conducted a study in a laboratory setting where they gathered walking data from 10 individuals diagnosed with PD. The purpose of the study was to ascertain the freezing of gait by utilizing the freezing index. However, this research only detected a single motor symptom. Pérez-Ibarra et al. [24] conducted a data collection study involving 5 healthy individuals and 7 PD. The participants were instructed by a professional to walk both on a treadmill and the floor. They devised a novel, real-time adaptive unsupervised algorithm to detect and classify gait events and phases using a solitary IMU mounted at the posterior aspect of the foot. Luis Sigcha et al. [25] leveraged the inertial sensors integrated into mainstream

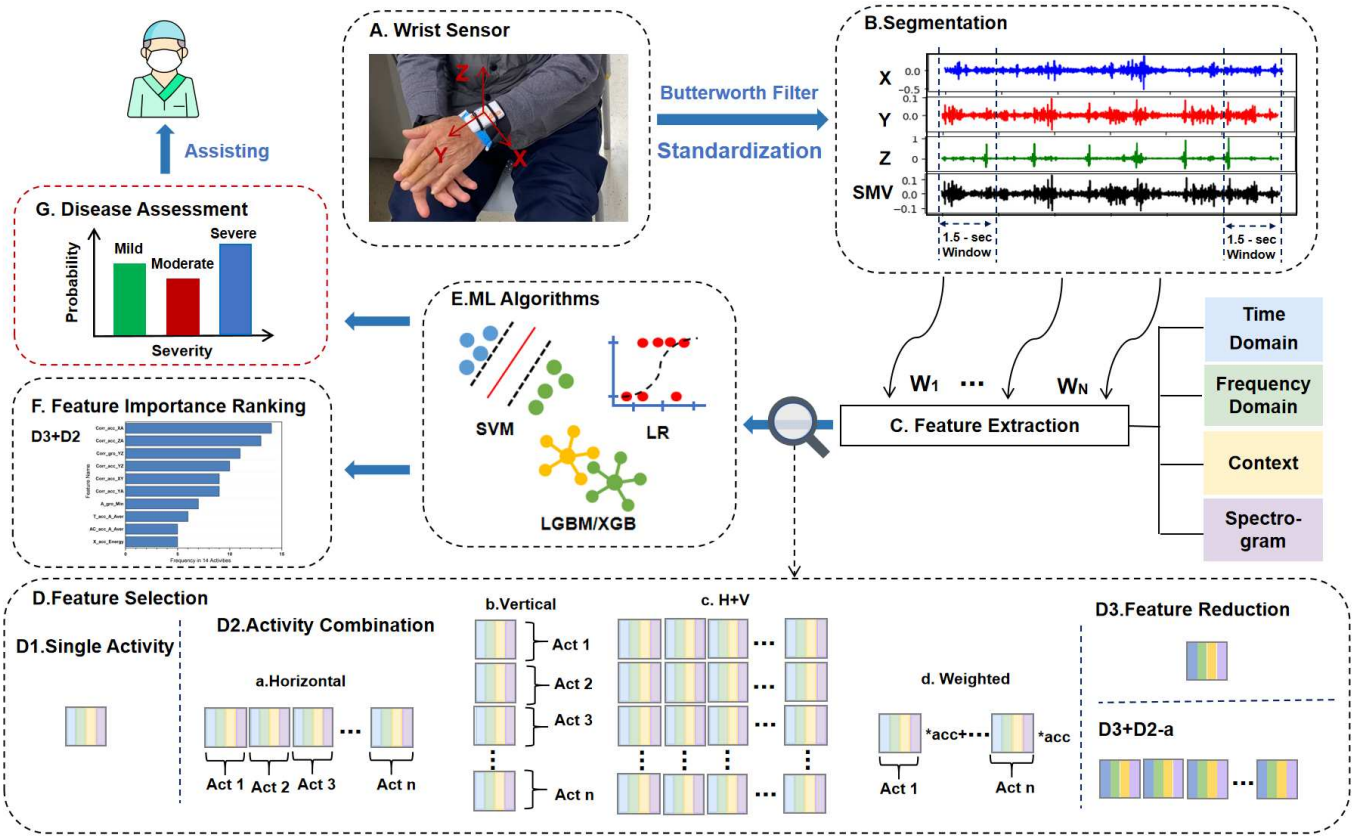


Fig. 1. Technical pipeline of our approach. Our core technical design is module D. Taking 174-dimensional features as an example, the horizontal concatenating directly concatenates Activity 1 features and Activity 2 features, increasing the dimension of features. At this time, the concatenated feature is 174+174=348 dimensions. The vertical concatenating adopts the form of upper and lower concatenating to concatenate features, and its features are converted from the original one-dimensional data (1 * 174) to (2 * 174). Feature fusion weights the features of Activity 1 and Activity 2, and their feature dimensions are consistent with the original activity dimensions.

consumer smartwatches and employed diverse ML models to detect and evaluate the intensity of bradykinesia in the superior extremities. The study encompassed six participants diagnosed with disease and seven healthy individuals of similar age, throughout a minimum period of six weeks, all participants were furnished with a consumer smartwatch and instructed to engage in a prescribed set of motor exercises. These studies only used a single activity, which may have overlooked the correlation between multiple activities, and they tested for one of the symptoms without providing an overall assessment of the condition.

B. Specific Scenario-based Diagnosis

The application scenarios can be divided into clinical, laboratory and free-living environments. The literature has examined a few potential PD biomarkers, among which cerebrospinal fluid-blood biomarkers 16 and neuroimaging 17 have demonstrated high. However, these biomarkers are costly, and invasive and require access to specialized medical centres. The researchers [26], [27] have explored several potential biomarkers for PD, including cerebrospinal fluid biomarkers, blood biochemical markers, and neuroimaging techniques. These biomarkers have demonstrated good accuracy in diagnosing PD. However, they often come with drawbacks such as high cost, invasiveness, and the need for access to specialized medical centres. Most of the research on the diagnosis of PD is carried out in the laboratory environment, As PD movement symptoms are diverse and fluctuating, requiring monitoring over multiple periods, another emerging research trend in PD assessment is the evaluation of PD symptoms and progression in a free-living environment. But still mainly focus on distinguishing between normal and PD patients, and single symptom assessment of PD. Chen et al. [8] developed an automated disease assessment framework that utilized smartphone sensors to classify PD patients and healthy individuals based on signals extracted during the performance of specific activities at home. However, the framework does not consider the abnormalities that may arise during the execution of these activities in a home environment. To address this issue and minimize anomalies during data collection in a home setting, further measures need to be implemented. Erb et al. [9], [28] put forth a scheme where patient logs were filled out by caregivers to monitor the daily activities, PD symptoms, and medication usage of patients. However, relatives often have a limited understanding of PD symptoms and struggle to accurately identify motor symptoms, resulting in misconceptions and mistakes in the recorded data. This highlights the need for improved caregiver education and additional methods to facilitate accurate symptom identification in order to enhance the reliability and usefulness of patient logs for tracking PD-

related information. Martin Ullrich et al. [10] conducted a data collection study involving 12 patients diagnosed with idiopathic PD. Over a span of two weeks, the patients were equipped with inertial measurement units to capture data. In addition to their routine daily activities, the patients performed a sequence of three successive 4 × 10-meter-walking tests at varying walking speeds. Griffiths et al. [29] conducted a study involving 34 patients with PD and 10 age-matched healthy controls. Data collection spanned a minimum duration of 10 days. The researchers utilized a wearable device called the Parkinson's Kinetigraph (PKG), which was worn like a wristwatch and incorporated a 3-axis iMEMS accelerometer. The PKG was employed to evaluate dyskinesia and bradykinesia, two characteristic symptoms of PD. Hammerla et al. [11] gathered data from 34 patients with PD for a minimum of 6 days within a home environment. They utilized movement sensors equipped with a tri-axial accelerometer and employed the RBM with a deep learning algorithm for training purposes. Most research datasets are collected for intervention guidance in a controlled environment, However, challenges in extending such schemes to out-of-hospital environments, the accurate identification of PD motor symptoms relies primarily on the quality and amount of recorded data. The variability of PD motor symptoms can significantly differ between individuals, and activities, and also change over time [30].

