A Wearable Multi-Segment Upper Limb Tremor Assessment System for Differential Diagnosis of Parkinson's Disease Versus Essential Tremor

Junjie Li, *Student Member, IEEE*, Huaiyu Zhu, *Member, IEEE*, Jiaxiang Li, Haotian Wang, Bo Wang, Wei Luo, and Yun Pan, *Member, IEEE*

***Abstract*— Upper limb tremor is a prominent symptom of both Parkinson's disease and essential tremor. Its kinematic parameters overlap substantially for these two pathological conditions, thus leading to high rate of misdiagnosis, especially for community doctors. Several groups have proposed various methods for improving differential diagnosis. These prior studies have attempted to identify better kinematic parameters, however they have mainly focused on single limb features including tremor intensity, tremor frequency, and tremor variability. In this paper, we propose a wearable system for multi-segment assessment of upper limb tremor and differential diagnosis of Parkinson's disease versus essential tremor. The proposed system collected tremor data from both wrist and fingers simultaneously. From this data, we extracted multi-segment features in the form of phase relationships between limb segments. Using support vector machine classifiers, we then performed differential diagnosis from the extracted features. We evaluated the performance of the proposed system on 19 Parkinson's disease patients and 12 essential tremor patients. Moreover, we also assessed the performance cost associated with reducing task load and sensor array size. The proposed system reached perfect accuracy in leave-one-out cross validation. Task reduction and sensor array reduction were associated with penalties of 2% and 9-10% respectively. The results demonstrated that the proposed system could be simplified for clinical applications, and successfully applied to the differential diagnosis of Parkinson's disease versus essential tremor in real-world setting.**

Manuscript received 31 March 2023; revised 28 June 2023 and 1 August 2023; accepted 9 August 2023. Date of publication 17 August 2023; date of current version 25 August 2023. This work was supported in part by the Zhejiang Provincial Key Research and Development Program under Grant 2021C03027, in part by the Zhejiang Provincial Natural Science Foundation of China under Grant LQ21F010016, and in part by the Key Laboratory of Collaborative Sensing and Autonomous Unmanned Systems of Zhejiang Province. *(Corresponding author: Yun Pan.)*

This work involved human subjects in its research. Approval of all ethical and experimental procedures and protocols was granted by the Ethics Committee of the Second Affiliated Hospital of Zhejiang University School of Medicine under Approval No. 2021-0450.

Junjie Li, Huaiyu Zhu, and Yun Pan are with the College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou 310027, China (e-mail: panyun@zju.edu.cn).

Jiaxiang Li, Haotian Wang, Bo Wang, and Wei Luo are with the Department of Neurology, Second Affiliated Hospital, School of Medicine, Zhejiang University, Hangzhou 310027, China.

Digital Object Identifier 10.1109/TNSRE.2023.3306203

***Index Terms*— Parkinson's Disease, essential tremor, tremor, multi-segment, machine learning.**

I. Introduction

PARKINSON'S disease (PD) and essential tremor (ET) are the two most common tremor syndromes [1]. In general, ET is characterized by postural tremor and kinetic tremor, whereas PD is characterized by resting tremor. However, 18% of ET patients may have resting tremor [2], and 90% of PD patients may develop postural tremor [3]. ET patients are also 3.5 times likely to develop PD than healthy people [4]. Thus, the differential diagnosis of PD versus ET is important in clinical practice, because a correct diagnosis guides adequate treatment and has a significant impart on prognosis [5], [6].

However, the differential diagnosis of PD versus ET could be challenging in clinical settings [7], especially for tremor dominated PD patients at early stage and patients without a family history [8]. Moreover, ET patients are at higher risk of developing PD, and must therefore be re-examined regularly. In clinical practice, differential diagnosis of PD versus ET relies on electrophysiological tests and molecular imaging [9]. Those methods can be only implemented in well-equipped hospitals, and molecular imaging can be expensive [8]. Thus, a convenient method that can be applied via telemedicine could be of great value in aiding clinical efforts.

Upper limb tremor is a prominent symptom of both PD and ET [9]. As a consequence, several groups have focused on developing noninvasive systems to obtain parameters relating to upper limb tremor for differential diagnosis. Surface electromyogram sensors, optical systems, and inertial sensors are widely adopted methodologies. In general, Surface electromyogram sensors are used to record the muscle bursting patterns, whereas optical systems and inertial sensors are used to monitor specific kinematic parameters of tremor, e.g. frequency and mean harmonic peak power.

The above kinematic parameters overlap substantially between PD and ET. Thus, some researchers have attempted to identify kinematic parameters that support better differentiation between PD and ET. For example, di Biase [10] proposed the tremor stability index, and Surangstrirat [11] proposed temporal fluctuation features. They both focused

on features associated with tremor variability within a single segment limb. To the best of our knowledge, no prior research has examined the potential of multi-segment features for differential diagnosis between PD and ET. We speculated that multi-segment tremor patterns, e.g. phase relationship, might differ between PD and ET patients as a consequence of their different pathogenesis.

Based on the above hypothesis, we propose a wearable system for multi-segment assessment of upper limb tremor and differential diagnosis of PD versus ET. We designed a wearable multi-sensor measurement system to collect tremor data from both wrist and fingers simultaneously. We then extracted single limb segment group features and multi-segment group features, and subjected them to the elastic net [12] and the sequential floating forward selection method (SFFS) [13] to reduce their dimensionality. Finally, we used a support vector machine (SVM) classifier to differentiate between PD and ET. To evaluate the proposed system, we enrolled 19 PD patients and 12 ET patients and included the tremor data of both left and right hands. Thus, We collected 38 PD tremor recordings and 24 ET tremor recordings for evaluation. We used leave-one-out cross validation (LOOCV) to obtain the performance, and achieved perfect accuracy (100.00%). Moreover, we investigated strategies for simplifying the proposed system. When adopting two tasks, the LOOCV accuracy decreased by approximately 2%. When relying only on single-segment information, the LOOCV accuracy decreased by 9-10%.

A. Related Work

Several groups have adopted inertial sensors or optical systems to perform differential diagnosis of PD versus ET. Muthuraman [14], WILE [5], Ghassemi [15], Barrantes [16], di Biase [10], Bove [17], Loaiza [18], Duque [8], Locatelli [19], Shahtalebi [20], and Su [21] collected tremor kinematic parameters from the back of the hand or the wrist using inertial sensors, smart watches, or smart phones to distinguish between PD and ET. The back of the hand and the wrist could be good choices for collecting tremor data, because sensors could be easily fixed on them. However, neurologists typically evaluate tremor by investing fingers tremor [22], [23].

