Computer Methods and Programs in Biomedicine 219 (2022) 106741

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

journal homepage: www.elsevier.com/locate/cmpb

Objective quantification of the severity of postural tremor based on kinematic parameters: A multi-sensory fusion study

Chenbin Ma[a,b,c,d], Peng Zhang[b,c], Jiachen Wang[e], Jian Zhang[e], Longsheng Pan[f], Xuemei Li[g], Chunyu Yin[g], Ailing Li[h], Rui Zong[f,*], Zhengbo Zhang[a,*]

[a] Center for Artificial Intelligence in Medicine, Medical Innovation Research Department, PLA General Hospital, 100853, Beijing, China
[b] Beijing Advanced Innovation Center for Biomedical Engineering, Beihang University, 100191, Beijing, China
[c] School of Biological Science and Medical Engineering, Beihang University, 100191, Beijing, China
[d] Shenyuan Honors College, Beihang University, 100191, Beijing, China
[e] Medical School of Chinese PLA, 100853, Beijing, China
[f] Department of Neurosurgery, First Medical Center of Chinese PLA General Hospital, 100853, Beijing, China
[g] Clinics of Cadre, Department of Outpatient, First Medical Center of Chinese PLA General Hospital, 100853, Beijing, China
[h] Pusheng Yixin (Beijing) Hospital Management Co., Ltd, 100020, Beijing, China

ARTICLE INFO

Article history:
Received 27 September 2021
Revised 27 February 2022
Accepted 8 March 2022

Keywords:
Essential tremor
Wearable sensor
Multi-sensory fusion
Rating of Severity
Machine learning

ABSTRACT

Background: Current clinical assessments of essential tremor (ET) are primarily based on expert consultation combined with reviewing patient complaints, physician expertise, and diagnostic experience. Thus, traditional evaluation methods often lead to biased diagnostic results. There is a clinical demand for a method that can objectively quantify the severity of the patient's disease.
Methods: This study aims to develop an artificial intelligence-aided diagnosis method based on multi-sensory fusion wearables. The experiment relies on a rigorous clinical trial paradigm to collect multi-modal fusion of signals from 98 ET patients. At the same time, three clinicians scored independently, and the consensus score obtained was used as the ground truth for the machine learning models.
Results: Sixty kinematic parameters were extracted from the signals recorded by the nine-axis inertial measurement unit (IMU). The results showed that most of the features obtained by IMU could effectively characterize the severity of the tremors. The accuracy of the optimal model for three tasks classifying five severity levels was 97.71%, 97.54%, and 97.72%, respectively.
Conclusions: This paper reports the first attempt to combine multiple feature selection and machine learning algorithms for fine-grained automatic quantification of postural tremor in ET patients. The promising results showed the potential of the proposed approach to quantify the severity of ET objectively.

1. Introduction

Essential tremor (ET) usually presents as postural and intentional tremor involving the upper extremities at a frequency of 4–12 Hz, with a prevalence of 0.4%–6% in adults [1]. The prevalence increases significantly with age, with 13.0–50.5/1000 in people over sixty. The prevalence rate of people over 80 years old is more than 20 times that of people between 20 and 39 years old [1,2]. Laboratory evaluation is the preferred method for diagnosing and rating the severity of ET symptoms. However, the effectiveness of such protocols depends on the clinical experience and expertise of the examining physician and additional review of the patient, and management of such disorders remains suboptimal [3]. In addition, it is challenging to assess the condition in seniors who have poor cognitive abilities or have limited mobility [4]. Neurologists observe patients performing specific tasks with the naked eye and score each task according to the Clinical Rating Scale for Tremor (CRST) criteria [5]. The variability of scores between different physicians or medical centers is high, and the condition of tremor symptoms can progress over time. Therefore, it is essential to adapt to the individual needs of ET patients, and accurate diagnosis of symptom severity is critical for prognosis and treatment [6].

Therefore, there is an urgent need for research to provide physicians with objective, multi-modal, quantitative metrics to enrich decision-aid information. Wearable device-based intelligent diag-

* Corresponding authors.
E-mail addresses: zongrui11@126.com (R. Zong), zhengbozhang@126.com (Z. Zhang).

https://doi.org/10.1016/j.cmpb.2022.106741
0169-2607/

nostic aids offer a powerful alternative to traditional, face-to-face clinical assessment strategies. Inertial measurement units (IMUs), typically integrated with accelerometers (ACC), gyroscopes (GYRO), and magnetometers (MAG), can be used for quantitative analysis of human posture [7–11]. In addition to this, sensors that measure bioelectrical activity (EMG, EEG, etc.) can also be integrated [12,13]. With the help of signal processing and feature extraction, multi-sensory fusion signals can be mapped into interpretable features with biological meaning. These features can characterize the degree of tremor in patients through a supervised machine learning model. Long-term collection of patient data in activities of daily living (ADL) can avoid patient recall bias and provide a basis for clinical assessment of the condition by longitudinally monitoring helpful information to assess the severity of tremors [7,9,14–18]. Many studies are based on smartphones or watches to monitor tremors [10,19–22]. These studies suggest the increasing feasibility of wearable technology for monitoring tremor symptoms in community and clinical applications.

However, most wearable devices are limited by power consumption and data storage, and the sampling rate and signal dimension tend to be insufficient. Even the few available medical-grade wearable systems for continuous monitoring ET tremor severity in ADLs, such as Kinesia HomeView™ [9], require patients to repeat standardized tremor assessment actions every hour using a tablet as prompted. These operations increase the burden of users and interfere with their daily life. Besides, Papadopoulos et al. [22] built a deep model and identified Parkinson's disease (PD) patients on datasets collected via smartphones, with a sensitivity of only 61.2% in the rest tremor task and a precision of only 57.6% in the postural tremor task. The limited recognition accuracy is difficult to apply to clinical scenarios due to the interference of environmental noise. Furthermore, It has been shown that the relative utility of data is highly dependent on the placement of sensing nodes in the body and the intended research question [23,24]. Hssayeni et al. [7] assessed resting tremors in PD patients' hands and found that tremor features in the hands were more sensitive and vital in predicting severity than those at the ankles. Sigcha et al. [25] collected ADL data from 18 PD patients based on a low-power accelerometer watch and graded the milder symptoms. However, this work uses an empirical threshold approach to determine the data labels that may deviate from the actual measurements, which may affect the reliability of the later modeling results. Moreover, uncontrolled ADL scenarios are typically mixed with random noise interference [18,21,22], and valuable information is difficult to support the fine-grained resolution required by clinical requirements. For example, Zhang et al. [18] train machine learning algorithms on laboratory data assuming that findings will translate to the wild. It turns out that this approach leads to large prediction errors. Related studies have shown [15,16,18,21,22] that algorithms trained on the ADL dataset applied directly to the canonical laboratory inspection tasks can significantly degrade performance. Although simplifying feature definitions enables lighter monitoring and analysis, it also loses its ability to upgrade. Accurate diagnostic efficiency requires the benefits of more regulated data acquisition, higher sensing resolution, and multi-dimensional signal channels. Therefore, it is better to use multi-sensory fusion in signal acquisition systems for laboratory diagnosis [7,26–30].

Studies have been quantifying the severity of symptoms in PD patients, for example, characterizing tremors by tasks such as gait analysis [29,31–35]. Butt et al. [29] proposed an adaptive neuro-fuzzy inference system that identified kinematic parameters to identify PD, although its correlation coefficient was only 0.814. Recently, Sigcha et al. [25] used deep neural networks, which may have the problem of learning too many parameters and even the risk of overfitting. And the classification results for resting tremors were limited to patients with mild symptoms, with a model sensitivity of only 89% and specificity of 89.1% for triple classification. However, although PD and ET are in the same movement disorder category, the tremor phenotype between the two is different [36]. On the other hand, the condition is both clinically and genetically heterogeneous, it remains clinically tricky to distinguish between the two disorders, and differences in tremor phenotype remain [10,11,26,36–39]. Therefore, quantifying the symptoms of ET cannot simply replicate the findings of PD. Further research is still needed on medical-grade diagnostic aids for an automated scoring method for the objective quantification of ET motor symptoms.

Overall, to our knowledge, few studies have used high-dimensional postural sensing features to assess the utility of quantitative analysis of ET symptoms. More experiments are needed to demonstrate the quantitative utility of inertial sensor-based derived kinematic parameters for ET symptoms. Firstly, the study relied on a rigorous clinical trial paradigm in which ET patients were screened by biochemical indicators, imaging, history taking, and expert diagnosis, and some were enrolled in the "MR-guided focused ultrasound (MRgFUS)" surgery. The ET patients included in this study wore a miniature IMU device mounted on the back of their hands and performed specified movements according to a clinical scale under the guidance of a neurologist. Simultaneously, video data were collected using a digital camera. Multiple experienced clinical experts independently score these videos to obtain consensus scores to minimize subjective bias and facilitate scientifically supervised learning. Then, this study extracts time-domain, frequency-domain, and nonlinear statistical features of the collected high-dimensional pose signals to characterize their tremor levels. In parallel, five machine learning models were developed, and the classification utility of the feature selection and optimization algorithms was evaluated based on comprehensive performance metrics.

2. Methodology

2.1. Subjects and protocol

This study was based on the clinical trial "Efficacy and Safety Study of ExAblate Transcranial MRgFUS Thalamic Disruption for Drug-Refractory Idiopathic Tremor" (trial protocol code ET002J) at PLA General Hospital. It was approved by the Ethics Committee of Chinese PLA General Hospital (S2018–021–00/01). The ET002J clinical trial is part of a prospective, single-arm, multicenter clinical trial bid by InSightec (ClinicalTrials.gov Identifier: NCT03253991). Patients were recruited remotely through questionnaires, telephone interviews, and video consultations. After a specialist consultation in the Department of Neurology at the PLA General Hospital, the initial determination was made. This data collection experiment was started on September 4, 2020. As of April 11, 2021, a total of 98 patients with ET (61.53 ± 9.77 years, 68 males) participated in this study, and they were enrolled voluntarily and signed informed consent. Participants were asked to fill in questionnaires (mini-mental state examination, Beck Depression inventory) before screening their health states. They required screening with biochemical tests, imaging examinations, and clinical scale evaluation; after then, 17 patients with drug-refractory type were finally treated with MRgFUS surgery. Table 1 shows the detailed basic information table of ET patients for this study. All of these patients met the criteria of having taken oral anti-tremor medication and were consistent with drug-refractory presentation (e.g., inadequate response to medication, intolerance due to adverse effects, etc.).

Data collection experiments are conducted in a closed, separate room to ensure that patients are not disturbed by their surroundings, and the laboratory examinations were guided throughout by a neurologist with expertise in movement disorders, and

Table 1
Basic information table of ET patents for this study.