C. Specific Data Source based Diagnosis

Currently, the data sources used for diagnosing PD can be categorized into three main types: sensor-based, video-based, and wireless signal-based. With advancements in wearable motion sensors, the evaluation of PD based on patients' movement patterns has emerged as a prominent research area. Ricci et al. [22] conducted a study involving 36 PD patients who wore 14 sensors while performing seven activities from the MDS-UPDRS. The patient's condition was evaluated using an SVM before and after medication intake to determine the stage of their disease. It is crucial to highlight that this study was conducted in a controlled laboratory environment, and potential factors such as abnormal activity during task performance and the potential stress experienced by patients wearing multiple sensors were not taken into consideration. Video-based deep learning algorithms have been increasingly utilized for the assessment of the MDS-UPDRS and other evaluation methods. Weiping Liu et al. [31] captured two videos obtained for the 3.17 and 3.15 tests to quantify PD tremor severity. Mandy Lu et al. [32]proposed the use of an ordinal focal neural network to estimate the MDS-UPDRS scores from input videos. In their study, they collected video recordings of gait and finger tapping, taking into account the ordinal nature of the MDS-UPDRS scores and addressing the challenge of class imbalance. By leveraging this approach, they aimed to improve the accuracy and reliability of evaluating PD severity based on video data. The third category is wireless signal. Yuzhe Yang et al. [33] conducted research on the identification and evaluation of PD using AI and nocturnal breathing signals. In their study, they explored the possibility of collecting breathing signals without the need for wearable devices. This was achieved by transferring a low-power radio signal and assessing the reflections of this signal off the person's body. By leveraging AI techniques, they aimed to develop a non-invasive and convenient method for detecting and assessing PD based on these breathing signals. Wearable devices' increasing popularity stems from their ease of operation, affordability, and ability to function in harsh environments without compromising others' privacy. The video-based method requires infrastructure support as it requires the installation of cameras in monitoring locations and is highly dependent on lighting. The wireless respiratory signal instrument has high environmental requirements, but it has disadvantages such as low detection accuracy and easy to be submerged by noise, the respiratory detection method is only applicable to a small indoor range and cannot be applied in outdoor scenes.

III. Materials

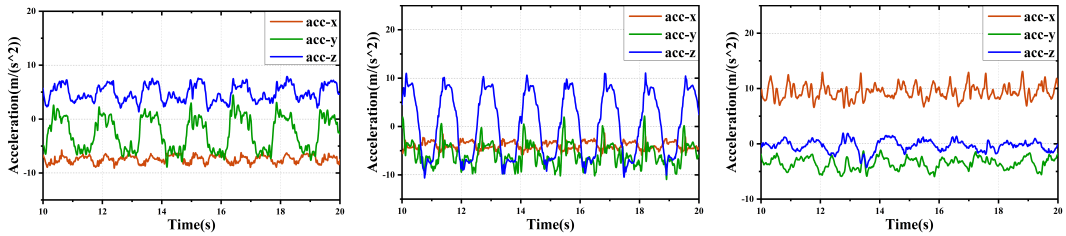
A. Participants

We have collected the signal data from 14 activities of 85 PD patients and 70 healthy controls in the past two years. All individual patient information has been anonymised. Table II shows the demographic information of the subjects. All subjects have identified and signed the informed consent form.

B. Data Collection

The activity data is collected by the wearable sensor shimmer3 IMU units with a sampling frequency of 204hz. These data include a three-axis accelerometer, three-dimensional gyroscope signal and three-axis magnetometer. Due to the lack of integration and high-efficiency consumption of sensors used in many studies, accelerometer signals are the main data source. In order to obtain more accurate motion signal data, we also use gyroscopes to assist the acceleration signal identification activities. Fig. 2 shows the waveform of activity data signals collected by the accelerometer in the sensor.

During the data collection, we placed sensors in five locations (waist, 2 feet, and 2 wrists) on the subjects. For imitating the use of a watch, only the signals of the right wrist were used. Then, under the guidance of professionals, the exercise tasks in Table I are executed in turn (all movements, except standing, walking and picking up things, are completed by the patient sitting on a chair with a height of 40cm). The collection process of UPDRS activities is shown in Fig. 3. It is synchronously transmitted to the computer through a Bluetooth connection, and the activity data is recorded (each action collected for 20s without special instructions), the total duration of the procedure is approximately 15 minutes. The accuracy of wearable devices was ensured by the use of equipment: professional Shimmer3 Inertial Measurement Unit (IMU) units. We chose the high sampling frequency of 204 Hz, which ensures that even the subtlest movements are accurately recorded. This high sampling rate is critical for our study's objective to assess complex PD motor symptoms accurately.



(a) Finger to nose-right (b) Hand rotation-right (c) Walk back and forth

Fig. 2. Active signal waveform diagram. the waveform of activity data signals collected by the accelerometer.

TABLE II
THE MDS-UPDRS SCORE DISTRIBUTION AND DEMOGRAPHIC INFORMATION OF SUBJECTS

UPDRS Score	PD	PD(1)	PD(2)	PD(3)	PD(4)	Healthy controls
N	85	18	34	19	14	70
Sex(F/M)%	39/46	9/9	14/20	9/10	7/7	32/38
Age(mean ± std)	67.2 ± 9.2	68.2 ± 10.4	65.8 ± 7.7	69.5 ± 6.3	67.5 ± 2.2	24.6±5.4
Height(mean ± std) cm	160.9 ± 7.4	163.0 ± 5.0	160.5 ± 5.9	161.0 ± 4.8	159.3 ± 4.3	171.33±3.2
Weight(mean ± std) kg	58.6 ± 8.7	61.0 ± 7.3	57.8 ± 6.9	58.8 ± 5.1	57.5 ± 3.9	63.75±3.4

Fig. 3. Data collection of fourteen activities according to the UPDRS-part III.

TABLE III
FEATURE EXTRACTION ON WIDE RANGE ACCELEROMETER AND GYROSCOPE

Feature	Dimension
Maximum, Minimum, Average, (X,Y,Z,A,T)	30
Variance, Standard Deviation, Amplitude(X,Y,Z,A,T)	
Skewness(X,Y,Z,A,T)	5
kurtosis(X,Y,Z,A,T)	5
Maximum and minimum autocorrelation coefficient(X,Y,Z,A,T)	10
Spectrum maximum, average(X,Y,Z,A,T)	10
Correlation coefficient(XY,XZ,XA,XT,YZ,YA,YT,ZA,ZT,AT)	10
Root mean square(X,Y,Z,A,T)	5
Energy value(X,Y,Z,A,T)	5
Entropy(X,Y,Z,A,T)	5
Dominant frequency (A,T)	2
Total	87

IV. METHODOLOGY

A. Signal Preprocessing

In order to maintain the authenticity of the original signal to a greater extent and reduce the interference of noise, some simple filtering processing is usually performed at the front end of the acquisition system. It is necessary to retain the tremor information and the main information of different movements since this study focuses on PD patients. Through signal spectrum analysis of the signals we collected and review of relevant literature, the tremor frequency of PD patients can be divided into three categories: resting tremor 3-6hz, postural tremor 4-12hz, and motor tremor 2-7hz [34]. Therefore, it is recommended to use a 2-12hz band-pass filter to filter the noise signals from the patient's motion. After filtering, the data of each axis were normalized by Z-score normalization [35]. This filtering technique is specifically designed to improve the reliability of these devices in capturing complex motor symptoms and mitigate the impact of external noise. Noting that we have taken great care to ensure that our filtering technique retains the vital tremor information. This was achieved by closely examining the overlap between the tremor frequency and the filter range.

B. Feature Extraction

After the operation of signal prepossessing, 87-dimensional features are extracted from the accelerometer and gyroscope. In the time domain, the features encompass the average, peak, nadir, dispersion, root mean square, variance, correlation, asymmetry, excess kurtosis, energy, median, and range. In the frequency domain, the features consist of spectral entropy, energy derived from the fast fourier transform, mean amplitude and maximum amplitude. A total of 87 dimensional features. Since we extract 87 dimensions from the accelerometer and gyroscope respectively, the total feature dimension is 174 dimensions. X, Y and Z respectively represent the three axes of the three-dimensional sensor, A is the fusion axis of the three axes, T is the inclination axis, and the fusion representation of the three axes is performed by calculating the signal amplitude vector. For the fusion axis, the fusion representation of the three axes is performed by calculating the signal amplitude vector (SMV), which avoids the user's change in a single direction, which helps to measure the overall intensity of the activity [36]. Detailed characteristics are shown in Table III.