Surangsrirat [11], Morrison [24], and Zhang [25] used inertial sensors to collect the acceleration of the forefinger or the middle finger. Oktay [6] used the Leap Motion Controller to collect the 3D positions of wrist, palm, and metacarpal bones for each finger, whereas Kovalenko [26] only extracted the 3D position of the wrist using a video camera. Except Oktay [6], most researchers have collected tremor of a single limb segment, even though there is no research before about which limb segment is best for differential diagnosis. Oktay [6] collected kinematic data from both wrist and fingers, however they did not investigate features related to multi-segment tremor patterns.

With a few exceptions [6], [20], most studies have adopted feature-based classifier methods for differential diagnosis of PD versus ET, because of the limited sample size available. Many research groups have therefore focused on identifying better kinematic parameters to distinguish between PD and ET, e.g. mean harmonic peak power [14], temporal fluctuation [11], regularity [24], and the tremor stability index (TSI) [10]. All these features reflect tremor variability for a single limb segment. To the best of our knowledge, there is no prior research on multi-segment tremor patterns.

B. Contributions and Article Structure

The main contributions of this study are as follow:

- We propose a wearable system for multi-segment assessment of upper limb tremor and differential diagnosis of PD versus ET, with promising results;
- We designed a wearable multi-sensor measurement system to collect the tremor data from wrist and fingers simultaneously for multi-segment tremor analysis;
- We investigated the performance of the proposed system when simplifications were applied, given the significant time pressure experienced in clinical practice especially for screening. We found a smaller impact of task reduction compared with sensor array reduction, indicating that task reduction may be a sensible choice in time-sensitive scenarios.

This article is organized as follow. In section II, we describe the proposed wearable system for multi-segment assessment of upper limb tremor and differential diagnosis of PD versus ET in detail. In section III we introduce the validation experiment and results. In section IV, we discuss the comparison with related work, the analysis of multi-segment features, the simplification of the proposed system, and study limitation.

II. Wearable System for Multi-Segment Assessment of Upper Limb Tremor

A. Wearable Multi-Sensor Measurement System

To collect tremor data, we designed a wearable multi-sensor measurement system. The hardware associated with this system contained five nine-degrees-freedom sensor units, one transmission node, one receiving node, and one personal computer, as shown in Fig. 1. The sensor unit is based on the MPU9250 chip, it measures 22 × 18×6 mm, and it weights 1.8 g, which is light enough to allow natural movement of the fingers [27]. The transmission node is based on a MPU9250 chip and a nRF52832 chip, it measures 39*41*11 mm, and it weighs 38.2 g. The transmission node collects all tremor data simultaneously at a sample rate of 100 Hz and transmits the data via Bluetooth. The receiving node (nRF-52-DK) worked as a Bluetooth serial port transparent transmission module. A personal computer was used to analyze the tremor data. As shown in Fig. 1, the sensor units were attached to the distal interphalangeal joints of the fingers, with Z axes perpendicular to the fingernails. The transmission node was attached to the wrist, with Z axis upward.

To estimate sensor orientation, we adopted the Mahony's complementary filter [28] to derive the direction cosine matrix $\boldsymbol{R}_{\mathbf{S}}^{\mathbf{E}}$. As suggested in [29], we decomposed the acceleration data to remove gravity acceleration as specified by:

$$\boldsymbol{a}_L^{\mathbf{S}} = \boldsymbol{a}^{\mathbf{S}} - \boldsymbol{a}_g^{\mathbf{E}} \boldsymbol{R}_{\mathbf{S}}^{\mathbf{E}}. \tag{1}$$

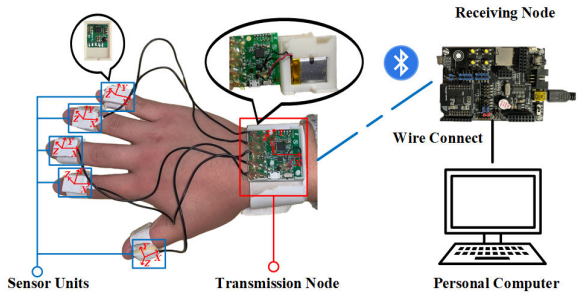


Fig. 1. Hardware overview of the wearable multi-sensor measurement system.

In the above expression, **E** and **S** represent the local tangent plane coordinate and the local coordination system of the sensor respectively; $a_g^{\mathbf{E}}$ indicates the gravitational acceleration with respect to **E**; $\boldsymbol{a}^{\mathbf{S}}$ is the output acceleration of the sensor; $\boldsymbol{a}_L^{\mathbf{S}}$ is linear acceleration with respect to **S**.

To improve interpretability, we converted direction cosine matrix to Euler angles in order of Z-Y-X. We therefore derived linear accelerations and Euler angles.

B. Preprocessing

First, we applied a five-level wavelet based on the db 4 wavelet, to remove signal trend in all components of linear acceleration and Euler angles. Inspired by [10], we then adopted principal component analysis to get the 1^{st} principal component ($\boldsymbol{V}_1$) by (2). Moreover, we chose the the coefficient corresponding to the Z-axis weight of the 1^{st} principal component (r_{31}) to represent tremor axis features, because the Z-axis is consistent between wrist and fingers.

$$[\,\boldsymbol{V}_1 \quad \boldsymbol{V}_2 \quad \boldsymbol{V}_3\,] = [\,\boldsymbol{V}_X \quad \boldsymbol{V}_Y \quad \boldsymbol{V}_Z\,]\boldsymbol{R} \tag{2}$$

$$\boldsymbol{R} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \tag{3}$$

In the above expression, $\boldsymbol{V}$ refers to acceleration or angle; $\boldsymbol{V}_1$, $\boldsymbol{V}_2$, and $\boldsymbol{V}_3$ represent the 1^{st}, 2^{nd}, and 3^{rd} principal component of $\boldsymbol{V}$ respectively; $\boldsymbol{V}_X$, $\boldsymbol{V}_Y$, and $\boldsymbol{V}_Z$ are the X, Y, and Z components of $\boldsymbol{V}$ respectively; $\boldsymbol{R}$ is the principal component coefficients matrix derived by PCA; r_{31} is the coefficient corresponding to the Z-axis weight of $\boldsymbol{V}_1$.

To remove noise, we identified the peak frequency (denoted f_0) of $\boldsymbol{V}_1$ between 3-12 Hz by Welch's power spectral density estimation, and filtered $\boldsymbol{V}_1$ using a band-pass critically damped digital filter [30] ranging between f_0-2 Hz and f_0+2 Hz [10]. In the following analysis, we only used the filtered signal ($\boldsymbol{V}_f$) and r_{31}.

C. Data Analysis

We adopted power spectral density, cross spectrum, and temporal analysis to analyze $\boldsymbol{V}_1$.