Gender	male (n = 68)	female (n = 30)
Demographic (mean ± SD)		
Age (years)	59.19 ± 9.48	64.75 ± 6.76
Disease duration (years)	5.93 ± 1.54	6.37 ± 0.88
Height (cm)	164.34 ± 5.52	158.63 ± 3.32
Weight (kg)	62.71 ± 7.96	54.93 ± 9.28
Scales (mean ± SD)		
MMSE	29.86 ± 0.48	29.38 ± 0.95
BDI	0.97 ± 0.95	0.71 ± 1.47
Severity Rating (mean ± SD)		
Wing-beat position	2.12 ± 1.44	1.97 ± 1.35
Arm extension	2.04 ± 1.39	2.00 ± 1.96
Finger-finger position	2.85 ± 1.39	2.77 ± 1.34

Notes: MMSE, mini-mental state examination; BDI, Beck depression inventory.

video data (CMOS camera, 48 MP, 1920 × 1080 HD, 60 fps) were recorded to support independent scoring (blinded with each other) by three neurologists. To ensure that the imaging quality was sufficient to support the neurologist's later diagnosis, the absolute position of the camera mount and chair during the experimental examination was fixed, and both of the patient's hands were displayed as large as possible within the camera's screen. To ensure signal quality and minimize individual differences in collected data due to random interference, patients were measured for at least 30 s under each postural task. If unexpected circumstances interrupted the experiment, it was required to recheck. Data that failed to meet the minimum measurement time would be deleted. During laboratory examinations, patients wear miniature posture sensors on the backs of their hands, and digital signals are transmitted wirelessly in a host computer via a Bluetooth connection. The upper computer displays the signal waveform changes in real-time and stores them in text form on the computer's hard disk. To ensure that a valid tremor signal is resolved, the sampling frequency of the device was set to 100 Hz, and the baud rate was set to 115,200 Byte per second.

The diagnostic scoring of the postural tremor was done independently by three neurologists and was performed according to standard diagnostic procedures (patient history, imaging examination, etc.). According to the CRST criteria [5], there are five levels of tremor severity. The specific scoring process based on CRST was as follows: 1) two neurologists scored the severity of the patient's postural tremor by viewing the video data and obtained two score lists blinded with each other. 2) a data analysis engineer counted the consensus scores of the two experts, and for actions with inconsistent scores, another experienced neurologist made the final decision until a reliable score was obtained. This experimental design also avoids training errors due to systematic bias [40] and makes the machine learning model more reliable.

2.2. Devices

The wearable system data acquisition core used in this experiment is the inertial sensing unit. The QFN packaged multi-chip MPU-9250 (InvenSense, USA), which integrates a 3-axis accelerometer, 3-axis gyroscope, and 3-axis magnetometer AK8963 (AKM Semiconductor, USA) via embedded microcontroller (MSP430). The monitored digital signals are transmitted wirelessly via Bluetooth protocol (BR-LE4.0-D2). The capacitance on each axis of the accelerometer measures the degree of deviation of the axes respectively; the gyroscope measures based on the Coriolis effect and can measure the angular velocity of each axis rotation; the magnetometer collects the electromagnetic intensity of the geomagnetic field based on the Hall effect and can estimate the attitude orientation of the carrier and other information. In addition, the host computer of the data acquisition system can be used to set the acquisition frequency, calibrate the sensors, visualize the acquisition signal and store the data (Fig. 1(a)).

2.3. Data preprocessing

This study focused on the quantitative analysis of postural tremor, i.e., tremor symptoms that occur when the affected part of the body actively holds a specific posture. According to the CRST standard scale Part A (Fig. 1(b-e)), the patient was asked to sit comfortably in a chair with arms outstretched, wrist mildly extended and fingers spread apart, and then maintaining three postures in sequence: 1) wing-beat position, with arms outstretched, arms mildly extended and fingers spread apart; 2) arm extension, with arms outstretched, arms straight and palms down; 3) finger-finger position, with arms outstretched, arms straight and palms down. The index fingers of the left and right hands are opposite, with the palms facing inward.

Fig. 2 shows the flow of data analysis, which is mainly divided into data processing and feature extraction. In this study, the raw IMU data obtained from the acquisition are filtered, sliced, and transformed into clean equal-length sequences for feature extraction. In summary, the experimentally acquired postural sensing signals include ET tremor, physiological tremor, and various random noises. Before feature extraction, it is necessary to filter out the interference signals other than idiopathic tremor as much as possible, such as the subtle physiological tremor caused by normal muscle activation, whose unidirectional tremor amplitude is usually around 150 μm with a frequency of 8–12 Hz [36]. In contrast, the ET tremor amplitude will be much larger with a frequency band of 4–12 Hz [41]. Therefore, we ensure stable sensing signal output in dynamic environments by integrating dynamic Kalman filtering algorithms on the hardware side. Furthermore, the attitude solver can filter random errors of sensing elements (including zero bias, temperature drift, inter-axis alignment errors, etc.) at the front end to ensure the reliability of the signal and retain the high signal-to-noise ratio of the tremor signal. Although a stable pose signal can be obtained, small amplitude human intentional motion and physiological tremor are also mixed during the tremor motion. The wavelet transform is suitable for non-stationary signals and has good time-frequency localization characteristics in problems such as signal mutation, compression reconstruction, and denoising. Therefore, in this study, the sym3 wavelet is chosen, and the noisy tremor signal is decomposed independently at two levels based on a soft threshold function to reconstruct useful pathological tremor information.

Next, we slice the filtered IMU signal using a sliding window to ensure that the sequence length is fixed at 5 s to facilitate feature extraction. To achieve simple and effective state recognition, while considering the reduced dependence on wearing position and pose variability, we computed the signal vector magnitude (SVM) of a three-axis sequence. SVM is a commonly used method for processing nine-axis sensing signals, reducing the complexity of vector operations for each IMU signal axis from a macroscopic perspective. This step is illustrated in this paper using a three-axis acceleration signal. The SVM of the three-axis acceleration signal can be expressed $\boldsymbol{a}_{svm}$, which can be obtained from Eq. (1):

$$\boldsymbol{a}_{svm} = \sqrt{\boldsymbol{a}_x^2 + \boldsymbol{a}_y^2 + \boldsymbol{a}_z^2} \tag{1}$$

where $\boldsymbol{a}_x$, $\boldsymbol{a}_y$, and $\boldsymbol{a}_z$ denote the time series of the x-axis, y-axis, and z-axis, respectively. For a sampled data segment of length N, the final SVM sequence of accelerations is obtained as $\boldsymbol{a}_{svm} = \{a_{svm}(i), i = 1, 2, \ldots, N\}$.

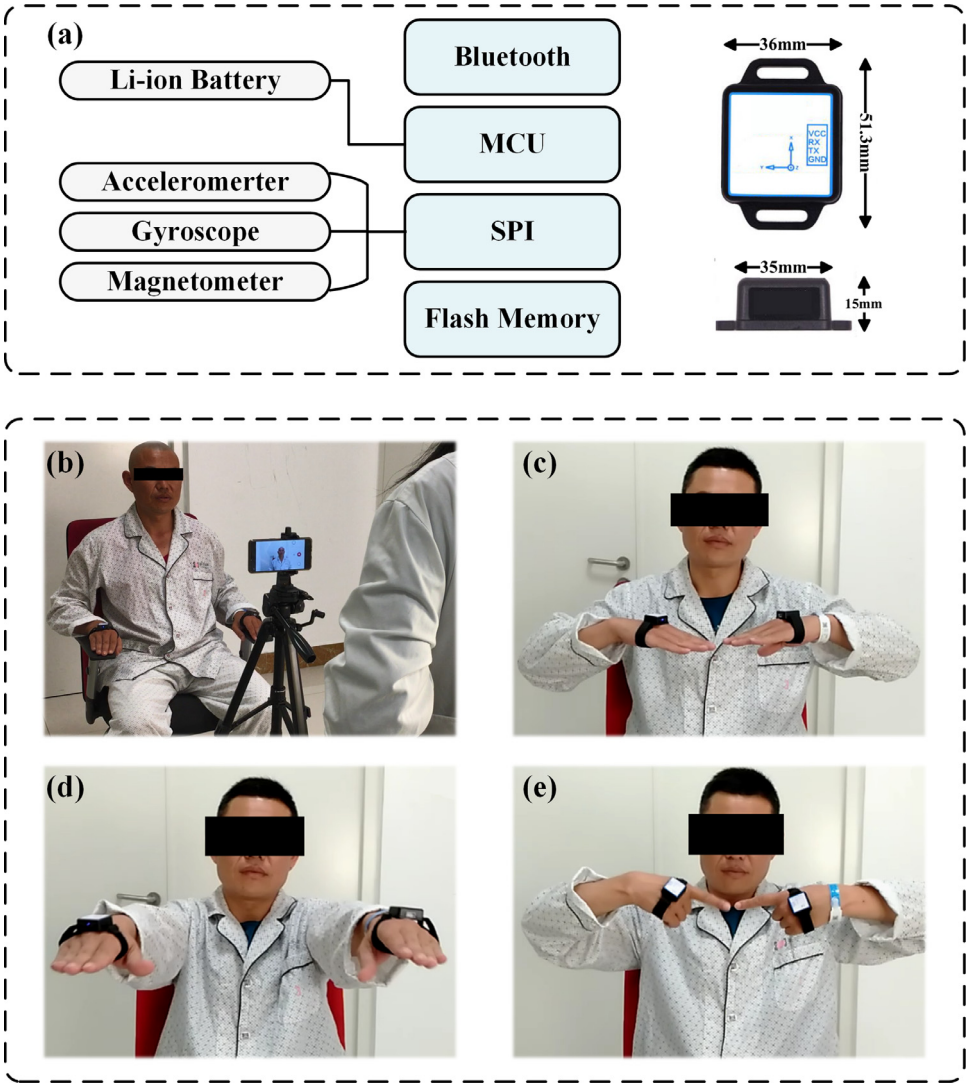

Fig. 1. Construction of the wearable device used in this study and the experimental demonstration. (a) shows the basic structure of the IMU-based wearable device. (b) demonstrates the experimental acquisition process. (c)-(e) demonstrate the three postural tremor tasks. **Notes**: MCU, micro controller unit; SPI, serial peripheral interface.

2.4. Feature metrics

In this study, 20 interpretable features, multi-domain features (including time-domain, frequency-domain, and nonlinear domain), were extracted from the SVM sequences of three-channel sensing signals (acceleration, angular velocity, and angle), respectively. The specific descriptions of the feature parameters are shown in Table 2. The time-domain parameters are all morphological features of the signal. Among them, skewness and kurtosis are the third-order standardized moment and fourth-order standardized moment of the signal, respectively, and their specific formulas are shown in Eq. (2-3):

$$a_{skew} = \frac{\frac{1}{N}\sum_{i=1}^{N}\left(a_{svm}(i) - a_{mean}\right)^3}{\left[\frac{1}{N}\sum_{i=1}^{N}\left(a_{svm}(i) - a_{mean}\right)^2\right]^{3/2}} \tag{2}$$

$$a_{kurt} = \frac{\frac{1}{N}\sum_{i=1}^{N}\left(a_{svm}(i) - a_{mean}\right)^4}{\left[\frac{1}{N}\sum_{i=1}^{N}\left(a_{svm}(i) - a_{mean}\right)^2\right]^2} \tag{3}$$

where a_{mean} denotes the sample mean of the SVM sequence of acceleration $\boldsymbol{a}_{svm}$.

