A Hierarchical Architecture for Multisymptom Assessment of Early Parkinson's Disease via Wearable Sensors

Chen Wang, Liang Peng, Zeng-Guang Hou, *Fellow, IEEE*, Yanfeng Li, Ying Tan, and Honglin Hao

Abstract—Parkinson's disease (PD) is the second most common neurodegenerative disorder and the heterogeneity of early PD leads to interrater and intrarater variability in observation-based clinical assessment. Thus, objective monitoring of PD-induced motor abnormalities has attracted significant attention to manage disease progression. Here, we proposed a hierarchical architecture to reliably detect abnormal characteristics and comprehensively quantify the multisymptom severity in patients with PD. A novel wearable device was designed to measure motor features in 15 PD patients and 15 age-matched healthy subjects, while performing five types of motor tasks. The abnormality classes of multimodal measurements were recognized by hidden Markov models (HMMs) in the first layer of the proposed architecture, aiming at motivating the evaluation of specific motor manifestations. Subsequently, in the second layer, three single-symptom models differentiated PD motor characteristics from normal motion patterns and quantified the severity of cardinal PD symptoms in parallel. In order to further analyze the disease status, the multilevel severity quantification was fused in the third layer, where machine learning algorithms were adopted to develop a multisymptom severity score. The experimental results demonstrated that the quantification of three cardinal symptoms was highly accurate to distinguish PD patients from healthy controls. Furthermore, strong correlations were observed between the Unified PD Rating Scale (UPDRS) scores and the predicted subscores for tremor ($R = 0.75$, $P = 1.40e - 3$), bradykinesia ($R = 0.71$, $P = 2.80e - 3$), and coordination impairments ($R = 0.69$, $P = 4.20e - 3$), and the correlation coefficient can be enhanced to 0.88 ($P = 1.26e - 5$) based on the fusion schemes. In conclusion, the proposed assessment architecture holds great promise to push forward the in-home monitoring of clinical manifestations, thus enabling the self-assessment of disease progression.

Index Terms—Machine learning, multilevel fusion, multisymptom assessment, Parkinson's disease (PD), wearable sensor system.

I. INTRODUCTION

PARKINSON'S disease (PD) is a progressive neurodegenerative disease characterized by pathological disappearance of dopaminergic neurons in substantia nigra [1], [2], and affects over six million worldwide [3]. The clinical phenotype of PD encompasses cardinal motor symptoms, such as tremor, bradykinesia, rigidity, postural instability, and gait disturbance [4], [5], which directly impact the quality of life for the millions of PD patients. These cardinal symptoms can be primarily treated by dopaminergic treatments (i.e., infusion of levodopa-carbidopa intestinal gel [6] and Deep Brain Stimulation [7]), and the clinical assessment of heterogeneous symptoms is highly valuable for optimal adjustment of treatment in patients with PD [8], [9]. Typically, continuous monitoring and assessment of early symptom development provide foundations for clinicians to detect motor alterations and predict the disease progression.

At present, the clinical assessment of PD symptoms is manually performed by clinicians according to rating scales [e.g., unified Parkinson's disease rating scale (UPDRS) [10], unified Dyskinesia rating scale (UDysRS) [11], and abnormal involuntary movement scale (AIMS) [12]], of which the UPDRS is most widely accepted. These conventional scales are mostly performed in the hospital not representative of the home environment, which also suffer of interrater and intrarater variability due to personal experience and perceptual bias [10]. Furthermore, discrete characteristics of rating scales lead to the ignorance of subtle alterations during PD progression [13], which compromises sensitive and quantitative multisymptom evaluation in patients afflicted by this neurodegenerative disorder. Therefore, automatically monitoring and evaluating PD-induced impairments have attracted considerable interest to address these unmet needs.

TABLE I
OVERVIEW OF STUDIES IN THE ASSESSMENT OF PD

Reference	Patient Set (PD/HC)	PD Symptoms	Sensors	Label Method	Model	Performance
[14]	30/0	Bradykinesia and Dyskinesia	Accelerometers	UPDRS	Convolutional Neural Networks	86.88% (Accuracy)
[15]	21/0	Freezing of Gait	Accelerometers	UPDRS	Support Vector Machines	80.10%/88.10% (Sensitivity/ Specificity)
[16]	13/12	Dyskinesia	Accelerometers, Gyroscopes, and Magnetometers	UDysRS	Sensor-based Metric Extraction	0.55-0.99 (Correlation Coefficients)
[17]	20/0	Essential Tremor	Accelerometers and Gyroscopes	Tremor Rating Scale (wTRS)	General Linear Model	0.67-0.90 (Correlation Coefficients)
[18]	949/866	Bradykinesia	Accelerometers and Touch-screens	Self-labeled	Deep Neural Networks	65.70%±1.05% (AUC)
[19]	21/23	Bimanual Coordination Problems	Accelerometers and Touch-screens	UPDRS	Statistical Analysis	81.00%/81.00% (Sensitivity/ Specificity)
[20]	25/20	Tremor	Accelerometers and Gyroscopes	UPDRS	Random Forest of Decision Trees	82.00%/90.00% (Sensitivity/ Specificity)

Recently, wearable technology has been widely employed in PD-related applications. Wearable sensors, such as accelerometers, gyroscopes, and motion tracking systems, were used to objectively measure motor abnormalities associated with PD out-of-clinic (e.g., see Table I). Among these studies, one or two specific motor symptoms have been considered for disease identification or severity estimation based on relatively limited tasks, including finger tapping tests, hand pronation/supination movements, and postural sway. Commonly, the accelerometer attached on joints of extremities is an effective means of tremor measurements (e.g., amplitude and frequency), and angular velocities recorded by gyroscopes can reflect the severity of bradykinesia. In order to comprehensively monitor and manage disease progression in the early stages of PD, simultaneous multisymptom quantification should be emphasized during activities of daily living.

The machine learning approach is an emerging technology typically used to quantify the severity of motor abnormalities in PD patients, which can be broadly summarized into two categories. The first extracts related parameters from external sensor data to predict the severity of PD symptoms based on UPDRS guidelines. For example, Kim et al. [21] derived the tapping velocities and range-of-motion of fingers from gyroscope signals during finger tapping tests, which showed a correlation with clinical scales. The second involves regression/classification models between sensor features and UPDRS scores, and focuses on distinguishing abnormal and normal movements for disease diagnosis. The systems in Patel et al. [9], for instance, used the support vector machine (SVM) to distinguish PD patients from healthy controls based on accelerometer data. Giuberti et al. [22] estimated the

leg agility using SVM and K nearest neighbors (kNNs) for automated UPDRS assessment. However, these systems only generate one-sided information for clinical decision making, which are not straightforward to be interpreted and associated with behavioral abnormalities out-of-clinic. Hence, we aim for an easy-to-understand assessment of multisymptom severity in which the level of motor impairments can be continuously analyzed for PD patients.