C. Activity Correlation Analysis

Various activities are not only distinct but also how their associations can reflect different aspects and severities of PD.

According to the UK Brain Bank criteria, a diagnosis of Parkinson's syndrome can be made if bradykinesia is present along with at a minimum the following: rigidity, rest tremor, and postural instability [37]. Part III of the MDS-UPDRS includes specific tests for tremors in PD patients. Test 3.17 assesses rest tremor, while test 3.15 evaluates postural tremor [38] [39]. In the finger tapping tests of the MDS-UPDRS, participants are requested to perform rapid and forceful tapping of their index finger to their thumb for a total of 10 repetitions. This assessment aims to evaluate the speed and amplitude of finger movements, serving as an indirect measure of the functional status of cortical motor areas. The finger tapping test is considered a valuable indicator of PD as it assesses bradykinesia through the observation of decreased rate or amplitude during repetitive finger movements. This test provides valuable insights into the motor impairment associated with PD and helps in diagnosing the condition. Other hand fine activities, such as wrist turnover, right-hand fingertip and nose tip, were evaluated for speed, amplitude, hesitation, pause and attenuation amplitude, and motion completion to assess the severity of muscle rigidity and quantify PD motor improvements [32]. So we think it would be better to evaluate the disease after integrating multiple symptoms, Fig.4 shows the correlation between activities. We use cosine similarity to calculate the cosine value of the angle between different activities' corresponding features. It measures the similarity in direction between two feature vectors, regardless of their absolute magnitudes. The specific steps are as follows, firstly, compute the dot product of feature vectors A and B, multiply the corresponding elements of vectors A and B, and sum up the products. This value represents the similarity between the two vectors in each dimension. Secondly, Computing the norm of vector A/B, summing up the squares of each element in vector A/B, and taking the square root of the result. This value represents the length or magnitude of vector A/B. Thirdly, by dividing the dot product obtained in step 1 by the product of the norms computed in step 2, this result represents the cosine similarity. From the figure, it can be seen that activities with significant differences in symptoms have a lower correlation. In the experimental section, we conducted a combined analysis of high/low correlation activities.

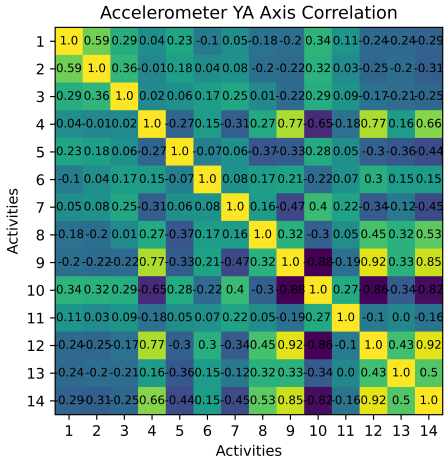


Fig. 4. Activity correlation analysis. The cosine similarity value ranges from -1 to 1, when the cosine similarity is close to 1, it indicates that the two vectors have a very similar direction; when the cosine similarity is close to -1, it indicates that the two vectors have very dissimilar directions; when the cosine similarity is close to 0, it indicates that there is no significant correlation between the directions of the two vectors.

D. Activity Combination Method

Different from traditional methods, activity combination steps are added after feature extraction. Because the information reflected by a single activity is limited, the UPDRS scale gives a comprehensive score based on all the activities of the patient, so our method tries to select representative activities to combine, concatenate multiple activity features, and add feature dimensions. In this way, it is more reasonable to evaluate the disease level of patients by integrating multiple activity information. Considering the correlation between activities and making full use of the data information of multiple activities, we try different feature concatenate methods on activities. This experiment adopts four concatenating strategies, cause we have multiple activities, we donate A_i as the activities, and F_{A_i} is the feature extracted from the activities, the four concatenating strategies are: (1) Horizontal concatenating: increase feature dimension;(2) Vertical concatenating: increase the number of samples;(3)H+V: Horizontal and vertical concatenating;(4) Weighted: active feature fusion.

$$\text{Concatenate\_}H\left(F_{A_1}, F_{A_2}, \cdots, F_{A_N}\right) \tag{1}$$

$$\text{Concatenate\_}V\left(F_{A_1}, F_{A_2}, \cdots, F_{A_N}\right) \tag{2}$$

$$\text{Concatenate\_}H+V\left(F_{A_1}, F_{A_2}, \cdots, F_{A_N}\right) \tag{3}$$

$$\text{Weight } = \sum_{\mathrm{i}=1}^{\mathrm{N}} \mathrm{F}_{\mathrm{A}_1} \cdot \mathrm{ACC}_{\mathrm{i}} \tag{4}$$

The extensive adoption of machine learning (ML) brings opportunities for a variety of real-life applications [40]–[45]. In order to assess the effectiveness of combination strategies, we utilised four classical ML algorithms: Support Vector Machines (SVM), K-Nearest Neighbors (KNN), Extreme Gradient Boosting (XGB), and Light Gradient Boosting Machine (LGBM). We chose activities 2(COA), 3(ALTER),4(HR-R),5(HR-L) for different combination strategies. The specific classification accuracy after concatenating is shown in Table IV. It can be seen that the accuracy of horizontal concatenating on the XGB model is the highest (74.07%), which is greater than the effect of vertical and weighted concatenating. The experimental findings indicated that the classification accuracy can be effectively improved by transforming features through horizontal concatenating. The reason why horizontal splicing is more effective than other splicing methods is that it expands the feature dimension, which can get uncorrelated information from the different activities which will lead to better accuracy.

TABLE IV
ACCURACY OF DIFFERENT CONCATENATING METHODS

	LGBM	SVM	KNN	XGB
V_245	62.96	44.44	46.3	67.59
H+V_245	67.59	53.7	61.11	73.15
W_245	62.04	48.15	51.85	65.74
H_2345	70.37	58.33	60.19	72.22
H_245	**69.44**	**62.96**	**63.89**	**74.07**

E. Significant Feature Selection

Due to the high dimensional curse brought about by directly conducting a combination of multiple activities, we are trying to reduce the dimensions and identify important features that are more relevant to the disease. After this dimensionality reduction shown in Fig.5, the dimension is small and the model performance is better.

Fig. 5. Feature dimensionality reduction

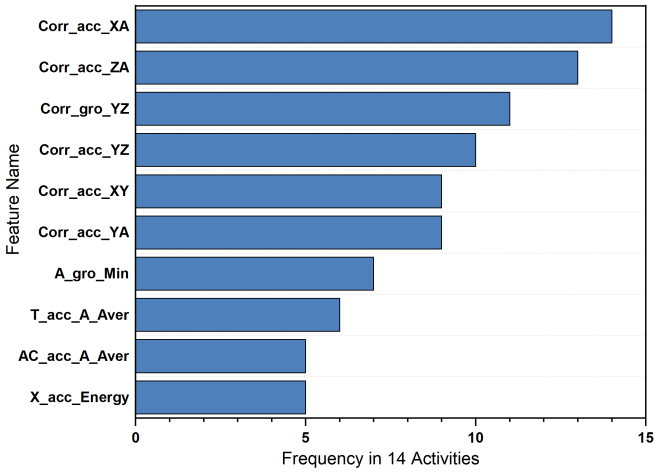


Fig. 6. Feature importance ranking. This figure shows the top 10 most important features jointly selected by the three models. Corr represents the axis correlation. The sensor contains three axes: X, Y, and Z to capture motion information. The features at the front of the figure are concentrated on the axis correlation, indicating that this feature plays a more important role in classification

After four model experiments and their common results, the feature importance ranking is shown in Fig. 6. Among the 10 features with the highest frequency in the four models, axis correlation is a feature worthy of attention. Axis correlation represents the correlation of the degree of change between the two axes, reflecting the degree of action completion, such as wrist turnover. When the x and z axes are negatively correlated when turning, when the patient is stuck in the middle of the illness when performing activities, the correlation will also change. Therefore, the correlation of different axes may be related to some delays, obstacles and coagulation during the movement.

V. EXPERIMENTS AND RESULTS

In this section, we introduced the validation methods and evaluation indicators, and then used a single activity and a selected combination of activities to test and verify the effectiveness of the activity combination, and analyzed its interpretability. Finally, we selected the most representative top 20 important features for clinical reference. It is worth mentioning that, our research content is fine-grained severity stage classification which involves the classification of mild(1+2), moderate(3), and severe symptoms(4).