In power spectral density analysis, we used Welch's periodograms to derive power spectral density by averaging the power spectra of each 50% overlapping 1 s segment. Following published methods by [31], we extracted peak power frequency (*F0*), peak power (*K*), median power frequency (*F50*), and frequency dispersion (*SF50*). *F50* is the frequency at which power spectral density is halved, and *SF50* is half of frequency band containing 68% of total signal power [32], i.e. frequency band between *F50*−*SF50* and *F50*+*SF50* contains 68% of total signal power.

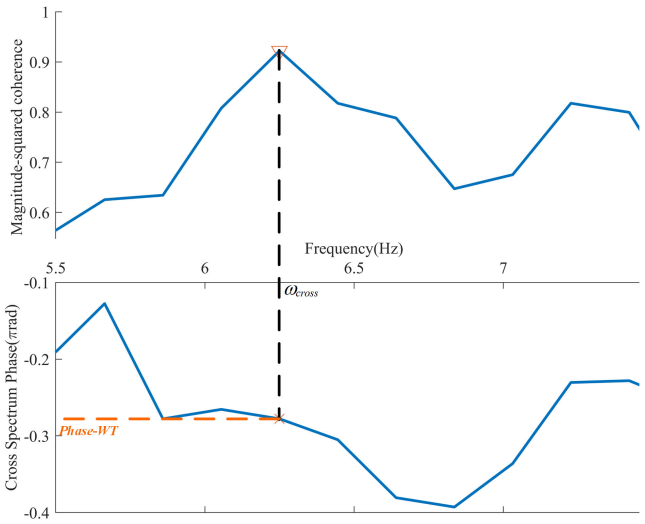


Fig. 2. An example of cross spectrum analysis.

As for cross spectrum analysis, we aimed to analyze the multi-segment pattern of hand tremor. Fig. 2 shows an example for the acceleration pattern between the wrist and the thumb. First, we derived the magnitude-squared coherence of wrist and thumb acceleration:

$$\boldsymbol{C}_{WT}(\omega) = \frac{|\boldsymbol{P}_{WT}(\omega)|^2}{\boldsymbol{P}_{WW}(\omega)\boldsymbol{P}_{TT}(\omega)}. \tag{4}$$

In the above expression, $\boldsymbol{P}_{WW}$ and $\boldsymbol{P}_{TT}$ represent the power spectral densities of wrist and thumb acceleration respectively; $\boldsymbol{P}_{WT}$ refers to the cross power spectral density. We then derived phase difference between wrist and thumb acceleration (*Phase-WT*):

$$\omega_{cross} = argmax(\boldsymbol{C}_{WT}(\omega)) \tag{5}$$

$$Phase\text{-}WT = angle(\boldsymbol{P}_{WT}(\omega_{cross})). \tag{6}$$

In temporal analysis, we could not apply the method proposed by [10] directly, because the sample rate of the proposed system is 100Hz. We therefore interpolated our data with a cubic spline and an upsampling factor 20, to improve frequency resolution. We then subdivided the tremor record into segments delimited by zero-crossings with positive gradient [10], as shown in Fig. 3.

Notably, the interval was constrained to be longer than 0.04 s. Thus, we could got the interval (T_n) and the magnitude (M_n) of the n^{th} tremor segment. We then derived instantaneous frequency (f_n), change in frequency (Δf_n), and change in magnitude (ΔM_n) of the n^{th} tremor segment as specified below:

$$f_n = \frac{1}{T_n} \tag{7}$$

$$\Delta f_n = f_{n+1} - f_n \tag{8}$$

$$\Delta M_n = M_{n+1} - M_n \tag{9}$$

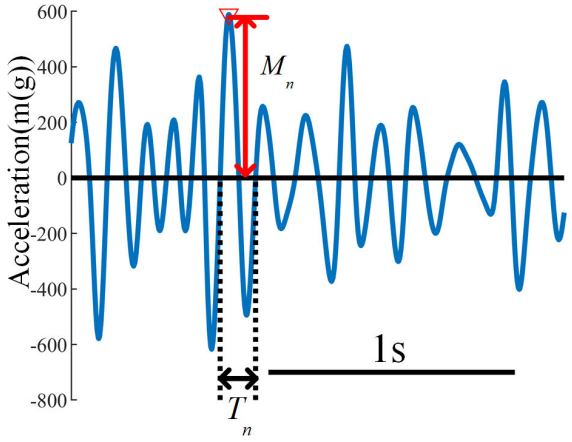


Fig. 3. An example of temporal analysis.

We then could derive frequency average (*FA*), frequency coefficient of variation (*FCV*), and frequency stability index (*FSI*) [10] by:

$$FA = \frac{1}{N}\sum_{i=1}^{N} f_i \tag{10}$$

$$FCV = \frac{\sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(f_i - FA)^2}}{FA} \tag{11}$$

$$FSI = iqr(\Delta f_n). \tag{12}$$

In the above expressions, N was is number of fragments and iqr indicates the interquartile range. We also derived magnitude average (*MA*), magnitude maximum (*MM*), magnitude coefficient of variation (*MCV*), and magnitude stability index (*MSI*):

$$MA = \frac{1}{N}\sum_{i=1}^{N} M_i \tag{13}$$

$$MM = \max_i M_i \tag{14}$$

$$MCV = \frac{\sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(M_i - MA)^2}}{MA} \tag{15}$$

$$\mathrm{MSI} = \frac{iqr(\Delta M_n)}{\sqrt{\sum_{i=1}^{N}\frac{\Delta M_i^2}{N}}} \tag{16}$$

Notably, we propose *MSI* to assess tremor magnitude variability, inspired by [10]. To remove the influence of tremor intensity, we normalized MSI by root mean square of magnitudes.

D. Feature Extraction

In this paper, we divided the extracted features into two groups: multi-segment group and single limb segment group. All extracted features were listed in Table I.

We extracted multi-segment group features by cross spectrum analysis, to capture the multi-segment tremor pattern. We extracted $C_6^2 = 15$ features, e.g *Phase-WT*, since there were six limb segments involved. We denoted those features as *Phase-IJ*: the phase difference between limb segment I and limb segment J. I and J could be W, T, F, M, R, and L, when referring to the wrist, the thumb, the forefinger, the middle finger, the ring finger, and the little finger, respectively. For example, *Phase-FR* referred to the phase difference between forefinger and ring finger.