In the frequency domain, the power spectrum is first calculated, and then the primary tremor frequency, peak power, and tremor stability index, which are the characteristics of the tremor frequency, are extracted. Moreover, frequency domain analysis gives more intuitive parameter characteristics than time-domain analysis in signal processing. Jeon et al. [8] concluded that the tremor signal is concentrated in the main frequency band. The peak power $P_m(f_p)$ within the signal band is defined as the area under the PSD curve within the main frequency interval and is calculated as expressed by Eq. (4):

$$\begin{aligned} a_{pwrp}(f_p, f_{th}) &= P_m(f_p) \\ &= \int_{f_p - f_{th}}^{f_p + f_{th}} \frac{1}{N^2}\left(S^*_{dft}(f) \times S_{dft}(f)\right) df \end{aligned} \tag{4}$$

where $f_p \pm f_{th}$ denotes the bandwidth of the peak power at the dominant frequency. Pierleoni et al. [42] argue that the peak power is preferable to the one-sided power spectrum of the sensor signal over a while (15 s) at the dominant frequency ±0.5 Hz around the length of the power estimate. $S^*_{dft}(f)$ is the complex conjugate of $S_{dft}(f)$, and $S_{dft}(f)$ denotes the discrete Fourier transform of the expressed power signal $\boldsymbol{a}_{svm}$, which can be calculated from Eq. (5):

$$S_{dft}(f) = \sum_{n=0}^{N-1} x_N \cdot e^{\frac{-i2\pi kn}{N}} \tag{5}$$

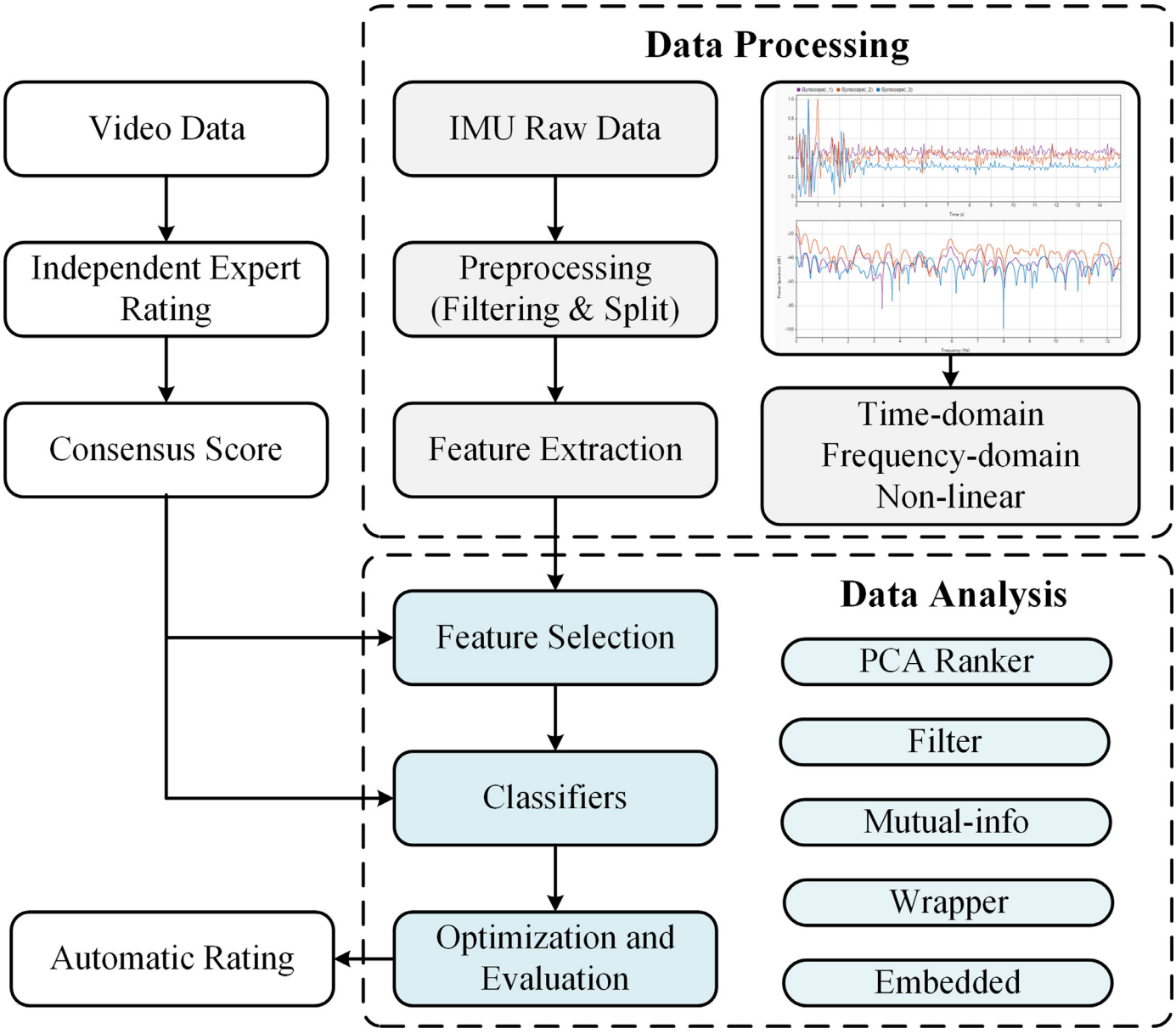

Fig. 2. Methodology flowchart. An automatic scoring system for ET patients with postural tremors was constructed based on two parts: data processing and data analysis.

Further from Eq. (6), we can calculate power spectral density (PSD) $P_S(f)$:

$$P_S(f) = \lim_{t\to\infty} \frac{1}{t}\left|S_{dft}(f)\right|^2 \tag{6}$$

PSD is widely defined as the signal power per unit frequency band, reflecting the frequency domain's signal power distribution. Besides, this study calculated the ratio of peak power to total power to confirm whether the signal is in tremor time, and Elble et al. [43] concluded that the percentage of peak power for the complete power estimate should be more significant than 85% to determine that the patient is in tremor. The tremor signal's peak power concerning the PSD estimate is shown in Fig. 3, and the shaded area indicates the tremor band of the primary frequency.

Finally, the nonlinear features of this study include approximate entropy (ApEn), sample entropy (SampEn), and fuzzy entropy (FuzzyEn). ApEn is a technique to quantify the degree of irregularity and unpredictability of time series data fluctuations [44]. In this study, For the acceleration sequence $\boldsymbol{a}_{svm} = \{a_{svm}(i), i = 1, 2, \ldots, N\}$, use a window truncation of length $m = 2$ to obtain the embedded sequence $\boldsymbol{X} = \{x(i), i = 1, 2, \ldots, N + m - 1\}$, the similarity tolerance is $r = 0.1 \times SD$(SD is the serial standard deviation), and ApEn is defined as

$$ApEn(\boldsymbol{a}_{svm}, m, r) = lnC_{Ap}^{m}(r) - lnC_{Ap}^{m+1}(r) \tag{7}$$

where $\boldsymbol{a}_{svm}$ represents a continuous 15 s sequence fragment, $C_{Ap}^{m}(r)$ denotes the average similarity rate for all m-length subfragments of the entire sequence under the similarity criterion r, the calculation formula is as follows:

$$C_{Ap}^{m}(r) = count(x(j)) \tag{8}$$

where $x(j)$ satisfy $max(|x(i) - x(j)|) \le r)/(N - m + 1)$.

The SampEn does not contain its vector's comparison when calculating the sequence self-similarity probability, so it is not limited by the data length. In contrast, fuzzy entropy proposes an unclear affiliation function that improves the binary process's similarity measure. This fuzzy boundary measure enhances signal complexity by fuzzy entropy and makes the entropy change more continuously and smoothly. According to a previous study [23], the

Table 2
Feature matrics of each channel sensing signal in the experimental phase and its definition.

Signal representation	Biomechanical features	Acronym	Formula
Time domain	Maximum peak value	Max	$a_{max} = max(a_{svm}(i))$
	Minimum valley value	Min	$a_{min} = min(a_{svm}(i))$
	Average value	Mean	$a_{mean} = \frac{1}{N}\sum_{i=1}^{N} a_{svm}(i)$
	Peak-to-peak amplitude	PeakP	$a_{peak} = max_i(\lvert a_{svm}(i)\rvert)$
	Average rectified value	Arv	$a_{arv} = \frac{1}{N}\sum_{i=1}^{N} \lvert a_{svm}(i)\rvert$
	Variance	Var	$a_{var} = \frac{1}{N}\sum_{i=1}^{N} (a_{svm}(i) - \mu)^2$
	Standard deviation	Std	$a_{std} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} (a_{svm}(i) - a_{mean})^2}$
	Kurtosis	Kurt	$a_{kurt} = \boldsymbol{E}[((a_{svm}(i) - \mu)/\sigma)^3]$
	Skewness	Skew	$a_{skew} = \boldsymbol{E}[((a_{svm}(i) - \mu)/\sigma)^4]$
	Root mean square	RMS	$a_{rms} = \sqrt{\frac{1}{N}\sum_{i=1}^{N-1} (a_{svm}(i))^2}$
	Shape factor	SF	$a_{sf} = a_{rms}/a_{arv}$
	Crest factor	CF	$a_{cf} = a_{peak}/(\frac{1}{N}\sum_{i=1}^{N} a_{svm}^2(i))^{1/2}$
	Impulse factor	IF	$a_{if} = a_{peak}/a_{arv}$
	Margin factor	LF	$a_{lf} = a_{peak}/\frac{1}{N}\sum_{i=1}^{N} (a_{svm}(i))^2$
Frequency domain	Tremor peak power in PSD	PwrP	$a_{pwrp}(f_p, f_{th}) = \int_{f_p - f_{th}}^{f_p + f_{th}} \frac{1}{N^2} (S_{dft}^*(f) \times S_{dft}(f))df$
	Peak power ratio	PwrP_R	$a_{pwrp\_r}(f_p, f_{th}) = a_{pwrp}(f_p, f_{th}) / \lim_{t\to\infty} \frac{1}{t}\lvert S_{dft}(f)\rvert^2$
	Principle peak in PSD	PrinP	$a_{prinp} = max_i(\lvert P_S(f)\rvert)$
Nonlinear domain	Sample energy	SampEn	$SampleEn(\boldsymbol{a}_{svm}, m, r) = lnC_{sample}^{m}(r) - lnC_{sample}^{m+1}(r)$
	Approximate entropy	ApEn	$ApEn(\boldsymbol{a}_{svm}, m, r) = lnC_{Ap}^{m}(r) - lnC_{Ap}^{m+1}(r)$
	Fuzzy entropy	FuzEn	$FuzzyEn(\boldsymbol{a}_{svm}, m, r) = lnC_{fuzzy}^{m}(r) - lnC_{fuzzy}^{m+1}(r)$