In this study, we introduce a hierarchical architecture to automatically quantify the severity of cardinal symptoms in PD patients. As tremor, bradykinesia, and postural disturbance are the common initial PD-induced symptoms, a novel wearable device was developed to collect multimodal signals during five types of motor tasks that lead to the manifestations of specific symptoms. In order to address the unmet clinical need of objective monitoring of multiple symptoms in daily activities of PD patients, five hidden Markov models (HMMs) were trained to automatically recognize symptom-specific movements in the first layer. Subsequently, the severity of single symptoms was independently quantified based on the predicted abnormality classes in the second layer. For the comprehensive multisymptom assessment, we further fused the multilevel severity estimations using machine learning algorithms to generate a multisymptom severity score (MSS), which was capable of providing a valid support for the early prediction of PD.

Our contributions in quantitatively evaluating the multiple motor impairments in PD patients out-of-clinic are composed of three key methodological achievements.

1) A measurement system was designed to track the motion of upper and lower extremities during different motor
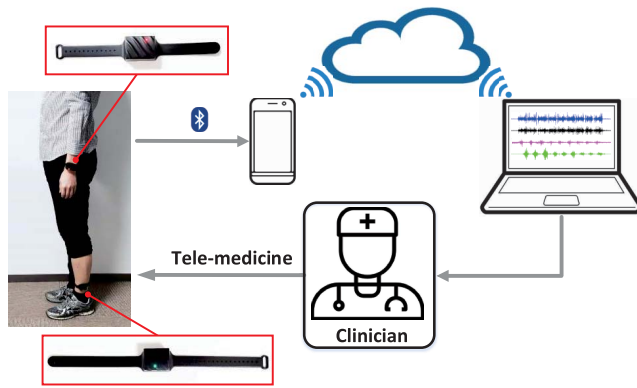
Fig. 1. Overview of the assessment system. Two wearable bands were attached at the wrist and ankle. Data gathered from all sensors were synchronously transmitted to the patient's smartphone and rendered on the cloud, which enables the clinician to remotely monitor the disease progression.

TABLE II
SUBJECT DEMOGRAPHICS

	PD Subjects		Healthy Controls	
	Males	Females	Males	Females
Numbers	9	6	8	7
Age (yrs)	72.1 ± 8.7	69.1 ± 5.1	66.3 ± 7.9	71.7 ± 9.5
UPDRS-III	22.3 ± 9.3	20.7 ± 7.1	-	-
H&Y	1.9 ± 1.3	2.2 ± 0.4	-	-

* UPDRS-III = Unified Parkinson's Disease Rating Scale, H&Y = Hoehn and Yahr scale; All standardized clinical tests were carried out prior to the experiment.

tasks, in which PD-induced symptoms became clinically manifest.

2) The simultaneously measured motor features were characterized by an HMM-based approach, and then different abnormality classes were detected to trigger the single-symptom models in the next layer.

3) The severity estimations of single motor symptoms were integrated by multilevel fusion schemes using the artificial neural network, and kernel-based and ensemble-learning algorithms, with the aim to yield a more comprehensive analysis of functional status.

The remainder of this study is organized as follows. Section II introduces experimental methods, consisting of the details of the wearable device, experimental setup, and data acquisition. Section III presents the hierarchical architecture for multisymptom assessment. Then, the experimental results are detailed in Section IV. Finally, Section V discusses the results and concludes this article.

II. EXPERIMENTAL METHODS

A. Measurement System

We have developed a novel wearable Parkinson's assessment device (wPAD) (see Fig. 1), including two bands worn at the patient's wrist and ankle, respectively. Each band is equipped with a MEMS inertial sensor (MPU9250, InvenSense) including a 3-D accelerometer and a 3-D gyroscope. The data range of the linear acceleration and angular velocity was ±8G ($G = 9.81$ m/s^2) and ±1000 dps (degree per second) in x-, y-, and z-axes, respectively. The accuracy, precision, and reliability of the data acquisition performance were validated in our previous study [23].

The acquisition of all data was performed at a sampling rate of 1000 Hz and synchronously transmitted to a personal computer at 200 Hz. A Bluetooth communication module (Bluetooth 4.0) was used to stream the data on a smartphone application, which allows PD patients engage in any daily activities. In addition, we adopted cloud services to wirelessly send the data to the personal computer of the clinician, and thus, provide foundations for telemedicine.

B. Participants and Clinical Testing

Fifteen patients diagnosed with PD were recruited from the Peking Union Medical College Hospital, and 15 healthy age-matched control subjects were also included in this study. The demographic information of 30 subjects is presented in Table II.

The inclusion criteria for the PD subjects selection were: 1) idiopathic PD diagnosed according to the U.K. PD Society Brain Bank criteria; 2) disease progression ranges from 1 to 2.5 according to the Hoehn and Yahr scale; 3) suffering from PD for less than three years; 4) no severe cognitive impairment; and 5) able to visit the clinic for experiments. All PD subjects were clinically evaluated by an experienced clinician using Hoehn and Yahr scale and UPDRS Sections II and III prior to participating.

The inclusion criteria for the control subjects were free of dementia, neurological, or psychiatric disorders, and had no first-degree relatives with PD. The study was reviewed and approved by the Ethics Committee of Chinese Academy of Medical Sciences and Peking Union Medical College, and written informed consent was provided by each subject after notified of experimental procedures.

C. Procedure

As PD progresses, motor symptoms often manifest themselves more severely on one side of the body than on the other side [24]. Therefore, the wPAD was worn on the wrist and ankle of the most affected side for PD patients. It should be noted that patients were recorded in the medication OFF state (i.e., at least 12 h after their PD medication), and healthy subjects accomplished all assessment by mounting the device on the nondominant body side.

All subjects were instructed to perform five motor tasks at a time for a minimum of 20 s for each task (see Fig. 2).

1) *Resting:* The subject stayed seated with both forearms resting on the arms of a chair, and maintained this posture for 20 s.

2) *Extending Hand Movement:* The subject stayed seated and fully extended both arms in front of them with palms facing toward the ground at the shoulder height level.