TABLE I
EXTRACTED FEATURES

Group	Subgroup	Feature	Symbol
Multi-Segment	Phase	phase difference between segment $I$and segment J	*Phase-IJ* $I,J\in\{W,T,F,M,R,L\}$
Single Limb Segment	Intensity	root mean square	*RMS*
		peak power	*K*
		magnitude average	*MA*
		magnitude maximum	*MM*
	Frequency	peak power frequency	*F0*
		median power frequency	*F50*
		frequency dispersion	*SF50*
		frequency average	*FA*
	Variability	magnitude coefficient of variation	*MCV*
		magnitude stability index	*MSI*
		frequency coefficient of variation	*FCV*
		frequency stability index	*FSI*
	Axis Pattern	Z-axis weight	*ZW*

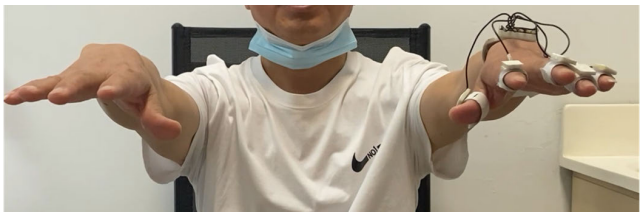
We extracted single limb segment group features from the acceleration or angle of a single limb segment, i.e. one finger or the wrist. Among these features, root mean square (*RMS*), peak power (*K*), magnitude average (*MA*), and magnitude maximum (*MM*) were adopted to assess tremor intensity. Peak power frequency (*F0*), median power frequency (*F50*), frequency average (*FA*), and frequency dispersion (*SF50*) were used to assess tremor frequency. Magnitude coefficient of variation (*MCV*), magnitude stability index (*MSI*), frequency coefficient of variation (*FCV*), and frequency stability index (*FSI*) [10] were used to assess tremor variability. Additionally, we propose Z-axis weight (*ZW*) to assess the axis pattern of tremor, considering pill rolling tremor in PD patients. We extracted 13 features from acceleration or angle for each limb segment. Totally, we extracted 78 features from acceleration or angle for the wrist and the five fingers.

E. Feature Selection and Classification

To remove scale effects, we firstly normalized features using Z scores. We then applied a regularization method to reduce the number of candidate features. More specifically, we used elastic net regularization method proposed in [12] which performs well when the number of features is larger than the sample size. When using this method, we adopted a penalty to the regression coefficients as specified below:

$$Penalty(\alpha,\lambda) = \lambda\sum_{i=1}^{p}\left(\frac{1-\alpha}{2}B_i^2 + \alpha\,|B_i|\right). \tag{17}$$

(a) An example of *P1* task



(b) An example of *P2* task

(c) An example of *P3* task

(d) An example of *P4* task

Fig. 4. Example of tasks.

In the above expression, α and λ are hyper parameters of elastic net regularization method; p is the number of features; B_i is the coefficient of the i^{th} feature. We adopted grid search to determine values for α and λ.

We adopted the wrapper method to select features, because it might produce better results without simplifying assumptions regarding feature independence. Specifically, we used SFFS, because it is one of the most promising feature selection methods for wrapper [13].

As for classification method, we used radial basis function kernel SVM [33] because the number of participants was relatively small and radial basis function kernel is a widely used choice [34]. Specifically, we adopted the grid searching method to optimize hyper parameters of SVM [33].

We used LOOCV to evaluate the performance of the proposed system. And we adopted accuracy as the evaluation criterion in this article. Accuracy was defined as the percentage of correctly classified recordings in validation groups using LOOCV. We also measured sensitivity, specificity, and F1 Score from recordings in validation groups, where we defined PD samples as positive and ET samples as negative. Moreover, we estimated confidence intervals by assuming that the LOOCV accuracy followed a binomial distribution.

TABLE II
DEMOGRAPHIC AND CLINICAL DATA OF DATASET

Variables	PD	ET	p-value
Sample N	19	12	
Gender No. men/women:	9/10	9/3	0.1581^a
Age years (mean±SD):	65.11 ± 11.42	55.08 ± 18.17	0.3200^c
P1 Score (mean±SD):	0.84 ± 0.97	0.71 ± 0.69	0.5602^b
P2 Score (mean±SD):	0.82 ± 0.73	1.20 ± 1.10	0.2080^c
P3 Score (mean±SD):	0.87 ± 0.81	1.29 ± 0.95	0.0591^c
P4 Score (mean±SD):	0.95 ± 0.87	1.38 ± 0.92	0.0650^b

[a] Fisher exact test;

[b] Two sample t-test;

[c] Kruskal-Wallis test; this test was adopted because assumptions of normality and homogeneity of variance were not satisfied.

III. VALIDATION EXPERIMENT

A. Data Collection Protocol

In the experiment, we adopted the following four tasks:

- *P1*: subjects put their arm on the chair arm with hand outstretched, as suggested in [22];
- *P2*: subjects outstretch their arm forward with palm downward, as suggested in [22];
- *P3*: subjects fold their arm in front of their chest with palm downward, i.e. wing beating task suggested in [35];
- *P4*: subjects fold their arm in front of their chest with palm inward, as suggested by clinicians.

Fig.4 showed examples for these tasks. The *P1* task was adopted to assess resting tremor, whereas *P2*, *P3*, and *P4* tasks were adopted to assess postural tremor.

We enrolled 19 PD patients and 12 ET patients to evaluate the proposed system, which was approved by the Ethics Committee of the Second Affiliated Hospital of Zhejiang University School of Medicine (2021-0450). The inclusion criteria were: (1) firm diagnosis of tremor dominated PD or ET following established diagnostic criteria [9], [36]; (2) patients with tremor symptoms. The exclusion criteria were: (1) inability to perform the assigned tasks; (2) cognitive dysfunction.

During the experiments, participants were asked to perform each task for 15 seconds. Meanwhile, an independent neurologist recorded videos using a smartphone, and evaluated tremor severity on a five-point scale from 0 to 4 according to MDS-UPDRS [22].

In total, we collected 38 PD and 24 ET inertial sensor recordings for each task from both hands. Table II summarized demographic and clinical data from all participants. Tremor intensity did not differ significantly between PD and ET recordings in our dataset, although the scores associated with *P3* and *P4* presented a larger difference between PD and ET compared with *P1* and *P2*. Notably, 2 ET patients and 1 PD patient presented similar symptoms, and were diagnosed based on the dopamine transporter imaging.