Validation. We validated our approach on datasets collected in a laboratory environment. All our experiments were carried out on an ordinary computer with a 2.6GHz CPU and 8GB memory. In the selection of verification methods, we choose the "Leave-One-Subject-Out Cross Validation".It is to make k equal to the number of data in the dataset, using only one test set at a time and the rest as the training set. The final result is the mean of these k verifications, and the result obtained by this method is closest to the expected value of the entire training test set. This is because a Semi-Non-Overlapping window is used when generating samples, so the same sample fragments may exist in the training set and the test set simultaneously, resulting in bias in the experiment. It is worth noting that the semi-non-overlapping window technique used in the leave-one-subject-out validation protocol is not influenced by bias. This is particularly desirable as the training and testing samples are separated by subjects. Consequently, the raw signal used to generate the samples, even if temporally close, will only appear in either the training or the testing set, but not in both. This approach ensures a more robust and unbiased evaluation of the model's performance.

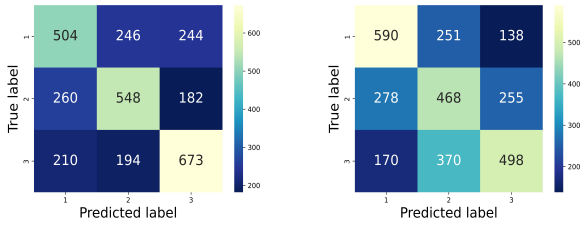
Evaluation metrics and models. Accuracy, F1-score, precision and recall are utilised in our experiments. We build the model on four ML algorithms: LGBM, SVM, KNN and XGB.For LGBM, the learning rate is 0.05, the maximum depth of trees is 2, the maximum number of trees is 300. For SVM, the kernel function is sigmoid, the penalty coefficient is 1. For KNN, the number of neighbors is 5, the weights distribution is uniform and the distance is 2. For XGB, the learning rate is [0.01,0.05], the number of iterations is [700,1500], and the minimum loss decrease is 0. One of the advantages of utilizing the gradient lifting algorithm is the direct obtainment of importance scores for each attribute once the lifting tree is constructed. This attribute-specific importance scoring feature enhances the interpretability of our results.

A. Construction of Balanced Dataset

During the data collection process of PD patients, we encountered the issue of data imbalance. Specifically, there was an uneven distribution among disease stages, with a higher number of patients classified as having mild disease (stages 1, 2) and a relatively small number of patients classified as having severe disease (stage 4). Therefore, the sample imbalance had a great impact on the experimental results, we chose to use the "Leave-One-Subject-Out" method to construct a balanced data set, which can decrease the negative influence of the data

TABLE V
CLASSIFICATION RESULTS OF HEALTHY INDIVIDUALS AND PD PATIENTS

Act_Num	Act_Name	Acc	F1	Pre	Rec
1	FT	87.50	87.40	87.85	87.45
2	COA	84.10	84.10	84.15	84.15
3	ALTER	83.30	83.25	83.55	83.30
4	HR-R	**92.50**	**92.45**	**92.90**	**92.45**
5	HR-L	**92.50**	**92.45**	**92.60**	**92.45**
6	FN-L	85.00	84.90	85.15	84.95
7	FN-R	**95.80**	**95.80**	**96.05**	**95.82**
8	STANDH	**95.80**	**95.75**	**96.25**	**95.80**
9	WA	86.60	86.50	87.90	86.65
10	AC	73.30	73.15	74.15	73.30
11	DRINK	85.80	85.75	86.30	85.80
12	PICK	91.10	91.05	91.40	91.05
13	SIT	**95.00**	**94.95**	**95.05**	**94.95**
14	STAND	78.30	78.20	78.60	76.65



(a) Clench and Open Alternately (b) Finger to Nose-Right

Fig. 7. The confusion matrix of the Clench and Open Alternately(COA) and Finger to Nose-Right(FN-R). The COA has a better classification effect in Critical patients (stage4), and the FN-R has a better classification effect in patients with mild symptoms (stages 1 and 2).

imbalance. More details are analysed in supplement material S1.

B. Detection and Classification of motor severity based on a single activity

On the basis of the verification method based on constructing balanced data sets, We conducted two major types of experiments, one is the detection experiment between healthy individuals and PD, which aims to accurately identify high-risk individuals who are considered healthy and tend to develop into early PD patients. Table V below shows the classification results of healthy individuals and PD patients. According to the experimental results, for early PD detection, the vast majority of UPDRS activities can be well detected, especially activities 7(FN-R) and 8(STANDH). Bold indicates an accuracy rate greater than 92%.

Another type of fine-grained disease severity classification for PD patients. We use single activity data for the experiment. Table VI shows the classification accuracy of 14 single activities. We found that most patients had obvious symptoms in the fine movements of the hand (clench and open alternately, hand rotation right), while the symptoms were suppressed in the whole-body movements (drinking water, rising from a chair) or static activities (standing with arms hold, sitting, standing). Such activities with no obvious symptoms would interfere with the experiment, So we aim to find the most representative ones among these activities. Our experimental findings align with this observation as well. The accuracy of hand movements is always higher than that of large amplitude or static movements. We choose this kind of hand-fine activity to carry out the following activity combination.

Fig.7 shows the confusion matrix of the Clench and Open Alternately(COA) and Finger to Nose-Right(FN-R). The COA has a better classification effect in critical patients (stage 4), and the FN-R has a better classification effect in patients with mild symptoms (stages 1 and 2). The reason why severe can not be correctly divided may be that there are too few patients in stage 4, severe patients cannot complete some activities, and there is no activity data. In the middle stage(stage 3), we also find the number of samples that were wrongly divided in many activities, The possible reason is that the patients with moderate disease are in adjacent stages and have similar conditions with mild and severe diseases, which is difficult to distinguish. The result revealed that Rapid alternating movements of hands outperform other activities with an accuracy rate of more than 73%.

C. Classification of motor severity based on Multi-Activity

From the results of the single activity experiment, we can see that several activities with high accuracy are concentrated in the hand fine activity, which is in line with the previous experimental conclusion of the correlation analysis of activity and disease, so the subsequent experiments mainly focus on the combination of hand fine activities, we selected seven activities: Finger taps(FT), Clench and open alternately(COA), Rapid alternating movements of hands (ALTER), Hand rotation-right(HR-R), Hand rotation-left(HR-L), Finger to nose-left(FN-L), Finger to nose-right(FN-R). After the correlation analysis between activities, we combined the high and low-correlation activities. The experimental findings are presented in Table VII. The results indicate that the combination of activities with low correlation exhibits a higher rising trend compared to activities with high correlation. This suggests that combining two activities with low correlation can effectively supplement the characteristic information to a certain extent. It is worth mentioning that when there are too many combined activities, the accuracy rate decreases, as shown in the last row of the Table VII. Therefore, selecting the appropriate activities for combination is crucial, while combining too many activities provides redundant and complex information that can interfere with the model. Therefore, we are committed to implementing as few activities as possible at home to timely assess the condition of patients in an out-of-hospital environment.