3) *Alternating Hand Movement:* While sitting, the subject first extended both arms in front of them, and then rhythmically pronated/supinated hands with palms facing up and down as fast as possible.
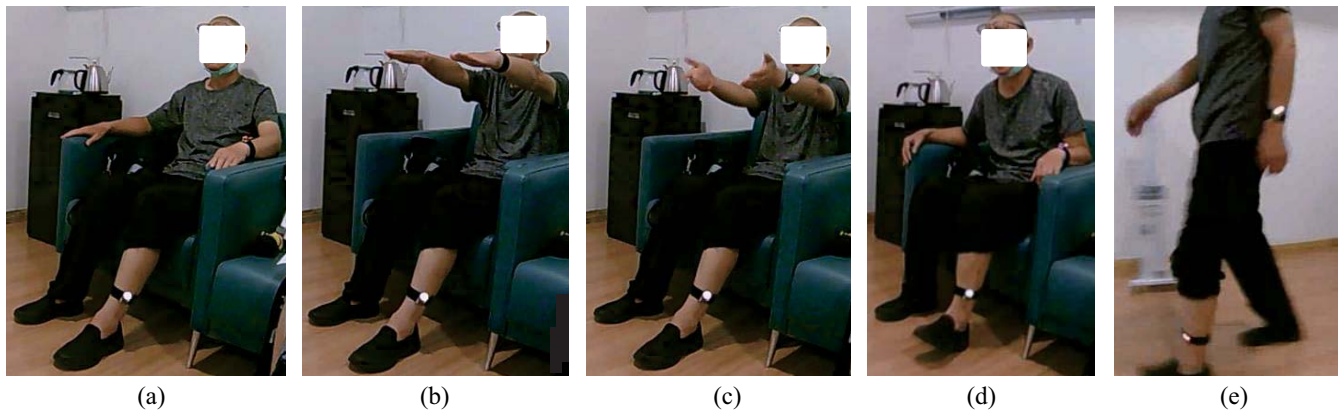
Fig. 2.   Overview of the experimental setup. All participants were asked to perform five types of motor tasks: (a) resting, (b) extending hand, (c) alternating hand, (d) tapping heel-toe, and (e) walking while wearing the developed Parkinson's assessment device.

4) *Tapping Heel-Toe Movement:* The subject was seated as in Task 1, and rhythmically taped heel-toe on the floor with their ankles rotating as complete as possible.
5) *Walking:* The subject started from a sitting position, walked straight ahead for 10 s, made a turn, and walked back for 10 s.

These tasks were selected according to standardized clinical practice with aims to assess multiple manifestations in patients affected by PD, such as rest tremor, postural tremor, bradykinesia, postural instability, and gait disturbance. More concretely, Task 1 and Task 2 are relevant to components III. 20 and III. 21 of the UPDRS, which focus on the evaluation of the symptom severity for rest and postural tremor, respectively. Task 3 and Task 4 correspond to components III. 25 and III. 26 of the UPDRS, which emphasize abnormalities in rhythm, velocity, and amplitude of upper and lower limb movements to estimate severity of bradykinesia. Task 5 can highlight the postural sway, trunk rotation, and gait problems (e.g., stride velocity and gait variability).

D. Data Preprocessing and Segmentation

In the following, signals in two modalities were smoothed using a fourth-order Butterworth low-pass filter (cutoff frequency: 10 Hz). After removing the noises due to high-frequency sudden changes, we normalized the data from three axes by min–max scaling for further data analysis. In order to obtain continuous features from five different tasks, the multimodal data were segmented into a series of fixed-length windows, in which the overlapping technique was employed to reduce the delay between adjacent windows [25]. The duration of each window was selected as 200 ms, and the overlapping ratio was 50%. Therefore, for an individual subject, 199 samples were generated for each motor task, and the total number of samples was 995 in the five-class problem, which was sufficient for severity estimation in our study.

III. HIERARCHICAL ARCHITECTURE FOR MULTISYMPTOM ASSESSMENT OF PARKINSON'S DISEASE

A. Symptom-Specific Movement Recognition

In this section, we presented an HMM-based approach to automatically recognize the abnormality classes of motor tasks

(i.e. Task 1-5), since each task induced specific clinical manifestation in PD patients, further symptom-specific quantification for rest tremor, postural tremor, upper-limb bradykinesia, lower-limb bradykinesia and coordination impairments can be triggered by this layer (see Fig. 3). In order to predict the underlying symptom class from continuous sensor data, we used HMMs to characterize the doubly embedded stochastic process [26], and estimate the precise parameters in this process.

An HMM is characterized by the observation sequence O and a hidden state sequence Q according to a state transition probability distribution $A = \{a_{ij}\}$, observation symbol probability distribution $B = \{b_j(k)\}$, and initial state distribution $\pi = \{\pi_i\}$ as follows [27]:

$$\lambda = (A, B, \pi) \tag{1}$$

where

$$a_{ij} = P\big[q_{t+1} = S_j | q_t = S_i\big], 1 \leq i, j \leq N$$
$$b_j(k) = P\big[v_k \, at \, t | q_t = S_j\big], 1 \leq j \leq N, 1 \leq k \leq M$$
$$\pi_i = P\big[q_1 = S_i\big], 1 \leq i \leq N$$

with

$$S = \{S_1, S_2, \ldots, S_N\}$$
$$V = \{v_1, v_2, \ldots, v_M\}$$

where N and M stand for the number of hidden states and distinct observation symbols per state in the model.

In this study, we encoded the probability density function of observed multimodal data using the Gaussian mixture model (GMM) [28]

$$b_j(O) = \sum_{c=1}^{C} \omega_{jc} \psi\left(O, \mu_{jc}, \sum_{jc}\right), 1 \leq j \leq N \tag{2}$$

where C stands for the number of mixture components, ω_{jc} is the kth mixture coefficient for state j, ψ and is the cth Gaussian component with mean vector μ_{jc} and covariance matrix $\sum_{jc}$. Specifically, the parameters of GMM were optimized based on the K-means algorithm.

Given the continuous observation symbols as training data, the HMM parameters were iteratively estimated based on the expectation–modification (EM) method [29] so that the model
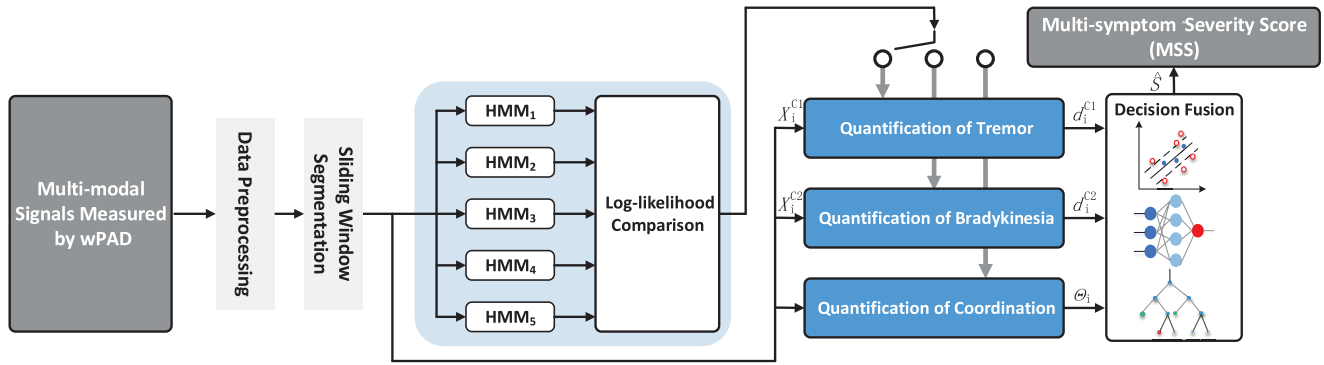
Fig. 3. Hierarchical architecture for assessing the multisymptom severity for patients with PD. The accelerometer and gyroscope data served as the input of the proposed architecture. After recognizing the symptom-specific movements in the HMM-based layer, the quantification of the tremor d_i^{C1}, bradykinesia d_i^{C2}, and coordination impairments Θ_i was motivated correspondingly in the second layer. Each of the single-symptom models estimate the severity in the range of 0–1. In the third layer, we fused multilevel severity estimations to generate the MSS for the comprehensive assessment.