B. Analysis of Extracted Features

We adopted significance tests to analyze the performance of extracted features at group level. Specifically, we conducted

(a) Results of Features Extracted from Acceleration in *P1*

(b) Results of Features Extracted from Angle in *P1*

(c) Results of Features Extracted from Acceleration in *P2*

(d) Results of Features Extracted from Angle in *P2*

(e) Results of Features Extracted from Acceleration in *P3*

(f) Results of Features Extracted from Angle in *P3*

(g) Results of Features Extracted from Acceleration in *P4*

(h) Results of Features Extracted from Angle in *P4*

Fig. 5. Significance test results of single segment limb group features (Blue Means $p > 0.05$; yellow means $p < 0.05$; red means $p < 0.01$.)

the one-factor analysis of variance when the normality and homogeneity of variance assumptions were satisfied. Otherwise, we conducted the Kruskal-Wallis test. Fig.5 showed significant results of single limb segment group features. We only listed features that showed significant differences in at least one limb segment between PD and ET patients. We found 46 significant ($p < 0.05$) single limb segment features, extracted from acceleration, and 42 significant features extracted from angle. Most significant features came from the *P1* task. The *P3* task was the largest contributor among the postural tremor tasks. Features from the forefinger contributed the most. Notably, features related to tremor intensity, namely *MA* and *MM*, showed significant differences between PD and ET only in the *P3* and *P4* tasks, consistent with the data in the Table II. The frequency stability index, i.e. a previously proposed tremor stability index [10], did not show significant differences between PD an ET patients in our dataset. This result may reflect the relatively low sample frequency or the symptom similarity between PD and ET patients in our dataset.

When assessing multi-segment group features, we found that *Phase-FR* and *Phase-FL* extracted from angle data in the *P1* task showed significant differences ($p < 0.05$). Furthermore, *Phase-WF*, *Phase-WR*, and *Phase-TM* extracted from acceleration data in the *P2* task showed significant differences. For *P3* and *P4*, multi-segment group features did not show significant difference.

C. Comparison Between Acceleration and Angle

To investigate which signal is better for differential diagnosis, we tested the performance of acceleration and angle in each task and all tasks. As shown in Table III, the classification performance was superior for models based on angle compared with those based on acceleration, as a consequence of the greater significance ($p < 0.01$) associated with features extracted from angle data. In view of tasks, *P2* and *P4* performed the best among those tasks.

Moreover, we tested the influence of random labels (Table IV), because the total number of features before feature selection was relatively larger than the sample size adopted in our study. As shown in the Table IV, the accuracy of the proposed model decreased significantly when data were randomly labeled. These results demonstrated that the proposed model did learn from natural data and delivered the promising results (III), which meant generalization of the proposed system was possible.

TABLE III
COMPARISON RESULTS BETWEEN ACCELERATION AND ANGLE

Task	Signal	Accuracy	Sensitivity	Specificity	F1-Score
P1	acc	0.8710	**0.9474**	0.7500	0.9000
	ang	**0.8871**	**0.9474**	**0.7917**	**0.9114**
P2	acc	0.9194	0.9474	**0.8750**	0.9351
	ang	**0.9355**	**0.9737**	**0.8750**	**0.9487**
P3	acc	0.7097	0.7368	**0.6667**	0.7568
	ang	**0.8065**	**0.9737**	0.5417	**0.8605**
P4	acc	0.7258	**0.9737**	0.3333	0.8132
	ang	**0.9355**	**0.9737**	**0.8750**	**0.9487**
All	acc	0.9355	**1.0000**	0.8333	0.9500
	ang	**1.0000**	**1.0000**	**1.0000**	**1.0000**

Acceleration and angle were denoted as "acc" and "ang" respectively;
Bold is the higher value between acceleration and angle;
There were 93 extracted features for each task and 372 extracted features for all tasks.

TABLE IV
RESULTS OF RANDOM LABEL

Task	Signal	Accuracy	Sensitivity	Specificity	F1-Score
All	acc	0.6290	0.5625	0.7000	0.6102
	ang	0.6613	0.6563	0.6777	0.6667

In random Labelling, half PD patients were labeled as ET, whereas half ET patients were labeled as PD;
Acceleration and angle are denoted as "acc" and "ang" respectively.

Considering the superior performance produced by angle data, we only investigated classification model based on angle information in the following analysis.

D. Comparison Between Task Combinations

In clinical practice, both resting tremor and postural tremor are important for differential diagnosis of PD versus ET [36]. For this reason, it appears reasonable to combine data from the resting tremor tasks with data from the postural tremor task. To identify the optimal combination, we tested the performance of our model for each combination. Fig. 6 shows The SFFS results for each combination. When adding 8 features, all combinations achieved optimal performance. We therefore presented detailed performance results from each combination in the Table V, when selecting 8 features. The *P2* and *P4* tasks barely improved when combined with the *P1* task, whereas the *P3* task improved substantially when combined with the *P1* task. Moreover, the proposed multi-segment feature, i.e. Phase-*FL-P1*, was positively selected for all task combinations.

E. Comparison Between Limb Segments

To investigate which limb segment is better for the proposed system, we tested the performance of each limb segment for the *P1&P3* task combination and for all tasks. The results (Table VI) demonstrate that the forefinger or the little finger may represent a better choice for the proposed system.

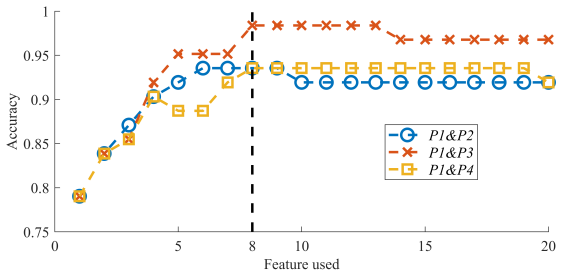


Fig. 6. SFFS results for each task combination.

TABLE V
FEATURES SELECTED FOR DIFFERENT COMBINATIONS AND ASSOCIATED RESULTS

Combination	*P1&P2*	*P1&P3*	*P1&P4*
1^{st} feature	*FA-F-P1*	*FA-F-P1*	*Phase-FL-P1*
2^{nd} feature	*Phase-FL-P1*	*Phase-FL-P1*	*FCV-F-P4*
3^{rd} feature	*RMS-T-P2*	*F0-L-P1*	*F50-L-P1*
4^{th} feature	*F0-L-P1*	*F50-F-P3*	*MM-R-P4*
5^{th} feature	*MCV-W-P2*	*F50-L-P1*	*F50-L-P4*
6^{th} feature	*MCV-L-P2*	*RMS-T-P3*	*SF50-R-P1*
7^{th} feature	*MA-W-P2*	*SF50-T-P3*	*FCV-R-P4*
8^{th} feature	*MA-M-P2*	*FA-L-P1*	*F50-F-P1*
Accuracy	0.9355	**0.9839**	0.9355
Sensitivity	0.9737	**1.0000**	**1.0000**
Specificity	0.8750	**0.9583**	0.8333
F1-Score	0.9487	**0.9870**	0.9500

Bold is the higher value among task combinations;
Features are denoted as "ζ-η-τ", in which ζ is the feature symbol, η is the segment limb, and τ is the task;
For example, *FA-F-P1* is frequency average of the forefinger in *P1* task.