Table VII shows the experimental accuracy of high /low correlation activity combinations. Taking activity 2(COA) as an example, activities 1(FT) and 3(ALTER) have a high correlation with it, while activities 4(HR-R), 5(HR-L), 6(FN-L), and 7(FN-R) have a low correlation. They form different combinations. At present, the combination with the best classification effect is 2(COA)+4(HR-R)+5(HR-L), with an accuracy rate of 74.07% in the XGB model. Compared with single activity classification, the accuracy rate has increased by 12.03%. The results clearly demonstrate that the horizontal concatenation method significantly enhances the correctness

TABLE VI
CLASSIFICATION ACCURACY OF SINGLE ACTIVITY

Num	Activity Name	LGBM				SVM				KNN				XGB			
		Acc	F1	Pre	Rec	Acc	F1	Pre	Rec	Acc	F1	Pre	Rec	Acc	F1	Pre	Rec
1	**FT**	42.59	41.89	41.73	42.59	27.78	27.72	27.84	27.78	35.19	33.90	34.81	35.19	44.44	44.28	44.99	44.45
2	**COA**	**60.19**	**59.55**	**59.90**	**60.19**	52.78	52.96	56.50	52.78	53.7	52.32	55.44	53.70	55.56	54.86	54.77	55.56
3	**ALTER**	**73.15**	**72.99**	**73.42**	**73.15**	49.07	49.08	49.46	49.07	54.63	54.50	56.37	54.63	67.59	66.88	66.70	67.59
4	**HR-R**	**62.04**	**62.03**	**62.60**	**62.03**	55.56	55.87	56.46	55.56	44.44	43.31	44.73	44.45	54.63	53.67	54.29	54.63
5	**HR-L**	**59.26**	**58.98**	**59.99**	**59.26**	44.44	43.87	43.66	44.45	46.3	44.55	44.31	46.30	62.04	61.23	61.41	62.04
6	**FN-L**	49.07	47.20	46.18	49.07	44.44	42.94	43.84	44.45	44.44	40.68	44.55	44.55	49.07	47.26	47.04	49.07
7	**FN-R**	52.78	52.74	53.94	52.78	35.19	34.10	34.76	35.18	38.89	36.65	37.94	38.89	57.41	57.83	58.93	57.41
8	**STANDH**	26.85	26.57	26.48	26.85	25.93	24.46	23.63	25.93	30.56	29.93	29.72	30.55	32.41	32.16	33.00	32.41
9	**WA**	47.22	45.68	44.71	47.22	52.78	51.96	52.64	52.78	53.7	52.39	52.60	53.70	48.15	47.25	47.52	48.15
10	**AC**	**61.11**	**65.33**	**64.66**	**66.67**	39.81	31.52	30.80	39.82	28.7	24.96	23.25	28.71	57.41	55.47	55.71	57.41
11	**DRINK**	47.5	59.4	55	65	25	27.10	30.82	25.00	40	39.67	39.96	40.00	50.83	49.03	48.14	50.83
12	**PICK**	54.17	52.76	61.79	46.88	42.71	41.00	47.23	42.71	50	50.11	51.82	50.00	58.33	58.71	60.84	58.33
13	**SIT**	36	33.87	46.33	28	25.33	23.40	26.00	25.33	21.33	19.85	20.73	21.33	21.33	19.17	17.55	21.33
14	**STAND**	19.44	13.69	15.28	12.5	19.44	18.20	18.24	19.44	22.22	19.98	18.52	22.22	19.44	19.82	21.86	19.45

TABLE VII
MULTIPLE ACTIVITY COMBINATION ACCURACY

Algorithm		LGBM				SVM				KNN				XGB			
Motor correlation		**Acc**	**F1**	**Pre**	**Rec**	**Acc**	**F1**	**Pre**	**Rec**	**Acc**	**F1**	**Pre**	**Rec**	**Acc**	**F1**	**Pre**	**Rec**
High	**1+2**	55.56	55.45	55.80	55.56	53.7	53.34	54.36	53.71	51.85	50.97	52.34	51.85	50.93	53.34	50.87	50.93
	2+3	62.04	62.02	62.77	62.04	53.7	54.22	56.36	53.70	57.41	57.02	59.40	57.41	62.96	62.05	62.01	62.96
Low	**2+6**	57.41	57.23	57.29	57.41	45.37	45.34	46.61	45.37	61.11	60.27	67.60	61.11	62.04	61.73	60.22	62.04
	2+7	63.89	63.77	64.17	63.89	51.85	52.03	55.29	51.85	46.3	44.22	46.43	46.30	51.85	52.45	53.12	51.85
	2+4	56.48	56.21	56.16	56.48	39.81	40.35	41.27	39.81	53.7	51.26	59.32	53.70	55.56	56.78	55.25	55.56
	2+5	63.89	63.84	65.15	63.89	63.89	63.06	65.00	63.89	61.11	59.33	66.23	61.11	71.3	72.42	71.7	71.3
	2+4+5	69.44	69.43	71.51	69.45	62.96	63.40	65.00	62.96	63.89	62.96	70.16	63.89	**74.07**	**75.05**	**74.23**	**74.07**
	2+3+4+5	70.37	70.29	70.79	70.37	58.33	59.21	61.35	58.33	60.19	60.60	64.59	60.19	**72.22**	**73.56**	**72.98**	**72.22**

TABLE VIII
COMPARISON OF ACCURACY BEFORE AND AFTER DIMENSION REDUCTION

2+4+5								
Feature Reduction	**Before**				**After**			
Algorithm	**Acc**	**F1**	**Pre**	**Rec**	**Acc**	**F1**	**Pre**	**Rec**
SVM	62.96	63.4	65	62.96	53.38	65.3	65.35	54.38
KNN	63.89	62.96	70.16	63.89	63.54	62	71.24	54.56
XGB	74.07	75.05	74.23	74.07	72.03	77.4	76.12	72.03
LGBM	69.44	69.43	71.51	69.45	**75.93**	**75.58**	**76.56**	**75.92**
2+3+4+5								
Feature Reduction	**Before**				**After**			
Algorithm	**Acc**	**F1**	**Pre**	**Rec**	**Acc**	**F1**	**Pre**	**Rec**
SVM	58.33	59.21	61.35	58.33	68.52	68.52	70.87	68.52
KNN	60.19	60.6	64.59	60.19	65.74	65.17	67.74	65.74
XGB	72.22	73.56	72.24	72.22	77.78	77.27	78.22	77.78
LGBM	70.37	70.29	70.79	70.37	**82.41**	**82.54**	**81.62**	**81.48**

of staging evaluation in PD patients. The second best combination is 2(COA) + 3(ALTER)+4(HR-R)+5(HR-L), with an accuracy rate of 72.78%.We can observe that it is not that the more activities you select, the better. When activity 3(ALTER) was included, the accuracy decreased. Possible reasons for this could be the introduction of redundant features and the curse of dimensionality caused by the high dimensionality resulting from activity combinations. To validate the effectiveness of activity combinations, we performed feature dimensionality reduction, which is explained in detail in section E of the experiment.

D. Feature Interpretability Analysis

At the same time, We analyzed the reasons why the combination of activities would increase. we found that after the combination of activities, the first 20 significant features come from different activities, as shown in Fig.8, which shows that the horizontal combination of activities provides more feature dimension information, which plays a certain complementary role among multiple activities. The red box, green box and yellow box are the characteristics of 2(COA),4(HR-R), and 5(HR-L).

On the other hand, we compare the importance weight of the most significant feature in the combined activities in different single activities as shown in Fig.9. The following two figures show the importance of the maximum value of the x-axis of the gyroscope in different activities (HMOV, HR-L, HR-R). It can be seen that the same feature has different importance in different activities, and the same feature extracted from different activities reflects different information. Hence, when the information dimension is expanded, it can provide more complementary disease information.

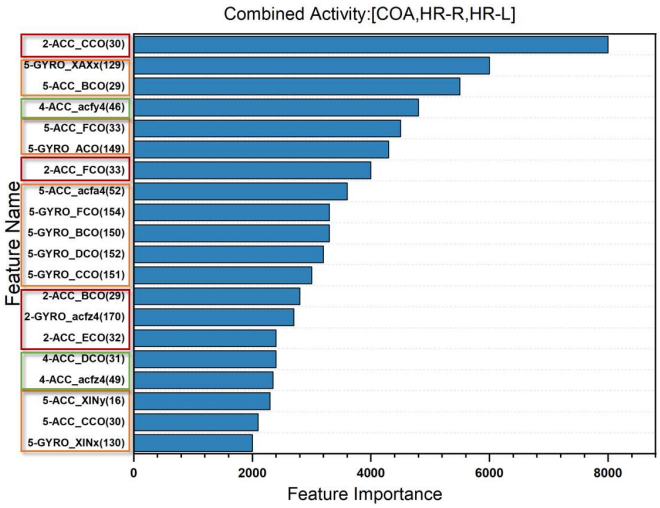
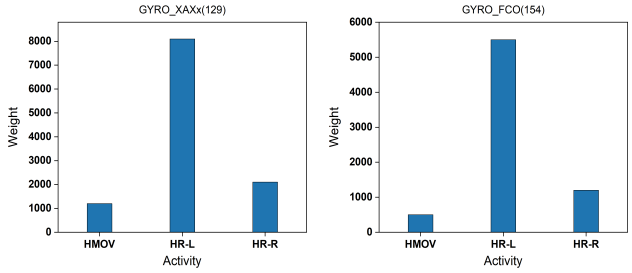


Fig. 8. Top20 Features of combined activities. The 20-dimensional important features after combining activities come from the significant features of different activities, indicating that horizontal activity combinations provide more feature dimensional information, which plays a certain complementary role among multiple activities.