TABLE III
PREDICTIONS OF THE HMM-BASED LAYER

	Task 1	Task 2	Task 3	Task 4	Task 5
TSC	1	1	0	0	0
BSC	0	0	1	1	0
CSC	0	0	0	0	1

* TSC = Tremor-Specific Classification, BSC = Bradykinesia-Specific Classification, CSC = Coordination-Specific Classification.

can best represent the time-varying nature of motion measurements. Note that the number of hidden states was set as five by fivefold cross-validation, while the number of Gaussian functions and the value of K were set as 5 to result in the good recognition performance but without increasing computational expensive. In the movement recognition layer, we started by developing five HMMs corresponding to symptom-specific movements, thereby motivating the recognition process for an unknown observation sequence. The probability of that sequence produced by each model was then computed using the forward–backward procedure, i.e., a score of how well each model matched a given sequence. By finding the highest log likelihood, the predicted class can be determined among five competing HMMs, and Table III describes all possible predictions made by this layer.

B. Single-Symptom Quantification

To estimate the severity of cardinal PD symptoms in parallel, we developed the single-symptom quantification layer using data in two modalities (see Fig. 3). More specifically, the quantification of the symptom severity for tremor and bradykinesia was performed by supervised machine learning approaches, in which the feedforward neural network (FFNN), SVM, and random forest (RF) served as representatives of feedforward artificial neural network, and kernel-based and ensemble-learning classification algorithms.

In case the tracked movements in current window was recognized as the tremor-related class (i.e., Task 1 and Task 2)

or bradykinesia-related class (i.e., Task 3 and Task 4) in the first layer, the tremor-specific classifier or bradykinesia-specific classifier was triggered accordingly. We extracted the feature vector from the ith segment by concatenating the preprocessed accelerometer and gyroscope data, and defined the corresponding labels as $+1$ for PD group and -1 for healthy group. Hence, the input vector $X_i^{C_m}$ fed to the single-symptom classifier C_m can be expressed as

$$X_i^{C_m} = \left[\zeta_1^{C_m}, \ldots, \zeta_n^{C_m}, \ldots, \zeta_{N_l^{C_m}}^{C_m} \right]^T \qquad (3)$$

where $\zeta_n^{C_m}$ is the nth motor feature, and the total number of features $N_l^{C_m}$ is 480 for two types of classifiers. The predicted probability vector $d_i^{C_m}$ for severity estimation of tremor and bradykinesia can be obtained

$$d_i^{C_m} = f^{C_m}\left(X_i^{C_m}\right) \qquad (4)$$

where f^{C_m} denotes the candidate models constructed by FFNN, SVM, and RF. Specifically, the FFNN classifier comprised an input layer with 480 nodes, a hidden layer with 27 nodes, and an output layer with 1 node. The last two layers selected the hyperbolic tangent and sigmoid as their activation functions. The SVM classifier was trained with a radial basis function kernel, and Platt's method [30] was adopted to transform outputs to posterior probabilities. The RF decided the symptom-specific classification result based on the majority vote of ten decision trees.

Suppose the current sample belongs to coordination-related class, the coordination of upper limb and lower limb movements was quantified by a synergy-based classifier. Synergy is a manifestation of the motor control strategies adopted by the central nervous system [31], [32], and our previous study has expanded the concept of interlimb synergies, which play a major role in motor function assessment [33]. In order to better analyze PD-induced postural characteristics during walking, we started by calculating the correlation matrix from the motion segment, thereby weighting each observed variable equally. The interlimb synergies were then identified by using principal component analysis

(PCA) to extract a set of uncorrelated variables from all 12 channels, and these new linear combinations can characterize the coordination ability in individuals affected by PD.

Therefore, instead of calculating principal components, we quantified the altered movement patterns based on k weighted angular similarity (kWAS) algorithm [34], which is capable of uncovering the angular similarity between the eigenvectors obtained from PD subjects and healthy controls

$$\Theta_i(S_i, S_{hc}) = 1 - \frac{1}{2}\sum_{p=1}^{n}\left[\left(\frac{\sigma_p}{\sum_{q=1}^{n}\sigma_q} + \frac{\lambda_p}{\sum_{q=1}^{n}\lambda_q}\right)|u_p \cdot v_p|\right] \quad (5)$$

where S_i denotes the motor synergies of the ith segment extracted from an individual subject, S_{hc} denotes the averaged synergies derived from healthy controls, u_p and v_p are the pth eigenvectors of the corresponding correlation matrices, and σ_q and λ_q are the respective eigenvalues. Note that n was set to 12 to involve all movement patterns derived from multimodal signals into comparison. It can be seen that the value of Θ_i ranges from 0 to 1, and becomes lower as the coordination ability approaches normal status.

C. Multilevel Fusion

Based on the quantification of the multisymptom severity (i.e., rest tremor, postural tremor, bradykinesia, and postural disturbance), we constructed a fusion layer to combine the multilevel features (see Fig. 3). Unlike the classification models in the second layer, the regression models based on FFNN, SVR, and RF were adopted to facilitate a more comprehensive quantification.

To begin with, the feature vector Y_j was constituted by concatenating the output vectors derived from tremor-specific classifier, bradykinesia-specific classifier, and coordination-specific classifier

$$Y_j = \left[\tau_1, \ldots, \tau_n, \ldots, \tau_{N_l^F}\right]^T \quad (6)$$

where τ_n is the severity estimation from the nth symptom-related class, and N_l^F stands for the five classes. Note that the predicted probabilities from each class were sequenced according to the chronological order of motion segments. We referred to ground truth as UPDRS-III scores given by the experienced clinician, therefore, the prediction for the overall motor abnormalities can be expressed as follows:

$$D_j = \eta^F(Y_j) \quad (7)$$

where η^F represent the candidate fusion algorithms. Concretely, the FFNN was defined as a standard three-layer feedforward network connected with hyperbolic tangent and linear transfer functions in the hidden and output layers. The node numbers of the three layers were determined as 5, 9, and 1, respectively. Support vector regression (SVR) consisted of a sigmoid kernel to train probabilistic models, and RF decided the regression result based on the majority vote given by four decision trees.

Furthermore, the severity predictions of multiple motor symptoms per patient were fused using the average rule [35]

$$\hat{S} = \frac{1}{J}\sum_{j=1}^{J}D_j \quad (8)$$

where $\hat{S}$ refers to the MSS that quantitatively assesses the motor abnormalities appearing mainly with PD patients in the early stages. Note that all MSSs are nonnegative, and a higher score indicates that the PD disease is more severe. As characteristics derived from three cardinal symptoms were all emphasized in the fusion layer, the MSS can be viewed as an overall understanding of the PD-induced symptom severity for a specific patient.