TABLE VI
COMPARISON BETWEEN LIMB SEGMENTS

Task	Limb	Accuracy	Sensitivity	Specificity	F1-Score
P1&P3	W	0.7903	0.9474	0.5417	0.8471
	T	0.8548	0.8947	**0.7917**	0.8831
	F	**0.8871**	0.9474	**0.7917**	0.9114
	M	0.8226	0.9211	0.6667	0.8642
	R	0.8226	**0.9737**	0.5833	0.8706
	L	**0.8871**	**0.9737**	0.7500	**0.9136**
All	W	0.7097	1.0000	0.2500	0.8085
	T	0.8871	0.8684	**0.9167**	0.9041
	F	**0.9194**	0.9737	0.8333	**0.9367**
	M	0.8871	0.9474	0.7917	0.9114
	R	0.9032	**1.0000**	0.7500	0.9268
	L	**0.9194**	0.9737	0.8333	**0.9367**

Bold is the higher value among limb segments;
W, T, F, M, R, L means wrist, thumb, forefinger, middle finger, and little finger respectively.

IV. DISCUSSION

A. Comparison With Related Works

In this article, we propose a wearable system for the multi-segment assessment ofr upper limb tremor and for differential diagnosis of PD versus ET. Table VII compared our work with related studies. Our system reached an accuracy of 91.94% (95% CI: 82.17 – 97.33%) when analyzing single limb segment, which is comparable to the prior results. Meanwhile, our system reached an accuracy of 100.0% (95% CI:

TABLE VII
COMPARISON WITH RELATED STUDIES

Year	Author	Subjects	Accuracy (95% CI)
2011	Muthuraman [14]	39 PD, 41 ET[a]	94%[c]
2014	WILE [5]	15 PD, 14 ET[a]	96%[c]
2016	Ghassemi [15]	13 PD, 11 ET[a]	83%[c]
2016	Surangsrirat [11]	32 PD, 20 ET[a]	100.00%[c]
2017	Morrison [24]	15 PD, 10 ET[b]	AUC= 0.97[c]
2017	Barrantes [16]	17 PD, 16 ET[a]	84.38%[c]
2017	di Biase [10]	16 PD, 20 ET[a]	92%[c]
2018	Bove [17]	20 PD, 20 ET[a]	95%*[c]
2018	Zhang [25]	26 PD, 24 ET[a]	96%[c]
2019	Loaiza [18]	17 PD, 16 ET[b]	$80-100\%$[c]
2020	Duque [8]	19 PD, 20 ET[b]	$77.8 \pm 9.9\%$[c]
2020	Oktay [6]	23 PD, 17 ET[a]	90%[d]
2020	Locatelli [19]	17 PD, 7 ET[a]	92.1%[c]
2021	Kovalenko [26]	42 PD, 13 ET[a]	$77 \pm 11\%$[c]
2021	Shahtalebi [20]	47 PD, 34 ET[a]	95.5%[c]
2021	Su [21]	48 PD, 48ET[a]	85%[c]
	Ours	19 PD, 12 ET	91.94%[c] (82.17 − 97.33%) 100.00%[d] (94.22 − 100.00%)

[a] not mentioned the tremor score of ET or PD;
[b] not assessed with consistent scales;
[c] analyzed single limb segment;
[d] analyzed more than one limb segment;
* the accuracy of PD versus ET and dystonic tremor.

94.22−100.0%) when analyzing more than one limb segment. In particular, we tested temporal fluctuation features [11] in our dataset and reached an accuracy of 72.58% (95% CI: 57.99 − 83.15%), which might result from the similarity in tremor intensity between PD and ET patients in our dataset.

Most published studies did not mention tremor scores of ET or PD patients. In some studies [8], [18], [24] PD patients were assessed according to the MDS-UPDRS [22], and ET patients were assessed according to the Fahn-Tolosa-Marin scale (FTMRS) [23]. However, scores from MDS-UPDRS are not consistent with those from FTMRS: score of 1 and 2 mean tremor smaller than 3 cm in MDS-UPDRS, and smaller than 1 cm in FTMRS. As a result, the estimated tremor intensity was not consistent between PD and ET patients, althogh the authors only included patients with a score of 1 or 2 on the MDS-UPDRS for PD, and on the FTMRS for ET [8], [18]. In this article, we enrolled PD and ET patients with similar hand tremor intensity according to MDS-UPDRS [22], especially in the *P1* and *P2* tasks suggested by [22]. Our dataset is therefore more representative of the clinical reality, and we obtained promising results with this dataset. To the best of ourknowledge, our study is the first to assess model performance for a dataset in which PD patients and ET patients present comparable hand tremor intensity.

B. Multi-Segment Features

In this study, we propose multi-segment features, i.e. the phase difference between limb segments. We found that the angle phase difference of PD patients was closer to zero in the *P1* task, as shown Fig. 7. This indicates that limb segments engage in synchronous tremor in PD patients.

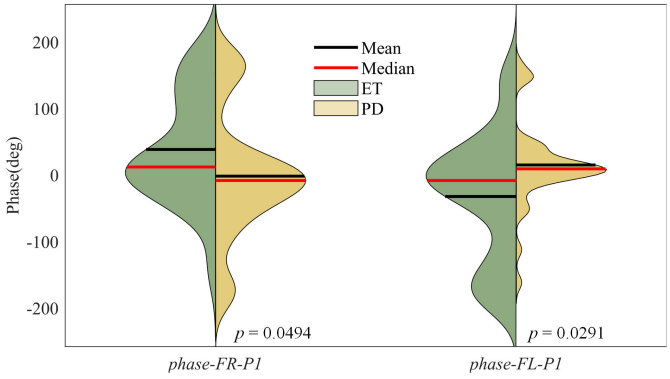


Fig. 7. Violin figures of phase differences between limb segment.

As shown in Table V, the multi-segment feature, *Phase-FL-P1*, was selected in all those combinations. We also investigated the performance of the proposed system without multi-segment group features (Table I). The LOOCV accuracy decreased to 85.48% (95% CI: 74.22−93.14%), 85.48% (95% CI: 74.22 − 93.14%), 77.42% (95% CI: 65.03 − 87.07%), 87.10% (95% CI: 76.15 − 94.26%), and 95.16% (95% CI: 86.50−98.99%) for *P1*, *P2*, *P3*, *P4*, and all tasks respectively. The LOOCV accuracy decreased approximately $3-8\%$ without multi-segment feature group. These results demonstrate the proposed multi-segment features contributed significantly to the promising performance of the proposed system.