(a) Gyro-XAxx in different activities (b) Gyro-FCO in different activities

Fig. 9. Comparison of feature importance weights in different single activities. The importance of the maximum value of the x-axis of the gyroscope in different activities (HMOV, HR-L, HR-R), the figures show the same feature has different importance in different activities.

E. Multi-Activity Important Features based Diagnosis

Due to the high latitude curse brought by the combination of activities, After selecting the most significant 20-dimensional features for the test, we observed a notable improvement in the accuracy rate when we eliminated other redundant features. This underscores the effectiveness of feature selection in enhancing the performance of the test. Table VIII evaluates the classification of the best activity combination(2(COA) + 3(ALTER)+4(HR-R)+5(HR-L)) before and after dimension reduction with four ML algorithms. We can see that after the dimensionality reduction, the performance increases to 82.41% in LGBM, which improves 12.04% compared to using all features and improves 10.19% compared to single activity in the XGB model.LGBM is superior to SVM and KNN in terms of performance, both before and after dimension reduction.This may be due to the dimensionality reduction operation eliminating a large amount of redundant information.The LGBM model is insensitive to noise and has stronger robustness. Outliers and missing values contain valuable information that can be learned, whereas SVM and KNN are more sensitive to noise, leading to a decrease in prediction performance. In contrast, LGBM demonstrates better interpretability and can assist in optimizing feature engineering to a certain extent. In summary, activities 2(COA) + 3(ALTER)+4(HR-R)+5(HR-L) are the representative activities that we need.

TABLE IX
COMPARISON OF PD ASSESSMENT ON SPECIFIC ACTION TYPES

References	Action Type	Assess symptoms
[15][27]	tremor type(rest / postural)	tremor
[21]	motor exercises / finger tapping	bradykinesia
[28]	wrist turnover / finger to nose	muscle rigidity
[19][20]	walking	freezing of gait

Furthermore, we estimated the performance of each severity grade before and after the feature reduction, using accuracy, precision, recall and F1-score. The accuracy before and after feature reduction was 70.37 and 82.41 respectively. As shown in Fig.10, we show the performance of accuracy, recall and F1-score on three classifications. In order to improve the confidence of the results, a more rigorous testing criterion is adopted. The estimated performance of the three grades was examined separately instead of averaging the statistics across all grades in Table.VIII. We observed that:

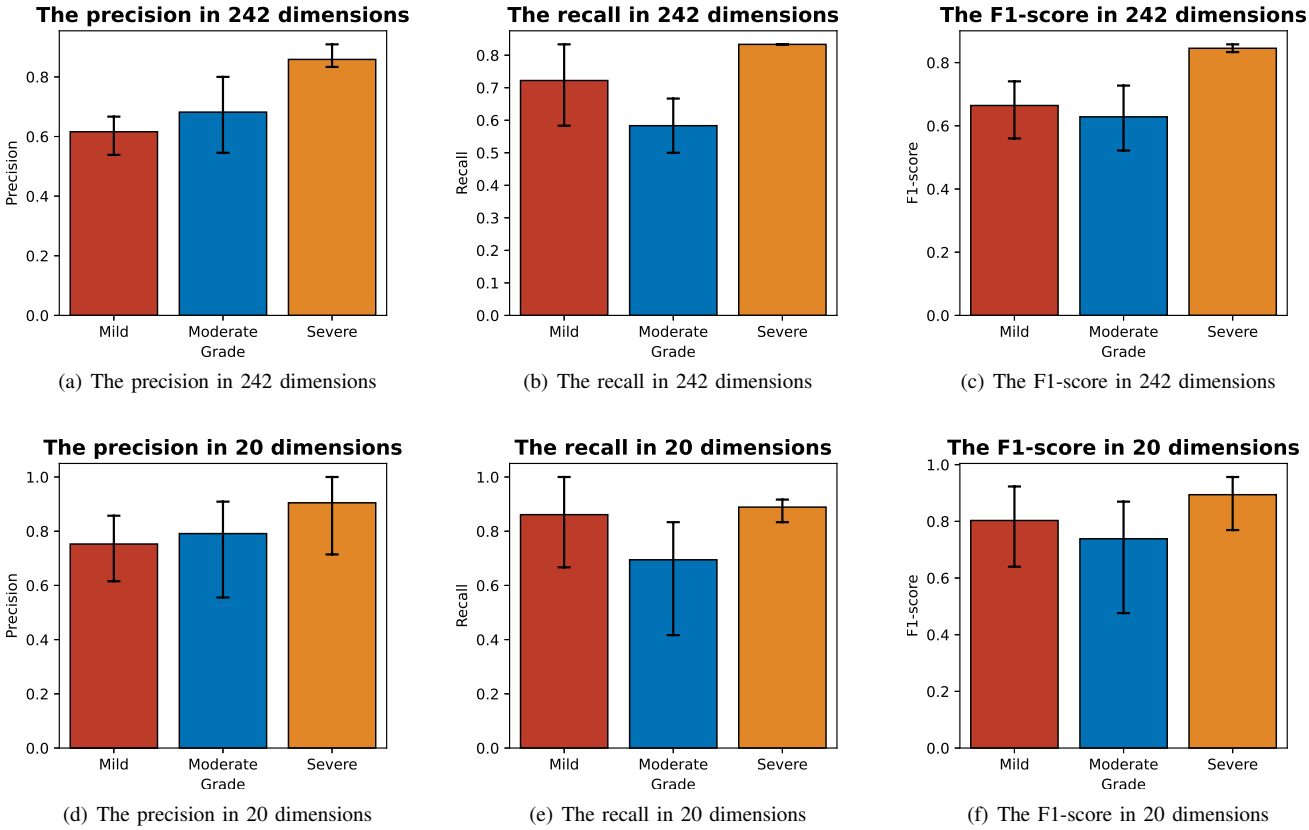
- The feature reduction brings about 12% improvement in performance while the confidence intervals (95% CI) are wider.
- Individuals in the severe grade have the highest classification performance, but feature reduction is not a major contributor.
- The performance of the moderate grade is the most challenging, despite performance gains from feature reductions.
- Feature reduction provides the greatest improvement in the classification of mild individuals, which is a key factor in the overall ability of the model estimation.

The experimental conclusion has significant clinical implications. It suggests that in clinical practice, doctors only need to focus on hand-related activities for diagnosis and assessment of the patient's condition. This saves a considerable amount of time and effort while making the evaluation process more objective. This finding can greatly assist in facilitating diagnosis and treatment procedures.

VI. DISCUSSION

Achieving self-assessment of PD using all types of UPDRS activity is an open question. Our proposed method has been able to identify the most representative activities under all standard UPDRS, which are verified to work well with common machine-learning algorithms. Although existing studies [19], [21], [23]–[25], [32] on PD assessment are limited to specific action types, as shown in Tabel IX, we will do an in-depth comparison to develop a representative self-assessment algorithm.

For the classification result of healthy individuals and PD patients in Table V, we used data from younger individuals as a comparison. Normal older adults have a higher risk of developing PD. More normal elderly subjects will be recruited to further develop an efficient early screening study.



(a) The precision in 242 dimensions (b) The recall in 242 dimensions (c) The F1-score in 242 dimensions

(d) The precision in 20 dimensions (e) The recall in 20 dimensions (f) The F1-score in 20 dimensions

Fig. 10. Bar charts for the three classifications. (a)-(c) show the precision, recall and F1-score for three classification with 242 dimensions.(d)-(f) show the precision, recall and F1-score for three classification with 20 dimensions.

In this paper, all subjects' activities were recorded using the Shimmer3 sensor. Moving towards self-assessment of PD in a free-living environment, we will utilise a consumer-level sensor device (e.g. a watch) to collect data as a means of validating the effectiveness of selected representative activities.

The heterogeneity of PD symptoms and the potential impact of external factors in uncontrolled environments may influence the results. In the process of feature extraction, tremors are characterized by features like Maximum, Minimum, and Average values, and bradykinesia which is analysed through frequency-domain features like the Energy value. In the future, the goal is to ensure that the selected activities and their corresponding features reflect a comprehensive view of the UPDRS scale. Additionally, future work will involve a more detailed investigation into these external factors(such as abnormal data detection [46]) to enhance the robustness of our data collection and analysis methods.