The hierarchical assessment architecture was trained and tested with the following strategy. The full data set, consisting of 29 850 segments collected from 30 individuals, was first separated into two nonoverlapping parts: one group comprised the training data set (i.e., a total of 17 910 segments from 18 individuals), and the other group including the remaining segments was used as the test data set, with the aim to demonstrate the generalization performance of the architecture for unknown patients. The five HMMs in the movement recognition layer were first trained independently using the training data set. For the single-symptom quantification layer, assuming that all segments were predicted correctly, we further split the training data set belonging to tremor and bradykinesia classes: 60% of the labeled data served as the input of the six candidate models for tremor-specific classifier or bradykinesia-specific classifier, and 40% were used to determine the optimal tremor-specific and bradykinesia-specific models. Subsequently, the predicted probabilities of these optimal models and their respective coordination ability scores were concatenated into the feature vectors for the three candidate fusion models, and the best regression model was selected based on the quantification performance. Finally, the remaining 40% of the full data set were used for test, which confirm the optimal model selection. It should be noted that hyperparameters of classification and regression models were optimized by fivefold cross-validation procedures. For example, the FFNN-based tremor assessment model was trained using the gradient descent (traingd) algorithm with a learning rate of 0.06. The penalty parameter and kernel function parameter of the SVM-based bradykinesia assessment model were, respectively, set as 0.315 and 0.002, as well as 0.100 and 0.160 for the SVR-based fusion model.

IV. RESULTS

A. Recognition Performance

The architecture generalizability to unknown subjects was first tested for each layer, and then the total performance over layers was evaluated. In this section, we reported the recognition ability of symptom-specific detectors in terms of the following performance metrics:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FN} + \text{FP} + \text{TN}} \quad (9)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (10)$$

Fig. 4. Recognition results for five types of symptom-specific movements that lead to manifestations of rest tremor, postural tremor, upper limb bradykinesia, lower limb bradykinesia and coordination impairments, respectively.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{11}$$

$$F1\text{ - Measure} = \frac{\text{Recall} \times \text{Precision} \times 2}{\text{Recall} + \text{Precision}} \tag{12}$$

where TP and FP denote the numbers of true and false PD patient identification, respectively; and TN and FN denote the numbers of true and false healthy person identification, respectively.

We started by analyzing the recognition performance among the five motor tasks using the independent test data set, and the confusion matrix of the estimated classes is shown in Fig. 4. The highest recall was obtained by HMM5 with an overall precision of 92.19%, mainly due to the discriminating features from trunk rotation, walking speed, and cadence. HMM3 performed poorly for the upper limb bradykinesia-related class, as 1.59% of the segments in this class were misclassified as the postural tremor-related class. The reason for the missclassification error can be explained by the similarity between extending and alternating hand movements for the PD patients in the severe stages. There were also limited confusions (1.33%) between the resting and extending hand movement, which have little effect on the severity estimation in the next layer.

In addition, Fig. 5 presents the confusion matrix for the three-class recognition problem (i.e., tremor, bradykinesia, and coordination-related movements). The recognition confusion mainly occurred between the tremor-related and bradykinesia-related classes, and the interlimb coordination class was more accurately recognized, which were consistent with the five-class recognition results. Consequently, for all PD and control subjects, the confusion matrices confirmed that motion segments can be accurately classified; therefore, this layer was capable of triggering the further quantification of corresponding motor symptoms.

B. Classification Performance

In the following, we investigate the classification reliability of PD patients versus healthy controls obtained with the

Fig. 5. Integrated recognition results for three types of symptom-specific movements that lead to manifestations of tremor, bradykinesia, and coordination impairments, respectively.

TABLE IV
OVERALL ACCURACY (%), PRECISION (%), RECALL (%), AND
$F1$-MEASURE (%) OF TREMOR-SPECIFIC CLASSIFIERS

Performance	Tremor-Specific Classification		
Measures	FFNN	SVM	RF
Accuracy	89.28	88.32	85.49
Precision	88.23	86.77	83.39
Recall	96.32	96.74	96.93
F1-Measure	92.10	91.48	89.65

TABLE V
OVERALL ACCURACY (%), PRECISION (%), RECALL (%), AND
$F1$-MEASURE (%) OF BRADYKINESIA-SPECIFIC CLASSIFIERS

Performance	Bradykinesia-Specific Classification		
Measures	FFNN	SVM	RF
Accuracy	82.43	85.13	82.60
Precision	82.32	85.02	85.72
Recall	95.33	95.56	90.11
F1-Measure	88.35	89.98	87.86

proposed architecture. Inside the single-symptom quantification layer, three types of classifiers estimated the symptom severity for tremor, bradykinesia, and coordination impairments in parallel. Table IV reports overall accuracy, precision, recall, and $F1$-measure of the tremor-specific classifier on the test data set, and the best performance among supervised machine learning algorithms is shown in highlighted cells. It can be observed that FFNN outperformed with the highest accuracy (89.28% compared to 88.32% and 85.49%) and the best $F1$-measure (92.10% compared to 91.48% and 89.65%). Similarly, Table V shows the overall performance of the bradykinesia-specific classifier, and the results derived from the best model are highlighted. For the bradykinesia estimation, SVM resulted in better performance with the accuracy as high as 85.13%, which outperformed RF and FFNN
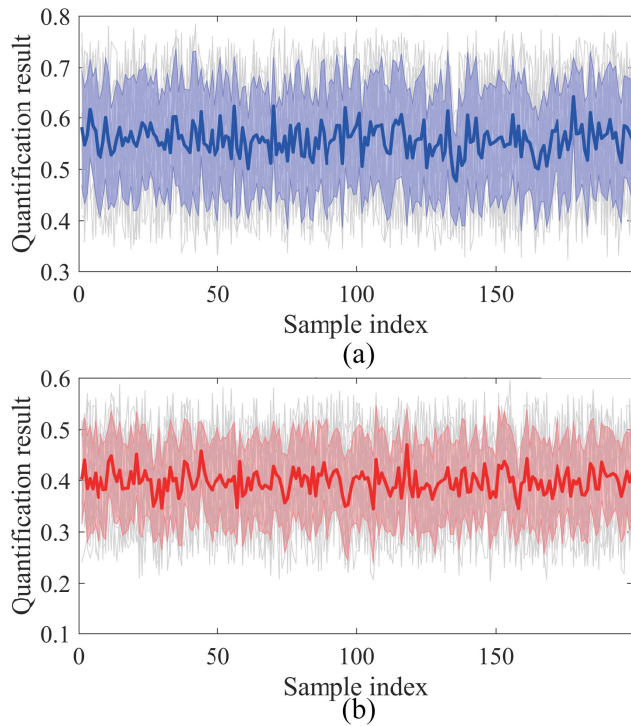
Fig. 6. Temporal evolutions for the quantification of interlimb coordination ability: (a) PD patients and (b) healthy controls. The thin gray lines indicate individual subject's quantification results. The blue line and shaded blue refer to the average evolution and standard deviation of quantification values across PD patients, respectively. The red line and shaded red refer to the average evolution and standard deviation of quantification values across healthy controls, respectively.
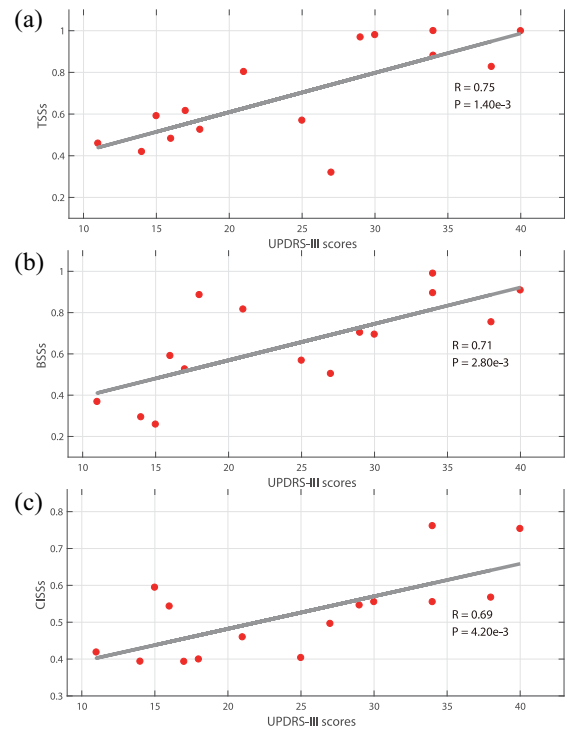