C. Simplification of the Proposed System

In clinical practice, time cost is significant especially for screening. Thus, ease of wearing could represent a priority for successful application of the proposed system. During the validation experiment, we kept the Z axes of the sensor units perpendicular to the fingernails. However, except for the Z-axis weight, all features of the proposed system are rotation invariant. As a consequence, the potential impact of different wearing configuration would depend only on the Z-axis weight. Except for *ZW-T-P2*, the Z-axis weight features extracted from angle were not significant(Fig. 5). Furthermore, Z-axis features were not selected, indicating that the proposed system was robust when the Z axes of the sensor units were not exactly perpendicular to the fingernails. These results indicated good compliance with different wearing configurations.

Our results also indicate that time costs may be effectively reduced by reducing tasks load and sensor array size. We investigated the performance penalty incurred when task reduction and sensor array reduction were applied. We estimated the time cost associated with the original configuration of six sensors & four tasks from the video recordings, and estimated the time cost associated with each simplified configuration. For example, we assumed that one sensor cost $1/6$ of the time associated with attaching sensors time in the original configuration, and two tasks cost $2/4$ of the time associated with performing tasks in the original configuration. The results were presented in Fig. 8.

We investigated performance for each single task separately (Table III), and for their combination (Table V). The *P2* and *P4* tasks, showed the best performance in single task settings,

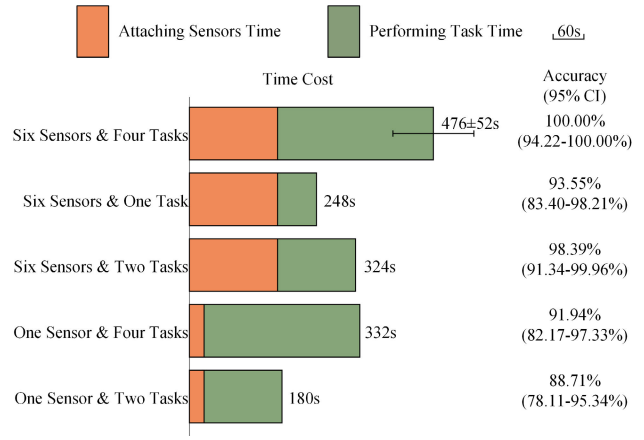


Fig. 8. Results of simplified system configuration.

with an accuracy of 93.55% (95% CI: 83.40 – 98.21%). Moreover, the *P3* task (wing beating task) showed the best performance when combined with resting tremor. As shown in the Fig. 6 and the Table V, the performance of combination *P1&P3* was superior to other combinations after including *F50-F-P3*, which indicating less correlation between features from the *P3* and *P1* tasks compared with other combinations. Overall, we achieved an accuracy of 93.55% (95% CI: 83.40–98.21%) when adopting only one task and an accuracy of 98.39% (95% CI: 91.34 – 99.96%) when combining *P1&P3*. Our results indicated that the combination of the *P1&P3* tasks could be adopted in time-sensitive scenarios.

During the experiments, attaching sensors took about 36% time. Thus, time cost and device cost could both be reduced if sensor array reduction was applied. We investigated performance of each single limb segment (Table VI) to estimate performance penalty when using only one single sensor. The little finger produced the best performance with an accuracy of 91.94% (95% CI: 82.17 – 97.33%) for all tasks, and 88.71% (95% CI: 78.11 – 95.34%) for the *P1&P3* combination. The performance of the forefinger was very close to that achieved by the little finger. These results indicated that a single inertial sensor attached to the forefinger or little finger could be adopted in performance-insensitive scenarios.

When comparing penalty in performance from task reduction and sensor array reduction, we found that the influence of sensor array reduction was larger than of that associated with task reduction. We speculated that the lower performance penalty (approximately 2%) associated with task reduction might reflect the fact that the *P2*, *P3*, and *P4* tasks are all postural tremor examination tasks, and are therefore highly correlated with each other. The higher performance penalty (9 – 10%) associated with sensor array reduction might reflect the fact that multi-segment features could not be derived from any single limb segment, and that the features extracted from each single limb segment were less correlated with each other.

D. Study Limitation

The results of this study are promising, however further researches are needed. The sample size involved in this study is relatively small, because we only involved tremor dominated PD patients and ET patients. Thus, whether the results could be generalized needed to be further investigated. Moreover, the pathological mechanism of the different multi-segment tremor pattern between PD and ET needed to be further investigated.

V. Conclusion

In this study, we propose a wearable system for multi-segment assessment of upper limb tremor and for differential diagnosis of PD versus ET. To evaluate the proposed system, we enrolled 19 PD patients and 12 ET patients. The results demonstrated that the proposed system successfully performed differential diagnosis of PD versus ET. Furthermore, we found that task reduction (adopting *P1&P3*) would lead to an accuracy reduction of 1.61%, and could therefore be adopted in time-sensitive scenarios. Finally, we found that multi-segment features had a significant impact on performance, and that a reduction in sensor array size would lead to a performance penalty of 9 – 10%. These findings indicate that the proposed system could be simplified for successful application in real-world clinical settings.

Acknowledgment

The authors would like to thank all participants who took part in the experiments.

References

[1] P.-C. Lin, K.-H. Chen, B.-S. Yang, and Y.-J. Chen, "A digital assessment system for evaluating kinetic tremor in essential tremor and Parkinson's disease," *BMC Neurol.*, vol. 18, no. 1, p. 25, Dec. 2018.

[2] O. A. Cohen, "Rest tremor in patients with essential tremor: Prevalence, clinical correlates, and electrophysiologic characteristics," *Arch. Neurol.*, vol. 60, no. 3, pp. 405–410, 2003.

[3] E. D. A. Louis, "Clinical correlates of action tremor in Parkinson disease," *Arch. Neurol.*, vol. 58, no. 10, pp. 1630–1634, 2001.

[4] J. Benito-Leon, E. D. Louis, and F. Bermejo-Pareja, "Risk of incident Parkinson's disease and parkinsonism in essential tremor: A population based study," *J. Neurol., Neurosurgery Psychiatry*, vol. 80, no. 4, pp. 423–425, Dec. 2008.

[5] D. J. Wile, R. Ranawaya, and Z. H. T. Kiss, "Smart watch accelerometry for analysis and diagnosis of tremor," *J. Neurosci. Methods*, vol. 230, pp. 1–4, Jun. 2014.

[6] A. B. Oktay and A. Kocer, "Differential diagnosis of Parkinson and essential tremor with convolutional LSTM networks," *Biomed. Signal Process. Control*, vol. 56, Feb. 2020, Art. no. 101683.