The integration of such technology into existing clinical workflows could significantly enhance patient care by providing continuous, real-time data that could lead to more timely and personalized interventions. For instance, it would involve evaluating current clinical workflows to identify areas where remote monitoring could augment or streamline processes, such as routine check-ups or symptom tracking [47] It also involves ensuring that technology interfaces seamlessly with electronic health records and existing IT infrastructure to facilitate efficient data transfer and analysis [48].

Ongoing and instantaneous surveillance of health status has the capacity to detect nuanced variations in disease progression that might not be evident during episodic clinical evaluations. Monitoring provides a constant stream of health data, ensuring that even the slightest changes in the patient's condition are documented over time. The discovery of representative activities will help to advance the development of continuous and timely home monitoring techniques.

VII. Conclusion

Accurately capturing motor symptom diagnosis of PD patients is particularly important to determine appropriate medication schedules. Although all MDS-UPDRS activities are designed for detecting related PD symptoms, the information about symptoms tends to be complementary and redundant with each other. In this paper, we propose a novel technical pipeline for fine-grained classification of PD severity grades and identifying the most representative activities combination, which effectively provides more representative information than enforcing all MDS-UPDRS activities. The experimental results show that the fine-grained classification accuracy of PD disease grade is 82.41% when combined 4 activities, and the accuracy improves by 11.02% over only using a single activity. Furthermore, we select the top 10 most significant features for dimension reduction, which not only achieve good results but also facilitate the later transplantation of lightweight equipment and provide reference for the PD self-assessment at

home environment. In the future, we will further explore the relationship between these standard actions and unrestricted automatic actions to achieve individual free-living monitoring. We hope that our experiments and theoretical method will make effects of proposed solutions for PD diagnosis in an uncontrolled environment setup, finally leading to self-PD both in and out of the hospital.

ACKNOWLEDGMENTS

This research was supported by the National Natural Science Foundation of China (No. 62301452, No. 62061050) and the 13th Postgraduate Research Innovation Project of Yunnan University 2021Y397. We thank Yunnan First People's Hospital for its strong support.

[1] E. Dorsey, T. Sherer, M. S. Okun, and B. R. Bloem, "The emerging evidence of the parkinson pandemic," *Journal of Parkinson's disease*, vol. 8, no. s1, pp. S3–S8, 2018.

[2] E. Leroy, D. Anastasopoulos, S. Konitsiotis, C. Lavedan, and M. H. Polymeropoulos, "Deletions in the parkin gene and genetic heterogeneity in a greek family with early onset parkinson's disease," *Human genetics*, vol. 103, pp. 424–427, 1998.

[3] X.-a. Bi, X. Hu, Y. Xie, and H. Wu, "A novel cernne approach for predicting parkinson's disease-associated genes and brain regions based on multimodal imaging genetics data," *Medical Image Analysis*, vol. 67, p. 101830, 2021.

[4] T. Asakawa, K. Sugiyama, T. Nozaki, T. Sameshima, S. Kobayashi, W. Liang, H. Zhen, C. Shujiao, L. Candong, and H. Namba, "Can the latest computerized technologies revolutionize conventional assessment tools and therapies for a neurological disease? the example of parkinson's disease," *Neurologia medico-chirurgica*, vol. 59, no. 3, p. 69, 2019.

[5] N. C. C. for Chronic Conditions *et al.*, "Symptomatic pharmacological therapy in parkinson's disease," *Parkinson's Disease. London: Royal College of Physicians*, pp. 59–100, 2006.

[6] S. Rahman, H. J. Griffin, N. P. Quinn, and M. Jahanshahi, "Quality of life in parkinson's disease: the relative importance of the symptoms," *Movement disorders: official journal of the Movement Disorder Society*, vol. 23, no. 10, pp. 1428–1434, 2008.

[7] K. P. Bhatia, P. Bain, N. Bajaj, R. J. Elble, M. Hallett, E. D. Louis, J. Raethjen, M. Stamelou, C. M. Testa, G. Deuschl, *et al.*, "Consensus statement on the classification of tremors. from the task force on tremor of the international parkinson and movement disorder society," *Movement Disorders*, vol. 33, no. 1, pp. 75–87, 2018.

[8] O. Y. Chén, F. Lipsmeier, H. Phan, J. Prince, K. I. Taylor, C. Gossens, M. Lindemann, and M. De Vos, "Building a machine-learning framework to remotely assess parkinson's disease using smartphones," *IEEE Transactions on Biomedical Engineering*, vol. 67, no. 12, pp. 3491–3500, 2020.

[9] M. K. Erb, D. R. Karlin, B. K. Ho, K. C. Thomas, F. Parisi, G. P. Vergara-Diaz, J.-F. Daneault, P. W. Wacnik, H. Zhang, T. Kangarloo, *et al.*, "mhealth and wearable technology should replace motor diaries to track motor fluctuations in parkinson's disease," *NPJ digital medicine*, vol. 3, no. 1, p. 6, 2020.

[10] M. Ullrich, A. Mücke, A. Küderle, N. Roth, T. Gladow, H. Gaßner, F. Marxreiter, J. Klucken, B. M. Eskofier, and F. Kluge, "Detection of unsupervised standardized gait tests from real-world inertial sensor data in parkinson's disease," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 29, pp. 2103–2111, 2021.

[11] N. Hammerla, J. Fisher, P. Andras, L. Rochester, R. Walker, and T. Plötz, "Pd disease state assessment in naturalistic environments using deep learning," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 29, 2015.

[12] P. Yang, M. Hanneghan, J. Qi, Z. Deng, F. Dong, and D. Fan, "Improving the validity of lifelogging physical activity measures in an internet of things environment," in *2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing*, pp. 2309–2314, IEEE, 2015.

[13] P. Yang, "Prls-inves: A general experimental investigation strategy for high accuracy and precision in passive rfid location systems," *IEEE Internet of Things Journal*, vol. 2, no. 2, pp. 159–167, 2014.

[14] Z. Deng, P. Yang, Y. Zhao, X. Zhao, and F. Dong, "Life-logging data aggregation solution for interdisciplinary healthcare research and collaboration," in *2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing*, pp. 2315–2320, IEEE, 2015.

[15] P. Yang, D. Stankevicius, V. Marozas, Z. Deng, E. Liu, A. Lukosevicius, F. Dong, L. Xu, and G. Min, "Lifelogging data validation model for internet of things enabled personalized healthcare," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 48, no. 1, pp. 50–64, 2016.

[16] S. Patel, K. Lorincz, R. Hughes, N. Huggins, J. Growdon, D. Standaert, M. Akay, J. Dy, M. Welsh, and P. Bonato, "Monitoring motor fluctuations in patients with parkinson's disease using wearable sensors," *IEEE transactions on information technology in biomedicine*, vol. 13, no. 6, pp. 864–873, 2009.

[17] B. M. Eskofier, S. I. Lee, J.-F. Daneault, F. N. Golabchi, G. Ferreira-Carvalho, G. Vergara-Diaz, S. Sapienza, G. Costante, J. Klucken, T. Kautz, *et al.*, "Recent machine learning advancements in sensor-based mobility analysis: Deep learning for parkinson's disease assessment," in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 655–658, IEEE, 2016.

[18] R. Powers, M. Etezadi-Amoli, E. M. Arnold, S. Kianian, I. Mance, M. Gibiansky, D. Trietsch, A. S. Alvarado, J. D. Kretlow, T. M. Herrington, *et al.*, "Smartwatch inertial sensors continuously monitor real-world motor fluctuations in parkinson's disease," *Science translational medicine*, vol. 13, no. 579, p. eabd7865, 2021.

[19] G. Rigas, A. T. Tzallas, M. G. Tsipouras, P. Bougia, E. E. Tripoliti, D. Baga, D. I. Fotiadis, S. G. Tsouli, and S. Konitsiotis, "Assessment of tremor activity in the parkinson's disease using a set of wearable sensors," *IEEE Transactions on Information Technology in Biomedicine*, vol. 16, no. 3, pp. 478–487, 2012.