Fig. 7. Correlation analysis between the UPDRS-III scores and the single-symptom quantification results: (a) TSSs, (b) BSSs, and (c) CISSs. The Pearson correlation coefficient (R-value) and the P-value are indicated in each subfigure, respectively.

by a margin of 2.70% and 2.53%, respectively. It is worth mentioning that the best models were in agreement with the optimal selections during the hierarchical architecture training, i.e., FFNN for tremor-specific classification and SVM for bradykinesia-specific classification.

To investigate the classification performance of the coordination-specific classifier, the contiguous quantification of interlimb coordination ability were concatenated into a feature vector per individual. The statistical features extracted from the evolution of quantification results verified the classifier's generalizability to discriminate PD patients with respect to healthy controls, since a lower value represents the better functional status and *vice versa* (see Fig. 6). Hence, these promising results demonstrated that the hierarchical architecture can reliably distinguish between PD versus healthy motor characteristics, which provide a practical tool for the early diagnosis of PD in home environments.

C. Quantification Performance

Since our purpose is to enable objective and quantitative monitoring of PD-induced motor symptoms, it is essential to investigate whether the proposed metrics (i.e., single-symptom quantification and multisymptom quantification) exhibit significant correlations with clinical scores.

For the single-symptom assessment, we carried out the Pearson correlation test [36] with 95% confidence intervals to analyze the correlation between clinician-administered

UPDRS-III scores and the quantification results of three types of severity estimation models. More specifically, the predicted probabilities produced by the optimal tremor-specific and bradykinesia-specific classifiers were, respectively, integrated using the average rule for each PD patient, and the outputs of the coordination-specific classifier were similarly integrated to generate a single-symptom score. Fig. 7 depicts the results of correlation analysis for the tremor severity score (TSS), bradykinesia severity score (BSS), and coordination impairments severity score (CISS), in which the TSS and BSS were derived from the FFNN-based tremor assessment model and SVM-based bradykinesia assessment model. The corresponding Pearson correlation coefficients were 0.75 ($P = 1.40e - 3$), 0.71 ($P = 2.80e - 3$), and 0.69 ($P = 4.20e - 3$), which indicate that three types of single-symptom scores are well correlated to the standardized rating scale. As can be seen, the FFNN-based model captured the tremor-related motor characteristics better than two other estimation models, which was consistent with the superior performance in identifying the subjects affected by PD.

For the multisymptom assessment, we started by testing three candidate regression models in terms of the root mean square error (RMSE)

$$S_{\text{RMSE}} = \sqrt{\frac{1}{K} \sum_{k=1}^{K} \left(S_k^{RM} - S_k^{\text{UPDRS}}\right)^2} \quad (13)$$

where K is the size of test data set, S_k^{RM} is the kth severity predictions yielded by FFNN, SVR, and RF, and S_k^{UPDRS}

TABLE VI
RMSE OF THE HIERARCHICAL APPROACH

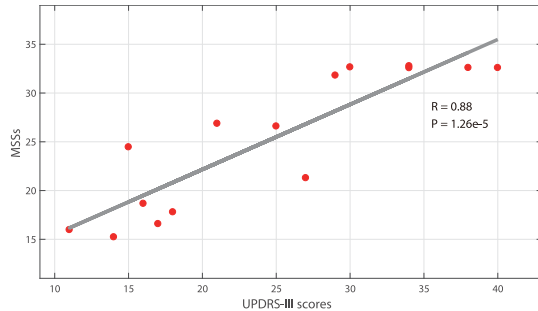| Performance Measures	Multi-Symptom Quantification		
	FFNN	SVR	RF
RMSE	4.75	3.55	4.17

Fig. 8. Correlation analysis between the UPDRS-III scores and the multisymptom quantification results. The MSSs were derived from the SVR-based model, which produced the best performance in the construction of the fusion layer.

is the corresponding UPDRS-III scores. Table VI reports the values of RMSE for all candidate models and highlights the best results in blue. The SVR-based model performed superiorly to both FFNN and RF, accompanied with the lowest RMSE. It is worth noting that SVR was also selected as the most suitable scheme for multilevel fusion on the training data set, in which the values of RMSE for the fusion models constructed using the FFNN, SVR and RF were 3.16, 2.81, and 2.97, respectively.

The correlation analysis was next performed to validate the association between UPDRS-III scores and MSSs for comparison purposes. Fig. 8 shows the results of correlation analysis and the corresponding Pearson correlation coefficients. We found that the predicted MSSs exhibited a significant relationship with the UPDRS-III scores ($R = 0.88$, $P = 1.26e\text{-}5$), which was even more successful in the quantification of motor symptom severity. Therefore, by fusing the complementarity of the multilevel motor characteristics, the clinical relevance of the proposed architecture can be enhanced. The results for single-symptom and multisymptom assessment have the potential to facilitate reliable medical-decision making, as well as implement the remote prediction of PD progression more easily in home environments.

V. DISCUSSION AND CONCLUSION

In this study, we proposed a hierarchical approach to comprehensively assess multiple motor symptoms in patients with early PD. Using a novel wearable device attached on the wrist and ankle, we were able to gather signals in two modalities during different motor tasks. As the PD-induced motor abnormalities are quite heterogeneous between patients, we put the emphasis on continuous severity estimation of cardinal motor symptoms in daily life. The experimental results demonstrated that the proposed architecture was capable of accurately identifying pathological motor characteristics and objectively quantifying the PD symptom severity.