[7] Y. S. A. Oh, "Color vision in parkinson's disease and essential tremor," *Eur. J. Neurol.*, vol. 18, no. 4, pp. 577–583, 2011.

[8] J. D. L. Duque, A. J. S. Egea, T. Reeb, H. A. G. Rojas, and A. M. González-Vargas, "Angular velocity analysis boosted by machine learning for helping in the differential diagnosis of Parkinson's disease and essential tremor," *IEEE Access*, vol. 8, pp. 88866–88875, 2020.

[9] K. P. Bhatia 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, Jan. 2018.

[10] L. D. Biase et al., "Tremor stability index: A new tool for differential diagnosis in tremor syndromes," *Brain*, vol. 140, no. 7, pp. 1977–1986, Jul. 2017.

[11] D. Surangsrirat, C. Thanawattano, R. Pongthornseri, S. Dumnin, C. Anan, and R. Bhidayasiri, "Support vector machine classification of Parkinson's disease and essential tremor subjects based on temporal fluctuation," in *Proc. 38th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug. 2016, pp. 6389–6392.

[12] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *J. Roy. Stat. Soc. Ser. B, Stat. Methodology*, vol. 67, no. 2, pp. 301–320, Apr. 2005.

[13] S. Kuchibhotla, H. D. Vankayalapati, and K. R. Anne, "An optimal two stage feature selection for speech emotion recognition using acoustic features," *Int. J. Speech Technol.*, vol. 19, no. 4, pp. 657–667, Dec. 2016.

[14] M. Muthuraman, A. Hossen, U. Heute, G. Deuschl, and J. Raethjen, "A new diagnostic test to distinguish tremulous Parkinson's disease from advanced essential tremor," *Movement Disorders*, vol. 26, no. 8, pp. 1548–1552, Jul. 2011.

[15] N. H. Ghassemi et al., "Combined accelerometer and EMG analysis to differentiate essential tremor from Parkinson's disease," in *Proc. 38th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug. 2016, pp. 672–675.

[16] S. Barrantes et al., "Differential diagnosis between Parkinson's disease and essential tremor using the smartphone's accelerometer," *PLoS ONE*, vol. 12, no. 8, Aug. 2017, Art. no. e0183843.

[17] F. A. Bove, "A role for accelerometry in the differential diagnosis of tremor syndromes," *Funct. Neurol.*, vol. 33, no. 1, pp. 45–49, 2018.

[18] J. D. L. Duque, A. M. González-Vargas, A. J. S. Egea, and H. A. G. Rojas, "Using machine learning and accelerometry data for differential diagnosis of Parkinson's disease and essential tremor," in *Applied Computer Sciences in Engineering*, J. C. Figueroa-García, Eds. Cham, Switzerland: Springer, 2019, pp. 368–378.

[19] P. A. Locatelli, "Classification of essential tremor and Parkinson's tremor based on a low-power wearable device," *Electronics*, vol. 9, no. 10, p. 1695, 2020.

[20] S. Shahtalebi, S. F. Atashzar, R. V. Patel, M. S. Jog, and A. Mohammadi, "A deep explainable artificial intelligent framework for neurological disorders discrimination," *Sci. Rep.*, vol. 11, no. 1, p. 9630, May 2021.

[21] D. Su et al., "Different effects of essential tremor and parkinsonian tremor on multiscale dynamics of hand tremor," *Clin. Neurophysiol.*, vol. 132, no. 9, pp. 2282–2289, Sep. 2021.

[22] C. G. Goetz et al., "Movement disorder society-sponsored revision of the unified Parkinson's disease rating scale (MDS-UPDRS): Scale presentation and clinimetric testing results," *Movement Disorders*, vol. 23, no. 15, pp. 2129–2170, Nov. 2008.

[23] W. Ondo et al., "Comparison of the Fahn-Tolosa-Marin clinical rating scale and the essential tremor rating assessment scale," *Movement Disorders Clin. Pract.*, vol. 5, no. 1, pp. 60–65, Jan. 2018.

[24] S. Morrison, K. M. Newell, and J. J. Kavanagh, "Differences in postural tremor dynamics with age and neurological disease," *Exp. Brain Res.*, vol. 235, no. 6, pp. 1719–1729, Jun. 2017.

[25] B. Zhang, F. Huang, J. Liu, and D. Zhang, "A novel posture for better differentiation between Parkinson's tremor and essential tremor," *Frontiers Neurosci.*, vol. 12, p. 317, May 2018.

[26] E. Kovalenko et al., "Distinguishing between Parkinson's disease and essential tremor through video analytics using machine learning: A pilot study," *IEEE Sensors J.*, vol. 21, no. 10, pp. 11916–11925, May 2021.

[27] J. A. Li, "Three-dimensional pattern features in finger tapping test for patients with parkinson's disease," in *Proc. 42nd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2020, pp. 3676–3679.

[28] S. A. Ludwig and K. D. Burnham, "Comparison of Euler estimate using extended Kalman filter, Madgwick and Mahony on quadcopter flight data," in *Proc. Int. Conf. Unmanned Aircr. Syst. (ICUAS)*, Jun. 2018, pp. 1236–1241.

[29] G. Cai et al., "Quantitative assessment of parkinsonian tremor based on a linear acceleration extraction algorithm," *Biomed. Signal Process. Control*, vol. 42, pp. 53–62, Apr. 2018.

[30] D. G. E. Robertson and J. J. Dowling, "Design and responses of Butterworth and critically damped digital filters," *J. Electromyogr. Kinesiol.*, vol. 13, no. 6, pp. 569–573, Dec. 2003.

[31] O. Bazgir, S. A. H. Habibi, L. Palma, P. Pierleoni, and S. Nafees, "A classification system for assessment and home monitoring of tremor in patients with Parkinson's disease," *J. Med. Signals Sensors*, vol. 8, no. 2, pp. 65–72, 2018.

[32] P. Pierleoni, L. Palma, A. Belli, and L. Pernini, "A real-time system to aid clinical classification and quantification of tremor in Parkinson's disease," in *Proc. IEEE-EMBS Int. Conf. Biomed. Health Informat. (BHI)*, Jun. 2014, pp. 113–116.

[33] C. Chang and C. Lin, "LIBSVM," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 1–27, 2011.

[34] C. A. Hsu, "A practical guide to support vector classication," *Bioinformatics*, vol. 1, pp. 1–16, Apr. 2003.

[35] A. Lenka and J. Jankovic, "Tremor syndromes: An updated review," *Frontiers Neurol.*, vol. 12, Jul. 2021, Art. no. 684835.

[36] R. B. Postuma et al., "MDS clinical diagnostic criteria for Parkinson's disease," *Movement Disorders*, vol. 30, no. 12, pp. 1591–1601, Oct. 2015.