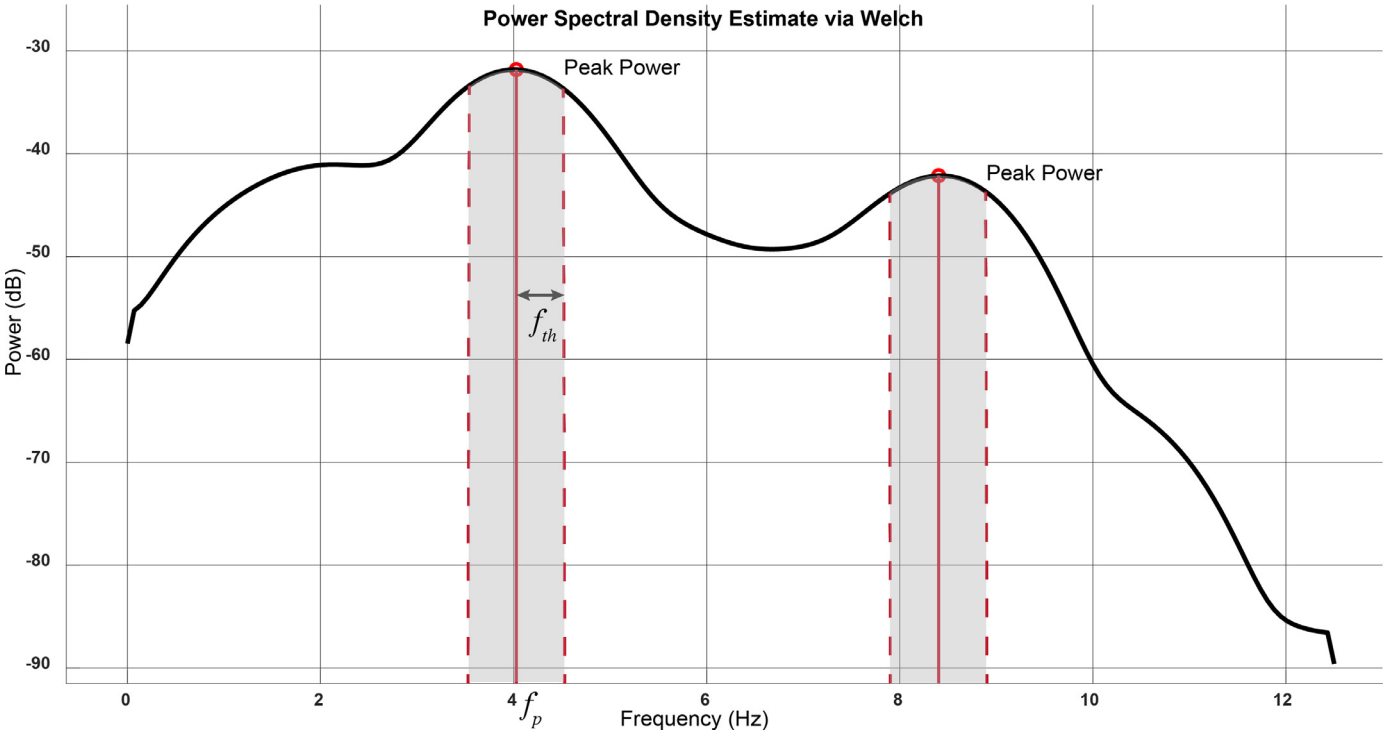

Fig. 3. Power spectral density function calculated from the acceleration signal's vector amplitude (intercepted 15 s signal segment). The method based on the power spectrum tolerance span is used to find the peak power, and the shaded area shows the tremor signal's peak power.

entropy feature characterizing the sequence complexity can sufficiently improve the tremor quantization model's performance, so it is all altogether estimated in this study.

2.5. Model selection

The automatic postural tremor severity assessment system is modeled based on machine learning algorithms. This study includes decision tree (DT), tree ensembles, support vector machine (SVM), Naïve Bayesian (NB), and k-nearest neighbor (KNN) algorithms. The optimal global solution for each model is selected using grid search and Bayesian optimization algorithms. The Bayesian optimization algorithm's collection function is expected to improve per second, and each model is iterated in 30 epochs to obtain the optimal solution. Grid search is preferred for models with fewer hyperparameters. The datasets are sliced by subject-wise cross-validation to ensure that postural signals of the same subject would not appear in the train and test data, which can avoid

Table 3
Model hyper-parameters of the classifiers.

Classifiers	Hyper-parameters Search Spaces	Optimizer Strategy	Feature Selection Methods
SVM	Kernel function: Gaussian, Linear, Quadratic, Ternary; Kernel scale: 0.001–200; Box constraint: 0.001–200	Bayesian optimization	PCA;
KNN	Neighbors: 1–15; Distance metric: City Block, Chebyshev, Correlation, Cosine, Euclidean, Hamming, Jaccard, Mahalanobis, Minkowski, Spearman	Grid Search	Filter (low variance); Mutual information selection;
Ensemble	Methods: AdaBoost, RUSBoost, Random Forest, XGBoost; Estimators: 10–500; Learning rate: 0.001–1; Max depth: 1–6935	Bayesian optimization	Wrapper (XGBoost);
DT	Maximum number of splits: 1–1500; Split criterion: Gini's diversity index, Towing rule, Maximum deviance reduction	Grid Search	Embedded (XGBoost);
NB	Dataset distribution assumptions: Gaussian and Multinomial Distribution; Kernel type: Gaussian, Box, Epanechnikov, Triangle	Grid Search	All Features

Notes: SVM, support vector machine; KNN, k-nearest neighbor; DT, decision tree; NB, naive Bayesian; PCA, principal components analysis.

sampling errors as much as possible [45]. All models are subjected to five-fold cross-validation to demonstrate the stability, and the mean and standard deviation of the evaluation metrics on the five test sets, are recorded. In general, the smaller the standard deviation of the test results means that predictions are credible.

Various feature selection methods are used for the models with the best classification performance, including 1) PCA, which selects different interpretable variance thresholds for feature selection of principal components. 2) Filter (low variance): the simplest feature selection method based on a low variance threshold eliminates all features with too slight variance. 3) Mutual information selection: scoring each feature according to the relevance of mutual information and selecting the top six features for modeling. 4) Wrapper (XGBoost): based on the hold-out method, for each subset of features to be selected, the base model is trained once on the training set (this experiment is chosen as XGBoost, a tree model-based Boosting algorithm that adds regularization to the cost function to prevent overfitting while supporting parallel operations and faster inference), and then selects a subset of features on the test set based on the error size. 5) Embedded (XGBoost): using XGBoost for the base model for training, the top six features are selected for modeling based on the weight coefficients of the features. The detailed model hyperparameter settings are shown in Table 3.

2.6. Performance metrics

Five main measures were commonly considered in the literature to evaluate the performance of classification results with the severity of tremor, including the accuracy (ACC), sensitivity (SEN), specificity (SPEC), precision (PRE), and F1 score. We choose the F1 score to balance false positive and false negative, which provides the global metrics of the classifiers [46]. In contrast, the receiver operating characteristic (ROC) curve considers classification thresholds that weigh sensitivity and specificity. Finally, the area under the curve (AUC) is often used as a performance metric, primarily when the database is unevenly distributed. Because the confusion matrix clearly illustrates the classification performance of each of the five severity levels, a global average was calculated for this study for the combined metrics of the classification models in this study.

3. Results

The feature parameters extracted from the three tasks of postural tremor were analyzed for statistical variability. Table 4 demonstrates the hand characteristic values of ET patients with significant differences before and after the MRgFUS procedure. All data are expressed as mean ± SD. The normality was tested using the Shapiro-Wilk test to compare the variability of the characteristic parameters between levels with a statistical significance level of $p < 0.05$. A paired t-test was performed on the parameters grouped surgery to see some changes in the severity of tremors. The results show that the kinematic parameters of the three sensing modalities are significantly different under all postural tremor tasks. The gyroscopic sensing signal seems critical in the arm extension, while the magnetometer seems more discriminative in the finger-finger position. Notably, most feature parameters exhibited significant differences in both the left and right hand in the wing-beat position and arm extension tasks, while the discriminative features in the finger-finger position task were primarily found in the left hand. This finding may be helpful for subsequent studies.

In this study, 60 time-domain, frequency-domain, and nonlinear features extracted from three sensors were automatically modeled to quantify the severity of postural tremor in ET patients. The hyperparametric search space of the five machine learning models is shown in Table 3, and the performance metrics of the optimized five classification problems are summarized in Table 5 Fig. 4.(a-c) represents the performance of the five machine learning models in wing-beat position, arm extension, and finger-finger position, respectively. Fig. 5(a-c) visualize the confusion matrix of the optimal integration algorithm overall feature sets.

As can be seen from Table 5, the accuracy of all algorithms can be maintained above 85.72%, indicating that the classification results are in good agreement with the actual scores. However, the scores of both sensitivity and precision are low, so the F1 score, as a reconciled average of sensitivity and precision, better shows the ability to control the risk and cost of decision simultaneously. Combining the F1 scores of the three tasks, we can see that AdaBoost has the best performance, with an average of 93.47%. The variance in the five-fold cross-validation of the performance matrix shows that the integrated learner has better stability, and its performance fluctuations in each task are minimal. In contrast, weak learners, such as NB, have lower fitting ability, resulting in unstable test results. In addition, AUC, as a metric that weighs sensitivity and specificity, is a better indicator of the model's generalization ability in an unbalanced dataset. According to the ROC curves in Fig. 4(a-c), it can be found that the AUC of AdaBoost can be above 99.1% regardless of the basic probability of each category, and the performance fluctuation is less compared with other machine learning models, and the test results are more stable. Fig. 5(a-c) plots the optimal AdaBoost algorithm and the confusion matrix in the three pose tremor classification tasks. It can be seen that the wing-beat position is the most challenging task for the AdaBoost model. However, except for the mild tremor CRST 1–2, which has the lowest sensitivity of 83.70% and 89.83%, all other classes have indexes above 90.04%. The detection accuracy of Arm extension was poor only for tremor class 1, with a sensitivity of 85.28%, 91.84%, 94.30%, and 94.69%.

The above results show that the ensemble learning methods based on the complete feature sets can effectively classify all five

Table 4
Statistical Significance of Features Analysis (mean ± SD).