[20] J.-W. Kim, J.-H. Lee, Y. Kwon, C.-S. Kim, G.-M. Eom, S.-B. Koh, D.-Y. Kwon, and K.-W. Park, "Quantification of bradykinesia during clinical finger taps using a gyrosensor in patients with parkinson's disease," *Medical & biological engineering & computing*, vol. 49, pp. 365–371, 2011.

[21] M. Giuberti, G. Ferrari, L. Contin, V. Cimolin, C. Azzaro, G. Albani, and A. Mauro, "Automatic updrs evaluation in the sit-to-stand task of parkinsonians: Kinematic analysis and comparative outlook on the leg agility task," *IEEE journal of biomedical and health informatics*, vol. 19, no. 3, pp. 803–814, 2015.

[22] M. Ricci, G. Di Lazzaro, V. Errico, A. Pisani, F. Giannini, and G. Saggio, "The impact of wearable electronics in assessing the effectiveness of levodopa treatment in parkinson's disease," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 7, pp. 2920–2928, 2022.

[23] Y. Guo, L. Wang, Y. Li, L. Guo, and F. Meng, "The detection of freezing of gait in parkinson's disease using asymmetric basis function tv-arma time–frequency spectral estimation method," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 27, no. 10, pp. 2077–2086, 2019.

[24] J. C. Perez-Ibarra, A. A. Siqueira, and H. I. Krebs, "Identification of gait events in healthy subjects and with parkinson's disease using inertial sensors: an adaptive unsupervised learning approach," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 28, no. 12, pp. 2933–2943, 2020.

[25] L. Sigcha, B. Domínguez, L. Borzì, N. Costa, S. Costa, P. Arezes, J. M. López, G. De Arcas, and I. Pavón, "Bradykinesia detection in parkinson's disease using smartwatches' inertial sensors and deep learning methods," *Electronics*, vol. 11, no. 23, p. 3879, 2022.

[26] J. Zhang, H. A. Mattison, C. Liu, C. Ginghina, P. Auinger, M. P. McDermott, T. Stewart, U. J. Kang, P. S. G. D. Investigators, K. C. Cain, *et al.*, "Longitudinal assessment of tau and amyloid beta in cerebrospinal fluid of parkinson disease," *Acta neuropathologica*, vol. 126, pp. 671–682, 2013.

[27] L. Parnetti, S. Paciotti, P. Eusebi, A. Dardis, S. Zampieri, D. Chiasserini, A. Tasegian, N. Tambasco, B. Bembi, P. Calabresi, *et al.*, "Cerebrospinal fluid β-glucocerebrosidase activity is reduced in parkinson's disease patients," *Movement Disorders*, vol. 32, no. 10, pp. 1423–1431, 2017.

[28] M. Heijmans, J. Habets, M. Kuijf, P. Kubben, and C. Herff, "Evaluation of parkinson's disease at home: Predicting tremor from wearable sensors," in *2019 41st annual international conference of the IEEE engineering in medicine and biology society (EMBC)*, pp. 584–587, IEEE, 2019.

[29] R. I. Griffiths, K. Kotschet, S. Arfon, Z. M. Xu, W. Johnson, J. Drago, A. Evans, P. Kempster, S. Raghav, and M. K. Horne, "Automated assessment of bradykinesia and dyskinesia in parkinson's disease," *Journal of Parkinson's disease*, vol. 2, no. 1, pp. 47–55, 2012.

[30] L. Lonini, A. Dai, N. Shawen, T. Simuni, C. Poon, L. Shimanovich, M. Daeschler, R. Ghaffari, J. A. Rogers, and A. Jayaraman, "Wearable sensors for parkinson's disease: which data are worth collecting for training symptom detection models," *NPJ digital medicine*, vol. 1, no. 1, p. 64, 2018.

[31] W. Liu, X. Lin, X. Chen, Q. Wang, X. Wang, B. Yang, N. Cai, R. Chen, G. Chen, and Y. Lin, "Vision-based estimation of mds-updrs scores for quantifying parkinson's disease tremor severity," *Medical Image Analysis*, p. 102754, 2023.

[32] M. Lu, Q. Zhao, K. L. Poston, E. V. Sullivan, A. Pfefferbaum, M. Shahid, M. Katz, L. M. Kouhsari, K. Schulman, A. Milstein, *et al.*, "Quantifying parkinson's disease motor severity under uncertainty using mds-updrs videos," *Medical image analysis*, vol. 73, p. 102179, 2021.

[33] Y. Yang, Y. Yuan, G. Zhang, H. Wang, Y.-C. Chen, Y. Liu, C. G. Tarolli, D. Crepeau, J. Bukartyk, M. R. Junna, *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.

[34] A. Beuter and R. Edwards, "Using frequency domain characteristics to discriminate physiologic and parkinsonian tremors," *Journal of clinical neurophysiology*, vol. 16, no. 5, p. 484, 1999.

[35] R. San-Segundo, H. Blunck, J. Moreno-Pimentel, A. Stisen, and M. Gil-Martín, "Robust human activity recognition using smartwatches and smartphones," *Engineering Applications of Artificial Intelligence*, vol. 72, pp. 190–202, 2018.

[36] R. Khusainov, D. Azzi, I. E. Achumba, and S. D. Bersch, "Real-time human ambulation, activity, and physiological monitoring: Taxonomy of issues, techniques, applications, challenges and limitations," *Sensors*, vol. 13, no. 10, pp. 12852–12902, 2013.

[37] W. Gibb and A. Lees, "The relevance of the lewy body to the pathogenesis of idiopathic parkinson's disease.," *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 51, no. 6, pp. 745–752, 1988.

[38] X. Li, Y. Xing, A. Martin-Bastida, P. Piccini, and D. P. Auer, "Patterns of grey matter loss associated with motor subscores in early parkinson's disease," *NeuroImage: Clinical*, vol. 17, pp. 498–504, 2018.

[39] C. G. Goetz, G. T. Stebbins, and B. C. Tilley, "Calibration of unified parkinson's disease rating scale scores to movement disorder society-unified parkinson's disease rating scale scores," *Movement disorders*, vol. 27, no. 10, pp. 1239–1242, 2012.

[40] Y. Zhang, V. Lanfranchi, X. Wang, M. Zhou, and P. Yang, "Modeling alzheimer's disease progression via amalgamated magnitude-direction brain structure variation quantification and tensor multi-task learning," in *2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 2735–2742, IEEE, 2022.

[41] P. Yang, J. Qi, S. Zhang, X. Wang, G. Bi, Y. Yang, and B. Sheng, "Feasibility study of mitigation and suppression intervention strategies for controlling covid-19 outbreaks in london and wuhan," 2020.

[42] Y. Zhang, X. Wang, T. Liu, R. Wang, Y. Li, Q. Xue, and P. Yang, "Sustainable fertilisation management via tensor multi-task learning using multi-dimensional agricultural data," *Journal of Industrial Information Integration*, vol. 34, p. 100461, 2023.

[43] T. Liu, X. Wang, P. Yang, S. Chen, and C. J. Harris, "Unsupervised transfer aided lifelong regression for learning new tasks without target output," *IEEE Transactions on Knowledge and Data Engineering*, 2024.

[44] M. Zhou and P. Yang, "Automatic temporal relation in multi-task learning," in *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 3570–3580, 2023.

[45] X. Wang, J. Qi, Y. Yang, and P. Yang, "A survey of disease progression modeling techniques for alzheimer's diseases," in *2019 IEEE 17th International Conference on Industrial Informatics (INDIN)*, vol. 1, pp. 1237–1242, IEEE, 2019.

[46] L. Tao, X. Wang, F. Nan, J. Qi, Y. Yang, and P. Yang, "Effective severity assessment of parkinson's disease with wearable intelligence using free-living environment data," in *2023 IEEE 32nd International Symposium on Industrial Electronics (ISIE)*, pp. 1–10, IEEE, 2023.

[47] S. Staras, J. S. Tauscher, N. Rich, E. Samarah, L. A. Thompson, M. M. Vinson, M. J. Muszynski, E. A. Shenkman, *et al.*, "Using a clinical workflow analysis to enhance ehealth implementation planning: tutorial and case study," *JMIR mHealth and uHealth*, vol. 9, no. 3, p. e18534, 2021.

[48] Y. K. Alotaibi and F. Federico, "The impact of health information technology on patient safety," *Saudi medical journal*, vol. 38, no. 12, p. 1173, 2017.