In more detail, for the classification performance, the hierarchical architecture generally performed well for tremor and bradykinesia, which classified PD versus healthy subjects with accuracies of 89.28% and 85.13%, respectively. For the quantification performance, a significant correlation was observed between the estimated MSSs and the UPDRS-III scores ($R = 0.88$, $P = 1.26e - 5$). Therefore, the MSSs can take a global view of the functional status and disease progression of the PD patient. Our approach also yielded independent subscores for tremor, bradykinesia, and coordination impairments (i.e., TSS, BSS, and CISS), which provide targeted information about PD-induced abnormalities that impair motor functions. Furthermore, these encouraging results can support the self-assessment of PD patients during daily living, as well as guide the individualization of the subsequent treatment methods.

To our knowledge, ours is the first study to integrate movement recognition into the automated assessment system for home monitoring of patients with PD. Since symptom manifestations vary across activities and the proposed assessment process was performed in the free-living environment, the tracked movements needed to be identified as different abnormality classes to ensure that certain symptoms can manifest themselves most obviously. After recognizing the symptom-specific movements, the single-symptom models were triggered for the evaluation of three cardinal symptoms, and the performance metrics presented in Tables IV and V substantiated the superior performance of FFNN and SVM in tremor and bradykinesia detection, respectively.

Unlike the previous studies that have been so far limited to estimate the presence of behavioral abnormalities in a specific symptom class, our assessment architecture used a fusion scheme to further exploit the estimations of single-symptom models. Moreover, the analysis of locomotor coordination ability, considering the altered synergy patterns between upper and lower extremities, served as an input feature of the multilevel fusion model to enhance clinical relevance. By applying three types of regression models, we found that the prediction of the SVR-based fusion model was even more successful in the severity quantification, and showed a better correlation with standardized clinical tests, as visualized in Fig. 8.

In general, we concluded the proposed approach provides an easy-to-use tool that can objectively assess and monitor the disease progression in PD, enabling the in-home assessment and cloud-based healthcare. Since the severity estimation of rigidity needs a trained clinician to achieve a full range of motion for the patient's major body joint and characterize the encountered resistance, the wearable device composed of force sensors will be further developed to support the remote assessment for rigidity. In addition, due to the limited sample size, this study only presents a preliminary demonstration, and future work with larger sample size can improve the generalization performance of the architecture.

REFERENCES

[1] W. Dauer and S. Przedborski, "Parkinson's disease: Mechanisms and models," *Neuron*, vol. 39, no. 6, pp. 889–909, 2003.

[2] L. M. De Lau and M. M. Breteler, "Epidemiology of Parkinson's disease," *Lancet Neurol.*, vol. 5, no. 6, pp. 525–535, 2006.

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

[4] D. J. Surmeier, J. A. Obeso, and G. M. Halliday, "Selective neuronal vulnerability in Parkinson disease," *Nat. Rev. Neurosci.*, vol. 18, no. 2, pp. 101–113, 2017.

[5] J. Jankovic, "Parkinson's disease: Clinical features and diagnosis," *J. Neurol. Neurosurg. Psychiatry*, vol. 79, no. 4, pp. 368–376, 2008.

[6] C. W. Olanow *et al.*, "Continuous intrajejunal infusion of levodopa-carbidopa intestinal gel for patients with advanced Parkinson's disease: A randomised, controlled, double-blind, double-dummy study," *Lancet Neurol.*, vol. 13, no. 2, pp. 141–149, 2014.

[7] L. di Biase and A. Fasano, "Low-frequency deep brain stimulation for parkinson's disease: Great expectation or false hope?" *Movement Disord.*, vol. 31, no. 7, pp. 962–967, 2016.

[8] S. S. Papapetropoulos, "Patient diaries as a clinical endpoint in Parkinson's disease clinical trials," *CNS Neurosci. Ther.*, vol. 18, no. 5, pp. 380–387, 2012.

[9] S. Patel *et al.*, "Monitoring motor fluctuations in patients with parkinson's disease using wearable sensors," *IEEE Trans. Inf. Technol. Biomed.*, vol. 13, no. 6, pp. 864–873, Nov. 2009.

[10] B. Post, M. P. Merkus, R. M. de Bie, R. J. de Haan, and J. D. Speelman, "Unified Parkinson's disease rating scale motor examination: Are ratings of nurses, residents in neurology, and movement disorders specialists interchangeable?" *Movement Disord. Off. J. Movement Disord. Soc.*, vol. 20, no. 12, pp. 1577–1584, 2005.

[11] C. G. Goetz, J. G. Nutt, and G. T. Stebbins, "The unified dyskinesia rating scale: Presentation and clinimetric profile," *Movement Disord. Off. J. Movement Disord. Soc.*, vol. 23, no. 16, pp. 2398–2403, 2008.

[12] M. R. Munetz and S. Benjamin, "How to examine patients using the abnormal involuntary movement scale," *Hosp. Commun. Psychiat.*, vol. 39, no. 11, pp. 1172–1177, 1988.

[13] G. A. Kang, J. M. Bronstein, D. L. Masterman, M. Redelings, J. A. Crum, and B. Ritz, "Clinical characteristics in early Parkinson's disease in a central California population-based study," *Movement Disord. Off. J. Movement Disord. Soc.*, vol. 20, no. 9, pp. 1133–1142, 2005.

[14] T. T. Um *et al.*, "Data augmentation of wearable sensor data for Parkinson's disease monitoring using convolutional neural networks," in *Proc. 19th ACM Int. Conf. Multimodal Interact.*, 2017, pp. 216–220.

[15] D. Rodríguez-Martín *et al.*, "Home detection of freezing of gait using support vector machines through a single waist-worn triaxial accelerometer," *PLoS One*, vol. 12, no. 2, 2017, Art. no. e0171764.

[16] M. Delrobaei, N. Baktash, G. Gilmore, K. McIsaac, and M. Jog, "Using wearable technology to generate objective Parkinson's disease dyskinesia severity score: Possibilities for home monitoring," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 10, pp. 1853–1863, Oct. 2017.

[17] C. L. Pulliam *et al.*, "Continuous in-home monitoring of essential tremor," *Parkinsonism Related Disord.*, vol. 20, no. 1, pp. 37–40, 2014.

[18] J. Prince and M. De Vos, "A deep learning framework for the remote detection of Parkinson's disease using smart-phone sensor data," in *Proc. 40th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, 2018, pp. 3144–3147.

[19] T. Arroyo-Gallego *et al.*, "Detection of motor impairment in parkinson's disease via mobile touchscreen typing," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 9, pp. 1994–2002, Sep. 2017.

[20] N. Kostikis, D. Hristu-Varsakelis, M. Arnaoutoglou, and C. Kotsavasiloglou, "A smartphone-based tool for assessing Parkinsonian hand tremor," *IEEE J. Biomed. Health Inform.*, vol. 19, no. 6, pp. 1835–1842, Nov. 2015.