Task	Features	Left			Right		
		Before	After	*p-value*	Before	After	*p-value*
Wing-beat position	ACC_Arv	−0.308 ± 0.302	0.774 ± 0.704	0.031*	−0.56 ± 0.362	0.545 ± 0.436	< 0.001*
	ACC_Var	0.526 ± 0.259	−0.549 ± 0.401	0.017*	0.416 ± 0.283	−0.567 ± 0.267	< 0.001*
	ACC_IF	0.054 ± 0.036	−0.499 ± 0.141	0.007*	0.097 ± 0.113	−0.595 ± 0.176	< 0.001*
	ACC_LF	0.111 ± 0.138	−0.447 ± 0.299	0.012*	0.125 ± 0.194	−0.514 ± 0.103	0.002*
	GYRO_Max	0.126 ± 0.063	−0.480 ± 0.167	0.009*	0.167 ± 0.113	−0.559 ± 0.131	0.001*
	GYRO_Min	0.143 ± 0.136	−0.498 ± 0.497	0.006*	0.226 ± 0.123	−0.603 ± 0.207	< 0.001*
	GYRO_Mean	0.026 ± 0.063	−0.489 ± 0.367	0.009*	0.027 ± 0.113	−0.559 ± 0.131	< 0.001*
	GYRO_Arv	0.136 ± 0.193	−0.471 ± 0.113	0.007*	0.211 ± 0.298	−0.581 ± 0.191	< 0.001*
	GYRO_Var	−0.408 ± 0.352	0.511 ± 0.495	0.012*	−0.345 ± 0.269	0.675 ± 0.311	0.003*
	GYRO_Kurt	0.114 ± 0.167	−0.482 ± 0.376	0.008*	0.169 ± 0.171	−0.564 ± 0.136	0.001*
	MAG_PwrP	0.247 ± 0.308	−0.464 ± 0.307	0.007*	0.227 ± 0.267	−0.334 ± 0.361	0.049*
	MAG_PrinP	0.135 ± 0.062	0.221 ± 0.231	0.064	0.125 ± 0.021	−0.373 ± 0.375	0.036*
	MAG_SampEn	0.017 ± 0.021	−0.504 ± 0.466	0.024*	0.013 ± 0.022	−0.566 ± 0.343	0.041*
	MAG_FuzEn	0.047 ± 0.027	−0.58 ± 0.549	0.010*	0.058 ± 0.042	−0.436 ± 0.763	0.013*
Arm extension	ACC_IF	0.153 ± 0.133	−0.405 ± 0.252	0.004*	0.129 ± 0.253	−0.535 ± 0.030	0.004*
	ACC_LF	0.051 ± 0.092	−0.391 ± 0.066	0.023*	0.099 ± 0.151	−0.421 ± 0.013	0.012*
	GYRO_Max	0.110 ± 0.133	−0.385 ± 0.212	0.010*	0.155 ± 0.117	−0.485 ± 0.021	0.006*
	GYRO_Min	0.225 ± 0.264	−0.380 ± 0.425	0.002*	0.289 ± 0.315	−0.381 ± 0.044	0.003*
	GYRO_Mean	0.113 ± 0.133	−0.385 ± 0.212	0.010*	0.455 ± 1.17	−0.485 ± 0.02	0.006*
	GYRO_PeakP	0.011 ± 0.013	−0.283 ± 0.114	0.006*	0.018 ± 0.008	−0.193 ± 0.163	0.064
	GYRO_Arv	0.317 ± 0.245	−0.392 ± 0.374	0.001*	0.655 ± 0.478	−0.578 ± 0.042	0.004*
	GYRO_Var	−0.492 ± 0.303	0.357 ± 0.312	0.001*	−0.499 ± 0.303	0.306 ± 0.431	0.051
	GYRO_Std	−0.497 ± 0.475	0.285 ± 0.142	0.038*	−0.471 ± 0.534	0.263 ± 0.286	< 0.001*
	GYRO_Kurt	0.123 ± 0.133	−0.388 ± 0.219	0.008*	0.164 ± 0.181	−0.492 ± 0.021	0.005*
	GYRO_PwrP	−1.021 ± 0.648	0.156 ± 0.166	0.003*	−0.931 ± 1.075	0.103 ± 0.184	0.025*
	GYRO_PwrP_R	0.424 ± 0.379	−0.686 ± 0.427	< 0.001*	0.512 ± 0.479	−0.597 ± 0.612	0.033*
	GYRO_PrinP	−1.342 ± 1.433	0.523 ± 0.578	0.003*	−1.326 ± 1.007	0.554 ± 0.391	0.013*
	GYRO_SampEn	−1.204 ± 1.104	0.582 ± 0.270	0.004*	−1.479 ± 0.747	0.519 ± 0.427	0.004*
	GYRO_FuzEn	−1.268 ± 0.979	0.613 ± 0.339	0.001*	−1.513 ± 0.856	0.533 ± 0.369	0.003*
	MAG_PwrP	0.087 ± 0.084	−0.631 ± 0.500	< 0.001*	0.825 ± 0.085	−0.450 ± 0.431	0.002*
	MAG_PrinP	0.238 ± 0.272	−0.482 ± 0.459	0.002*	0.257 ± 0.174	−0.519 ± 0.316	0.001*
	MAG_SampEn	0.164 ± 0.113	−0.657 ± 0.665	0.001*	0.183 ± 0.173	−0.541 ± 0.541	< 0.001*
	MAG_ApEn	0.247 ± 0.156	−0.395 ± 0.285	0.031*	0.238 ± 0.225	−0.486 ± 0.034	0.006*
	MAG_FuzEn	0.174 ± 0.151	−0.652 ± 0.513	< 0.001*	0.125 ± 0.161	−0.614 ± 0.614	< 0.001*
Finger-finger position	ACC_Max	−0.299 ± 0.147	−0.301 ± 0.801	0.993	−0.105 ± 0.167	−0.568 ± 0.138	0.006*
	ACC_ApEn	0.105 ± 0.136	−0.357 ± 0.207	0.300	0.175 ± 0.053	−0.473 ± 0.536	0.019*
	GYRO_LF	0.051 ± 0.047	0.003 ± 0.022	0.043*	0.052 ± 0.073	0.011 ± 0.032	0.095
	MAG_Max	0.662 ± 0.496	0.112 ± 0.016	0.032*	0.671 ± 0.678	0.271 ± 0.273	0.092
	MAG_Min	0.575 ± 0.504	0.056 ± 0.081	0.042*	0.509 ± 0.669	0.053 ± 0.093	0.061
	MAG_Skew	0.573 ± 0.504	0.055 ± 0.089	0.042*	0.508 ± 0.569	0.051 ± 0.089	0.062
	MAG_ApEn	0.015 ± 0.007	−0.685 ± 0.142	0.048*	0.015 ± 0.006	−0.638 ± 0.225	0.002*

Table 5
Classification performance (mean ± SD) of each optimized machine learning method in three postural tremor tasks.

Task	Classifiers	Optimal Hyperparameters	ACC (%)	SEN (%)	SPEC (%)	PRE (%)	F1 score (%)
Wing-beat position	SVM	Quadratic Kernel Function; 0.0098 Box constraint level; One-vs-One	85.72 ± 1.45	58.87 ± 2.13	90.82 ± 0.83	64.42 ± 3.17	58.79 ± 2.16
	KNN	5-Neighbors; Cosine; Squared inverse	89.52 ± 1.28	66.44 ± 1.69	93.03 ± 1.23	68.14 ± 2.17	67.13 ± 1.59
	Ensemble	AdaBoost; 498 estimators; 0.9247 learning rate	**97.71 ± 0.13**	**92.05 ± 0.51**	**98.44 ± 0.06**	**94.92 ± 1.28**	**93.37 ± 0.85**
	DT	Max deviance reduction; 640 split number	94.19 ± 0.88	83.58 ± 1.56	96.13 ± 0.68	83.41 ± 1.96	83.48 ± 1.31
	NB	Gaussian Kernel Function	86.26 ± 1.57	62.49 ± 2.95	90.99 ± 1.36	61.04 ± 2.55	61.49 ± 2.13
Arm extension	SVM	RBF Kernel Function; 165.522 Kernel Scale; 966.689 Box Constraint; One-vs-All	89.56 ± 1.23	67.23 ± 2.16	92.99 ± 1.50	71.20 ± 2.11	68.81 ± 2.09
	KNN	5-Neighbors; Spearman Distance; Inverse Distance Weighted	89.47 ± 1.33	68.43 ± 2.06	92.88 ± 1.09	71.19 ± 1.83	69.55 ± 1.91
	Ensemble	AdaBoost; 324 estimators; 0.9917 learning rate	**97.54 ± 0.21**	**92.51 ± 1.37**	**98.35 ± 0.33**	**93.94 ± 0.63**	**93.17 ± 1.08**
	DT	Max deviance reduction; 747 split number	94.43 ± 0.59	83.58 ± 1.22	96.31 ± 0.89	84.18 ± 1.96	83.86 ± 0.85
	NB	Gaussian Kernel Function	86.17 ± 1.44	67.20 ± 3.17	91.26 ± 1.28	62.26 ± 3.25	63.23 ± 3.23
Finger-finger position	SVM	RBF Kernel Function; 2.7194 Kernel Scale; 11.5243 Box Constraint	88.05 ± 1.19	63.95 ± 3.96	92.01 ± 1.06	67.46 ± 3.07	65.32 ± 3.28
	KNN	6-Neighbors; City Block Distance; Inverse Distance Weighted	91.98 ± 0.84	76.27 ± 2.06	94.60 ± 1.23	78.59 ± 2.33	77.23 ± 2.17
	Ensemble	AdaBoost; 94 estimators; 0.9592 learning rate	**97.72 ± 0.31**	**93.24 ± 0.55**	**98.47 ± 0.66**	**94.58 ± 0.93**	**93.86 ± 1.22**
	DT	Max deviance reduction; 560 split number	94.48 ± 0.93	84.38 ± 1.57	96.33 ± 0.84	85.05 ± 1.63	84.70 ± 1.29
	NB	Gaussian Kernel Function	86.33 ± 3.28	67.72 ± 4.58	91.33 ± 2.34	62.77 ± 4.59	63.82 ± 4.66

Notes: SVM, support vector machine; KNN, k-nearest neighbor; DT, decision tree; NB, naive Bayesian.

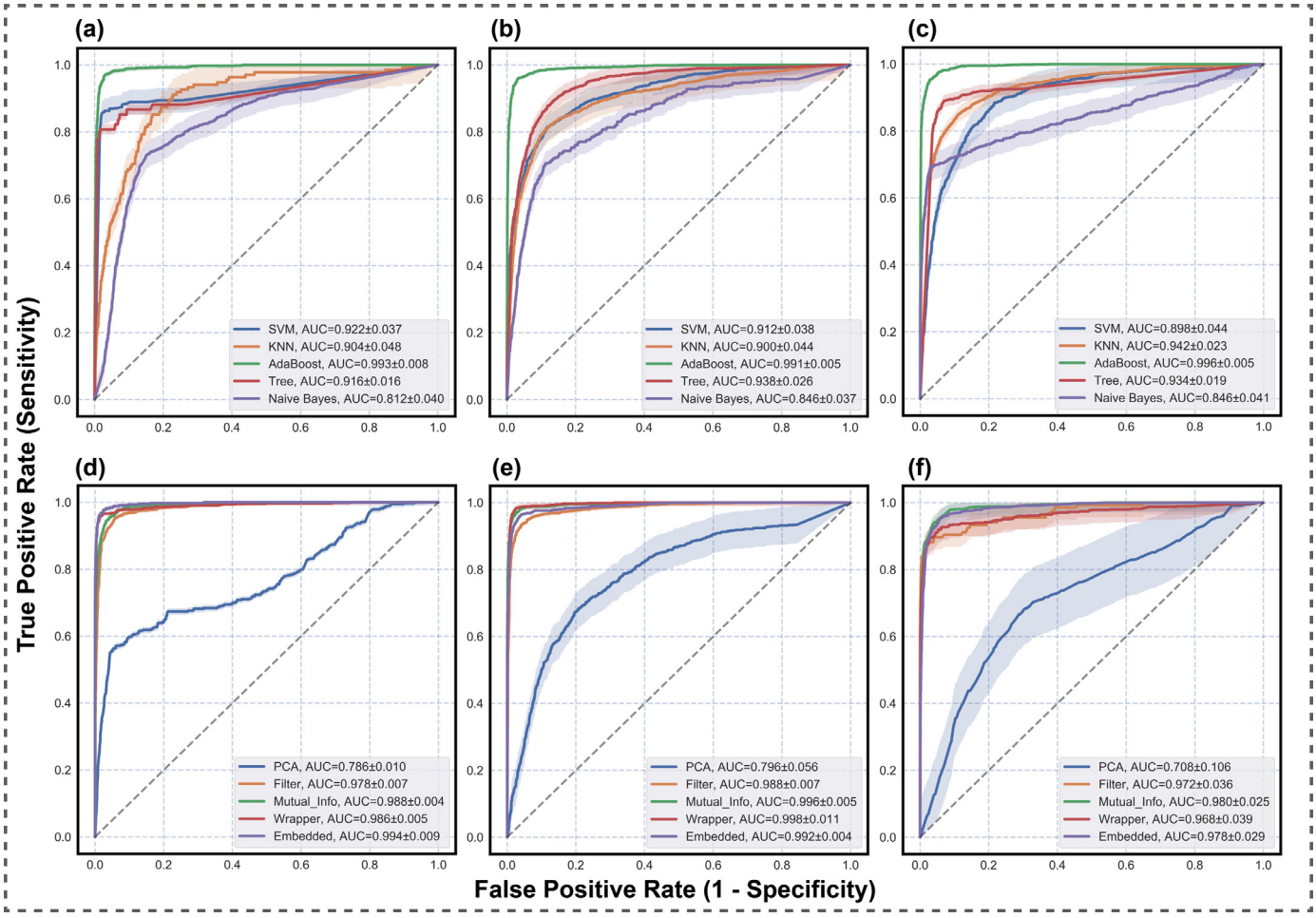

Fig. 4. Receiver operating characteristic figures of the five classification models, which show the results of the five-fold cross-validation. (a-c) ROC curves of the five machine learning models on all feature sets in the wing-beat position, arm extension, and finger-finger position tasks, respectively. (d-f) ROC curves of the integrated learning tree models based on the five feature selection methods in the three tasks. In addition, this study shows the ROC curves of the optimal Ensemble model based on different feature subsets for the three tasks (Fig. 5(d-f)).

severity levels in all three postural tremors. Furthermore, this study explores the optimization of five machine learning models through different feature selection methods to evaluate meaningful clinical features and speed up the inference of the algorithms. Table 6 summarizes the classification results of five machine learning models based on PCA ranker, Filter low variance, Mutual information, Wrapper subset evaluation and Embedded selected feature subsets to optimize classification results. The global metrics AUC and accuracy show that the Ensemble algorithm performs best in all feature subsets for all three pose tremor tasks. PCA reduces the number of features by 93.33% by selecting 95% of the interpretable variance and compressing the features to less than four. The AUCs of this method for the wing-beat position, arm extension and finger-finger position tasks were 0.870, 0.926 and 0.754, respectively. Although there is a slight loss in performance compared to other feature selection methods, it can significantly speed up the model operations.

For different tasks, the final selection of features and algorithms differ. Finally, the study visualizes the confusion matrix of the optimized ensemble learning methods obtained by training based on the optimal feature selection method by Fig. 5(d-f). Combined with Table 6, it can be seen that the experimental results found that the optimal classification results of RUSBoost based on Embedded way in wing-beat position task, with a total of 16 feature parameters selected (73.33% of features eliminated) and five severity levels of specificity between 97.46%–99.54%. In contrast, using the AdaBoost model based on the Wrapper method in the arm extension task, the macro F1 score can reach 94.49% by selecting only 12 feature parameters. Even for the category with severity level 4, the accuracy, sensitivity, specificity, precision, and F1 scores were 99.45%, 96.32%, 99.72%, 96.57%, and 96.44%, respectively. The Bagging model based on the Mutual information method was used in the finger position task. The model readily misidentified CRST 1 as free of tremor symptoms, but the F1 scores for the other types ranging from 93.21%–94.47%.

4. Discussion

4.1. Clinical application

This study proposes an automatic quantitative assessment of postural tremor symptom severity based on a posture sensor, and a sizable amount of ET patient data is collected for modeling and testing the evaluation in a real clinical scenario. In addition, the extracted useful features are analyzed with the help of statistical methods and multiple feature selection algorithms. These features can provide helpful reference indicators for the clinical diagnosis of neurologists. Compared with existing methods, our approach can

Table 6
Predictive performance of the machine learning classifiers based on different feature selection methods.

Task	Feature selection method	Selected features	Classification method				
			SVM	KNN	Ensemble	DT	Naïve Bayes
Wing-beat position	PCA ranker	Principal Components (PC1-PC4)	AUC = 0.870 ACC = 0.662	AUC = 0.888 ACC = 0.746	**AUC = 0.936** **ACC = 0.808**	AUC = 0.886 ACC = 0.753	AUC = 0.832 ACC = 0.652
	Filter	ACC_SampEn; ACC_Arv; ACC_Var; ACC_Kurt; ACC_SF; ACC_CF; ACC_IF; GYRO_PwrP_R; GYRO_PrinP; GYRO_FuzEn; GYRO_Mean; GYRO_Arv; GYRO_CF; MAG_PwrP; MAG_PwrP_R; MAG_Var; MAG_RMS; MAG_SF	AUC = 0.920 ACC = 0.763	AUC = 0.954 ACC = 0.862	**AUC = 0.978** **ACC = 0.882**	AUC = 0.902 ACC = 0.833	AUC = 0.816 ACC = 0.665
	Multual information	ACC_RMS; ACC_SF; ACC_CF; ACC_IF; ACC_LF; ACC_Max; ACC_Mean; GYRO_Max; GYRO_Min; GYRO_Mean; GYRO_PeakP; GYRO_Arv; GYRO_Var; GYRO_Skew; GYRO_LF; MAG_RMS; MAG_Max	AUC = 0.944 ACC = 0.803	AUC = 0.980 **ACC = 0.963**	**AUC = 0.988** ACC = 0.929	AUC = 0.934 ACC = 0.872	AUC = 0.826 ACC = 0.677
	Wrapper subset evaluation	ACC_RMS; ACC_Max; ACC_Min; ACC_Mean; ACC_PeakP; ACC_Arv; ACC_Var; ACC_Skew; ACC_RMS; ACC_SF; ACC_CF; ACC_IF; ACC_LF; GYRO_RMS; GYRO_Max; MAG_RMS; MAG_FuzEn	AUC = 0.908 ACC = 0.751	AUC = 0.982 **ACC = 0.973**	**AUC = 0.986** ACC = 0.906	AUC = 0.900 ACC = 0.823	AUC = 0.802 ACC = 0.644
	Embedded	ACC_RMS; ACC_Min; ACC_Arv; ACC_LF; GYRO_RMS; GYRO_Max; GYRO_Min; GYRO_Arv; GYRO_SF; GYRO_LF; MAG_FuzEn; MAG_RMS; MAG_Max; MAG_Mean; MAG_SF; MAG_CF	AUC = 0.972 ACC = 0.872	AUC = 0.942 ACC = 0.915	**AUC = 0.988** **ACC = 0.928**	AUC = 0.928 ACC = 0.876	AUC = 0.820 ACC = 0.674
Arm extension	PCA ranker	Principal Components (PC1-PC4)	AUC = 0.926 ACC = 0.749	AUC = 0.920 ACC = 0.749	**AUC = 0.934** **ACC = 0.776**	AUC = 0.872 ACC = 0.728	AUC = 0.830 ACC = 0.659
	Filter	ACC_PwrP; ACC_ApEn; ACC_Min; ACC_PeakP; ACC_Arv; ACC_Kurt; GYRO_PwrP; GYRO_Max; GYRO_PeakP; GYRO_Std; MAG_RMS	AUC = 0.888 ACC = 0.712	AUC = 0.816 ACC = 0.626	**AUC = 0.964** **ACC = 0.846**	AUC = 0.896 ACC = 0.771	AUC = 0.838 ACC = 0.621
	Multual information	ACC_RMS; ACC_Max; ACC_Mean; ACC_LF; GYRO_Max; GYRO_Min; GYRO_Mean; GYRO_PeakP; GYRO_Skew; GYRO_LF; MAG_RMS; MAG_Max	AUC = 0.938 ACC = 0.786	AUC = 0.976 ACC = 0.883	**AUC = 0.996** **ACC = 0.948**	AUC = 0.946 ACC = 0.882	AUC = 0.838 ACC = 0.635
	Wrapper subset evaluation	ACC_RMS; ACC_Min; ACC_Arv; ACC_LF; GYRO_Max; GYRO_Min; GYRO_Arv; GYRO_LF; GYRO_RMS; MAG_FuzEn; MAG_RMS; MAG_Max	AUC = 0.974 ACC = 0.838	AUC = 0.964 ACC = 0.864	**AUC = 0.994** **ACC = 0.936**	AUC = 0.946 ACC = 0.881	AUC = 0.872 ACC = 0.68
	Embedded	ACC_RMS; ACC_Max; ACC_Mean; ACC_Arv; ACC_RMS; ACC_SF; ACC_LF; GYRO_RMS; GYRO_Min; GYRO_Mean; GYRO_Arv; GYRO_LF; MAG_FuzEn; MAG_RMS; MAG_Max; MAG_SF; MAG_CF	AUC = 0.962 ACC = 0.849	AUC = 0.942 ACC = 0.909	**AUC = 0.994** **ACC = 0.940**	AUC = 0.936 ACC = 0.876	AUC = 0.832 ACC = 0.620
Finger-finger position	PCA ranker	Principal Components (PC1-PC2)	**AUC = 0.754** ACC = 0.574	AUC = 0.750 ACC = 0.583	AUC = 0.738 **ACC = 0.584**	AUC = 0.742 ACC = 0.576	AUC = 0.698 ACC = 0.558
	Filter	ACC_PwrP; ACC_Max; ACC_Arv; ACC_Std; ACC_Skew; ACC_SF; ACC_LF; GYRO_PrinP; GYRO_SampEn; GYRO_ApEn; GYRO_FuzEn; GYRO_Arv; GYRO_Kurt; GYRO_CF; GYRO_IF; GYRO_LF; MAG_FuzEn; MAG_RMS; MAG_Max; MAG_Min; MAG_Mean; MAG_PeakP; MAG_Var; MAG_SF	AUC = 0.880 ACC = 0.690	AUC = 0.822 ACC = 0.767	**AUC = 0.954** **ACC = 0.868**	AUC = 0.862 ACC = 0.768	AUC = 0.698 ACC = 0.550
	Multual information	ACC_Max; ACC_Mean; ACC_SF; ACC_CF; ACC_IF; ACC_LF; GYRO_Max; GYRO_Min; GYRO_Mean; GYRO_PeakP; GYRO_LF; MAG_RMS; MAG_Max	AUC = 0.824 ACC = 0.689	AUC = 0.932 ACC = 0.840	**AUC = 0.980** **ACC = 0.929**	AUC = 0.918 ACC = 0.839	AUC = 0.696 ACC = 0.544
	Wrapper subset evaluation	ACC_RMS; ACC_Min; ACC_Mean; ACC_Arv; ACC_LF; GYRO_RMS; GYRO_Max; GYRO_Min; GYRO_Mean; GYRO_LF; MAG_FuzEn; MAG_Max	AUC = 0.828 ACC = 0.713	AUC = 0.850 ACC = 0.714	**AUC = 0.968** **ACC = 0.893**	AUC = 0.880 ACC = 0.793	AUC = 0.702 ACC = 0.522
	Embedded	ACC_RMS; ACC_Min; ACC_Mean; ACC_Arv; ACC_SF; ACC_IF; ACC_LF; GYRO_RMS; GYRO_Max; GYRO_Min; GYRO_Mean; GYRO_Arv; GYRO_RMS; GYRO_LF; MAG_FuzEn; MAG_Max; MAG_Min	AUC = 0.910 ACC = 0.775	AUC = 0.904 ACC = 0.890	**AUC = 0.980** **ACC = 0.928**	AUC = 0.910 ACC = 0.843	AUC = 0.732 ACC = 0.545