[21] J.-W. Kim *et al.*, "Quantification of bradykinesia during clinical finger taps using a gyrosensor in patients with Parkinson's disease," *Med. Biol. Eng. Comput.*, vol. 49, no. 3, pp. 365–371, 2011.

[22] M. Giuberti *et al.*, "Automatic UPDRS evaluation in the sit-to-stand task of parkinsonians: Kinematic analysis and comparative outlook on the leg agility task," *IEEE J. Biomed. Health Inform.*, vol. 19, no. 3, pp. 803–814, May 2015.

[23] R. Liu, L. Peng, L. Tong, K. Yang, and B. Liu, "The design of wearable wireless inertial measurement unit for body motion capture system," in *Proc. IEEE Int. Conf. Intell. Safety Robot. (ISR)*, Shenyang, China, 2018, pp. 557–562.

[24] R. Djaldetti, I. Ziv, and E. Melamed, "The mystery of motor asymmetry in Parkinson's disease," *Lancet Neurol.*, vol. 5, no. 9, pp. 796–802, 2006.

[25] V. Kober, "Fast algorithms for the computation of sliding discrete sinusoidal transforms," *IEEE Trans. Signal Process.*, vol. 52, no. 6, pp. 1704–1710, Jun. 2004.

[26] L. Rabiner, *Fundamentals of Speech Recognition*. Upper Saddle River, NJ, USA: Prentice-Hall, 1993.

[27] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proc. IEEE*, vol. 77, no. 2, pp. 257–286, Feb. 1989.

[28] D. A. Reynolds, "Gaussian mixture models," in *Encyclopedia of Biometrics*, vol. 741. Boston, MA, USA: Springer, 2009.

[29] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Stat. Soc. B, Methodol.*, vol. 39, no. 1, pp. 1–38, 1977.

[30] J. Platt *et al.*, "Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods," *Adv. Large Margin Classifiers*, vol. 10, no. 3, pp. 61–74, 1999.

[31] N. A. Bernstein, *The Co-Ordination and Regulation of Movements*. New York, NY, USA: Pergamon Press, 1966.

[32] W. A. Lee, "Neuromotor synergies as a basis for coordinated intentional action," *J. Motor Behav.*, vol. 16, no. 2, pp. 135–170, 1984.

[33] C. Wang, L. Peng, Z.-G. Hou, J. Li, T. Zhang, and J. Zhao, "Quantitative assessment of upper-limb motor function for post-stroke rehabilitation based on motor synergy analysis and multi-modality fusion," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 4, pp. 943–952, Apr. 2020.

[34] C. Li, S. Q. Zheng, and B. Prabhakaran, "Segmentation and recognition of motion streams by similarity search," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 3, no. 3, p. 16, 2007.

[35] J. Mendes-Moreira, C. Soares, A. M. Jorge, and J. F. D. Sousa, "Ensemble approaches for regression: A survey," *ACM Comput. Surv.*, vol. 45, no. 1, pp. 1–40, 2012.

[36] J. Benesty, J. Chen, Y. Huang, and I. Cohen, "Pearson correlation coefficient," in *Noise Reduction in Speech Processing*. Heidelberg, Germany: Springer, 2009, pp. 1–4.

Chen Wang received the B.E. degree in automation from China University of Mining and Technology, Jiangsu, China, in 2016, and the Ph.D. degree in control theory and control engineering from the University of Chinese Academy of Sciences, Beijing, China, in 2021.

She is an Assistant Professor with the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing. Her research interests include motor function assessment and rehabilitation robot control.
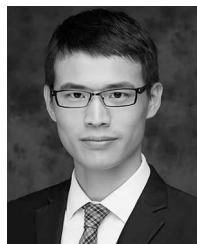
Liang Peng received the B.E. degree in automation from the North University of China, Taiyuan, China, in 2009, the M.S. degree in control theory and control engineering from Beijing Institute of Technology, Beijing, China, in 2012, and the Ph.D. degree in control theory and control engineering from the Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 2016.

He is an Associate Professor with the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences. His current research interests include medical robotics and human–robot interaction.

Zeng-Guang Hou (Fellow, IEEE) received the B.E. and M.E. degrees in electrical engineering from Yanshan University (formerly, North-east Heavy Machinery Institute), Qinhuangdao, China, in 1991 and 1993, respectively, and the Ph.D. degree in electrical engineering from Beijing Institute of Technology, Beijing, China, in 1997.

From May 1997 to June 1999, he was a Postdoctoral Research Fellow with the Key Laboratory of Systems and Control, Institute of Systems Science, Chinese Academy of Sciences, Beijing. He was a Research Assistant with Hong Kong Polytechnic University, Hong Kong, from May 2000 to January 2001. From July 1999 to May 2004, he was an Associate Professor with the Institute of Automation, Chinese Academy of Sciences, and has been a Full Professor since June 2004. From September 2003 to October 2004, he was a Visiting Professor with the Intelligent Systems Research Laboratory, College of Engineering, University of Saskatchewan, Saskatoon, SK, Canada. He is a Professor and the Deputy Director of the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, and the Center for Excellence in Brain Science and Intelligence Technology, Chinese Academy of Sciences, and also with the School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing.

Prof. Hou was an Associate Editor of the *IEEE Computational Intelligence Magazine* and the IEEE TRANSACTIONS ON NEURAL NETWORKS, the Chair of Neural Network Technical Committee of Computational Intelligence Society (CIS), and the Adaptive Dynamic Programming and Reinforcement Learning Technical Committee of CIS. He is an Associate Editor of the IEEE TRANSACTIONS ON CYBERNETICS. He is an Editorial Board Member of *Neural Networks*.

Ying Tan received the M.D. degree in neurology from Peking Union Medical College, Chinese Academy of Medical Sciences, Beijing, China, in 2020.
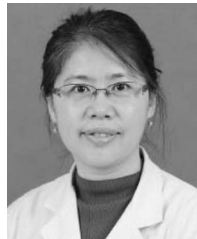
She has been an Attending Physician of the Department of Neurology, Peking Union Medical College Hospital, Beijing, since 2016. Her research interests include neurodegenerative diseases and neuroimmune diseases.

Yanfeng Li received the Ph.D. and M.D. degrees from the Graduate School, Peking Union Medical College (PUMC), Beijing, China, in 1996.

He is currently working with PUMC Hospital, Beijing, as a Neurologist. His research interests include: Alzheimer disease, Parkinson's disease, and other neurodegenerative diseases.

Dr. Li is the President of the Beijing Neurodegenerative Disease Society.

Honglin Hao received the M.D. degree with the Graduate School, Peking Union Medical College, Beijing, China, in 2009.

She is an Associate Chief Physician of the Department of Neurology, Peking Union Medical College Hospital, Beijing. Her research interests include Parkinson's disease, sleep disorders, and Alzheimer's disease.