Notes: SVM, support vector machine; KNN, k-nearest neighbor; DT, decision tree; NB, naive Bayesian; AUC, the area under curve; ACC, accuracy.

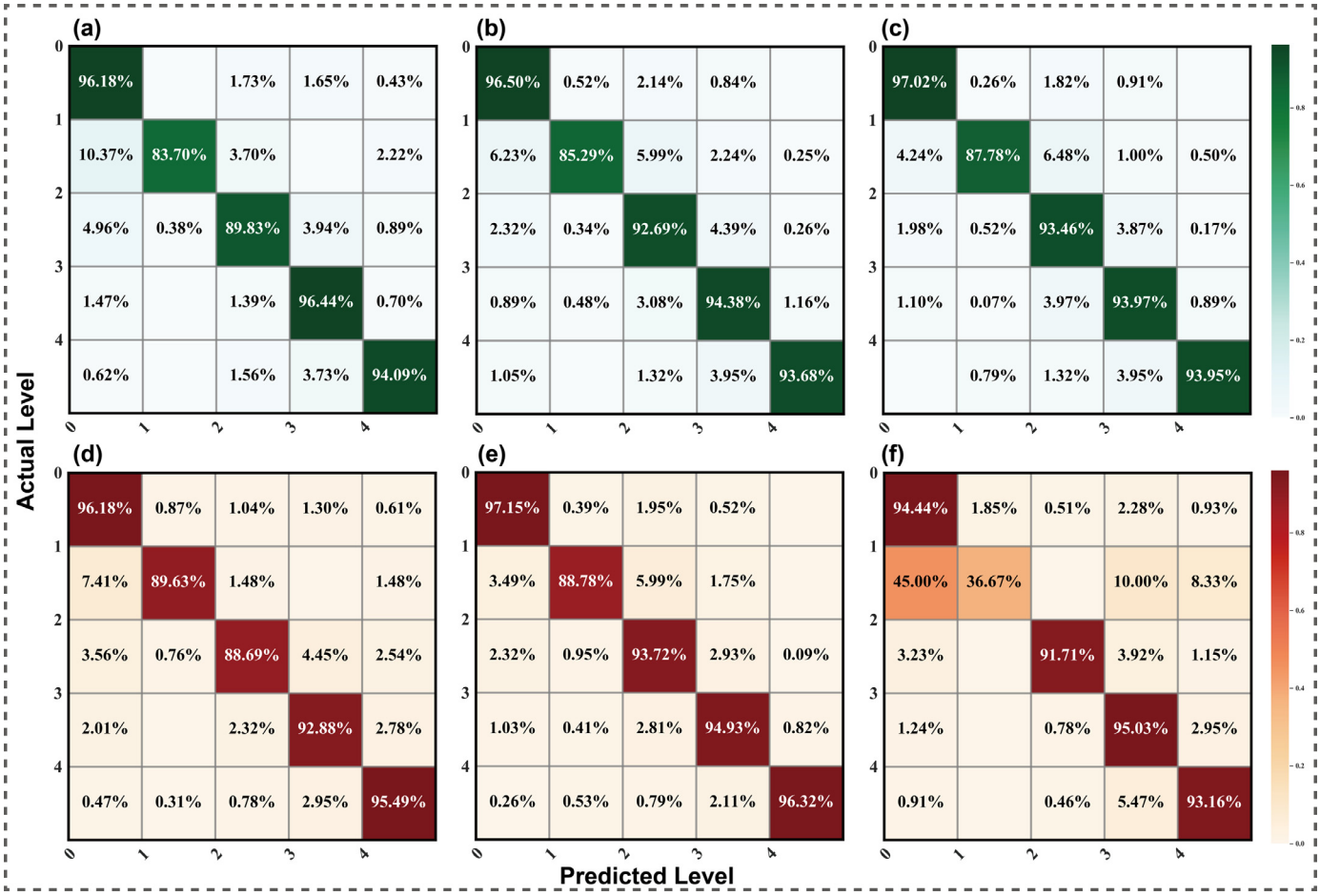

Fig. 5. Confusion matrices of the optimal five-classification model. (a-c) Confusion matrices of the optimal classifier AdaBoost overall feature sets for the Wing-beat position, Arm extension, and Finger-finger position tasks. (d-f) denote the optimal integrated learning tree model's confusion matrices based on the optimal feature selection method in the three tasks. Among them, (d) indicates the classification results of RUSBoost based on the Embedded method; (e) represents the classification results of AdaBoost based on the Wrapper method; (f) denotes the classification results of Bagging based on the Mutual information method.

provide an estimated CRST score with high accuracy and minimal error margin after learning a large amount of data collected by IMU-based wearable devices with high dimensionality and sensitivity.

As demonstrated by the studies summarized in Table 7, the collection of tremor activity data from patients in ADL seems promising for assessing the severity of symptoms through longitudinal monitoring to provide a basis for clinical assessment of the condition [7,9,14–18,25,30,47–52]. Many studies have even used sensors embedded in cell phones for collection [10,19–22], and these devices allow for collecting large amounts of reference data more quickly. However, due to limitations in power consumption and data storage, these wearable devices tend to have low sampling rates and signal dimensions and have difficulty resolving noise interference in everyday life. Furthermore, access to expert-annotated data is also a significant challenge and leaving room for improvement in their findings [25,47–49,51–53]. On the other hand, most studies are limited to determining whether the tremor occurs [18,21,22] and did not include a more detailed diagnosis of disease severity typing. In conclusion, accurate assessment of the disease requires standardized measurements in a relatively controlled experimental paradigm that can provide clinically meaningful statistical characteristics and automatically quantify tremor severity to better track disease progression.

Our experimental results demonstrate that spatio-temporal features extracted from multi-sensory fusion signals can effectively characterize the severity of the postural tremor. However, the need for higher diagnostic accuracy could benefit from more regulated data acquisition, higher sensing resolution, and more multi-sensory fusion signal channels [7,26–29]. This will increase the available data and improve the quantity and quality of the information obtained. Combined with Table 7, it can be seen that several studies have already focused on the issue of automated scoring of PD patients. However, these studies still address gait-related tasks and motor delays [31–35]. Unfortunately, the tremor phenotypes of ET and PD are different, although the differential diagnosis between the two remains a topical issue [10,11,26,37–39]. Therefore, despite the relevance of such studies, we believe that the direct application of models or research patterns that automatically quantify PD severity [25] to ET diagnosis may lead to even inferior results. Thus, to our knowledge, this is the first study to incorporate a high-dimensional IMU sensor to automatically predict the five severity levels of the CRST scale for postural tremors using multiple machine learning classifiers guided by the entire clinical trial paradigm.

4.2. Feature analysis

Previous studies have demonstrated the effectiveness of time-domain and frequency-domain features for monitoring and typing movement disorders [7–10,15,23,37]. Therefore, commonly used statistical features are also extracted in this paper. In general, ma-

Table 7
Summary of research on wearable devices applied to neurodegenerative disorders.

Work (Year)	Subjects	Sensors	Data Collection Paradigm	Research Objectives	Algorithm	Optimal Results
Hoff et al. (2004) [53]	15 PD	ACC	24-hour activity monitoring	Distinguish on and off states (binary classes)	Experience-based threshold	SEN 71%, SPEC 76%
Salarian et al. (2007) [56]	10 PD, 10 HC	GYRO	45-minute laboratory tasks	Distinguish PD and HC	Experience-based threshold	SEN 99.5%, SPEC 94.2%
Zwartjes et al. (2010) [48]	6 PD, 7 HC	ACC	8 days of ADL monitoring	Quantify PD severity (regression)	Experience-based threshold	Rest tremor: UPDRS scores with a correlation of 0.84
Rigas et al. (2012) [49]	18 PD, 5 HC	ACC	5 days of ADL monitoring	Quantify PD severity (4 classes)	Hidden Markov Model	ACC 87%, SEN 84.5%, SPEC 84.5%
Pulliam et al. (2017) [50]	13 PD	ACC, GYRO	6 days of ADL monitoring	Tremor detection (binary classes)	Experience-based threshold	SEN 90%, SPEC 80%, AUC 0.89
Rovini et al. (2018) [28]	30 HC, 30 IH, 30 PD	ACC, GYRO	2-minute laboratory tasks	Distinguish patients and HC, Quantify IH severity (3 classes)	SVM, Random forest, NB	Quantify IH severity: ACC 78%
Khodakarami et al. (2019) [51]	172 PD	ACC	6 days of ADL monitoring	Identification of suitable DAT candidates (binary classes)	Logistic regression, SVM	SEN 89%, SPEC 86.6%
Butt et al. (2020) [29]	64 PD, 50 HC	ACC, GYRO	2.5-minute laboratory tasks	Quantify PD severity (regression)	SVM, Adaptive Neuro-Fuzzy Inference System, Random forest, LR	ACC 81.4%, RMSE 0.101
Locatelli et al. (2020) [26]	17 PD, 7 ET	ACC, GYRO	15-minute laboratory tasks	Distinguish PD and ET	SVM, KNN, NB, DA, DT	ACC 89.5%, SEN 91.7%, SPEC 85.7%
Mahadevan et al. (2020) [30]	60 HC, 35 PD	ACC, GYRO, MAG	45-minute of simulated ADL monitoring	Distinguish PD and HC	Random forest	ACC 83%, SEN 86%, PRE 86%, F1 83%
Sigcha et al. (2021) [25]	18 PD	ACC	5.8 weeks of ADL monitoring	Quantify PD severity (3 classes)	LOSO, SVM, Ensemble, CNN	SEN 86.1%, SPEC 86.1%, AUC 0.936
Proposed	**98 ET**	**ACC, GYRO, MAG**	**2-minute laboratory tasks**	**Quantify ET severity in three postural tasks (5 classes)**	**SVM, KNN, Ensemble, NB, DA**	**ACC 99.6%, SEN 95.7%, SPEC 99.4%, PRE 99.1%, F1 97.3%, AUC 0.936**

Notes: the length of data collection for individual studies was calculated using the average time spent. ACC, Accelerometer; GYRO, Gyroscope; MAG, Magnetometer; DA, Discriminant analysis, Activities of Daily Living; DAT, Device-assisted therapies; IH, Idiopathic hyposmia; UPDRS, Unified PD Rating Scale.

chine learning models are data-driven, and classification performance often depends on meaningful feature design and feature selection methods. In this study, features with statistical differences in ET patients before and after MRgFUS surgery are summarized in Table 4. These features contain time-domain, frequency-domain, and nonlinear parameters of multi-sensory fusion modalities and characterize the postural tremor's severity. This result is consistent with the study by Butt et al. [28,29].

Further, this study trained five machine learning models using all features and a subset of features constructed by multiple feature filtering methods for automatic quantitative scoring of three postural tremor tasks, respectively. It was found that selecting different feature subsets for the various postural tremor tasks may help improve the performance of the models. For example, only 20% of the original features were extracted using the Wrapper method in the arm extension task. However, the mean value of AUC improved from 99.1% to 99.8% (Fig. 5(b) and (e)). As shown in Table 6, the sensing information of the three modalities seems essential to present the redundant information of the original feature set and enhance the classification results with minimal computational cost. However, it is possible that PCA, which maps features to high dimensions, should be chosen with caution, despite its ability to explain most of the variance information through a few features.

4.3. Classifier performance

In this study, a comprehensive evaluation of multiple classifiers was performed using global performance metrics such as F1 score, AUC, etc., considering the balance between false alarm rate and accuracy in model decision making. In clinical practice, patients with more severe postural tremors tend to be fewer than those with CRST 1–2, as most severe tremors are treated with pharmacological or surgical interventions. Thus, a smaller CRST 3–4 tremor distribution is inevitable. It is noteworthy that more deficits or even zero recognition of tremor with higher tremor severity have been reported in other similar studies [8,9,40,54,55]. However, the machine learning classifier designed in this study has a generalized ability to discriminate between small samples of data (e.g., patients with higher severity).

As shown in the Results section, the ensemble learning method has the best classification performance. The bagging or boosting algorithm aggregates multiple decision tree models by averaging or optimizing the weights, providing better generalization capability than a single classifier. The Naïve Bayes classifier is a simple probabilistic classifier based on Bayes' theorem, which assumes that features are independent of each other. However, the observed variables, in reality, are consistently correlated with each other. Similar to the KNN algorithm, both are prone to overfitting in tasks

with unbalanced categories. In addition, with a high-dimensional feature space, the SVM classifier performs similarly to the first two because of the increased dimensionality of the sum function mapping. However, as shown by the feature selection experiments in Table 6, decreasing the number of features can improve the SVM classification very well. In addition, different pose tremor tasks combined with multiple feature subsets can also affect the selection of the final ensemble models. For example, the wing-beat position task may be suitable for using the Embedded method-based RUSBoost classifier, the arm extension task can use the Wrapper method-based AdaBoost model based on the Wrapper, while the finger-finger position task is best performed using the Bagging algorithm based on the Mutual information method. All these experimental results contribute to establishing an intelligent quantitative scoring system.

4.4. Limitations and future work

A review by Dr. Bhidayasiri, a neurologist, showed that postural tremor in ET patients usually affects the hands and is mostly bilateral and largely symmetrical [36]. In clinical practice, however, neurologists evaluate patients' hands separately. Therefore, our study focused on isolated scored products at the time of the visit, which may have overlooked the association between right- and left-hand tremor symptoms within individuals. Although the results of our statistical analysis of the characteristics suggest that the kinematic parameters of patients' hands do not differ significantly, future studies should design more practical features to mark this interindividual correlation. Furthermore, the research paradigm of using machine learning methods to analyze fine-grained tremor information obtained from postural sensors to provide meaningful aided diagnostic results is equally applicable to assessing action tremors. However, subjects perform action tasks with mixed motor disturbances, and it is difficult to maintain a consistent posture for each individual. Thus, the collected data have high variability. Considering the specificity of action tasks, we believe that future work should design more specific features for these tasks, for example, calculating the average duration of each movement completion in finger-to-nose tasks or writing tasks, to characterize the severity of tremor more efficiently. We plan to conduct future work with higher quality, larger data sets, develop an end-to-end comprehensive tremor severity scoring system, and receive ongoing feedback from neurologists working in the Department of Neurology at the PLA General Hospital. We plan to conduct future work with higher quality, larger data sets, develop an end-to-end tremor severity scoring system, and receive ongoing feedback from neurologists working in the Department of Neurology at the PLA General Hospital.

Another limitation is that the data used in this study were collected in a standardized movement assessment, which is essential for the rigor of clinical diagnosis. As previously mentioned, models trained under standardized movement datasets do not always generalize successfully to data from ADL scenarios [15,16,18,21,22,57]. Indeed, longitudinal data collection in unsupervised scenarios in patients with movement disorder-like disorders can help track progression and precise treatment [57]. Future studies should consider how to guide patients to standardize using this device in their daily lives and explore its potential as a complementary diagnostic tool.

5. Conclusions

This paper presents an automated quantification method for postural tremors developed based on a wearable device. The study collects postural signals for three tasks from patients diagnosed with ET based on a rigorous experimental paradigm. It builds a database (including HD video, electronic case information, and high-dimensional postural sensing signals) that three neurosurgeons jointly and independently score. The study explored SVM, KNN, NB, decision tree, and ensemble learning methods to select the best hyperparameters through Bayesian optimization and grid search in the data analysis phase. These machine learning algorithms are validated using a five-fold crossover method to enhance their generalization capability. Further, we explore the characterization ability of a subset of parameters constructed using different feature selection methods for the pose tremor task and demonstrate the effectiveness of the extracted features.

Combining the experimental results, we suggest that the optimal models are Ensemble models, using the RUSBoost model based on Embedded method in Wing-beat position task, AdaBoost model based on Wrapper method in Arm extension task, and Finger-finger position task using the Bagging algorithm based on the Mutual information method. The accuracy of the three postural tremor tasks ranged from 97.25%−97.98% with an AUC of 0.980–0.997, with a small probability of classification error, which is the best accuracy we know for automatic scoring of ET symptoms.

In conclusion, the overall results obtained in this paper are superior compared to existing studies. However, research conducted in future developments (e.g., sample expansion and training improvements) will strengthen the results of this study, enhance the robustness of the model, and potentially further improve accuracy.

Author contributions

Chenbin Ma, Rui Zong and Zhengbo Zhang designed the project. Chenbin Ma and Ailing Li collected the data. Chenbin Ma and Peng Zhang analyzed and interpreted the recorded data and wrote the paper. Rui Zong, Longsheng Pan, Xuemei Li, and Chunyu Yin recruited the subjects as a neurologist, scored the recorded information using a clinical rating scale. Jiachen Wang and Jian Zhang evaluated the project's hardware. Zhengbo Zhang participated in the conception and reviewed the study.

Funding

This work was supported by Natural Science Foundation of China (6217012292), Beijing Municipal Science and Technology (Z181100001918023) and Big Data Research & Development Project of Chinese PLA General Hospital (2018MBD-08, 2018MBD-09).

Acknowledgements

The authors declare no conflict of interest.

References

[1] E.D. Louis, The Roles of Age and Aging in Essential Tremor: an Epidemiological Perspective, Neuroepidemiology 52 (2019) 111–118.
[2] E.D. Louis, R. Ottman, W.Allen A. Hauser, How common is the most common adult movement disorder? Estimates of the prevalence of essential tremor throughout the world, Movement Disord. 13 (1998) 5–10.
[3] J. Kassubek, Diagnostic procedures during the course of Parkinson's Disease, Basal Ganglia 4 (2014) 15–18.
[4] M.T. Hu, R. Butterworth, V. Kumar, J. Cooper, E. Jones, L. Catterall, Y. Ben-Shlomo, How common and what are the determinants of sub-optimal care for Parkinson's disease patients: the Milton Keynes community study, Parkinsonism Rel. Disord. 17 (2011) 177–181.
[5] H.H. Fernandez, D. Stamler, M.D. Davis, S.A. Factor, R.A. Hauser, J. Jimenez-Shahed, W.G. Ondo, L.F. Jarskog, S.W. Woods, D. Bega, Long-term safety and efficacy of deutetrabenazine for the treatment of tardive dyskinesia, J. Neurol., Neurosurg. Psychiatry 90 (2019) 1317–1323.
[6] C. Ossig, H. Reichmann, Treatment strategies in early and advanced Parkinson disease, Neurol. Clin. 33 (2015) 19–37.
[7] M.D. Hssayeni, J. Jimenez-Shahed, M.A. Burack, B. Ghoraani, Wearable Sensors for Estimation of Parkinsonian Tremor Severity during Free Body Movements, Sensors 19 (2019) 4215.

[8] H. Jeon, W. Lee, H. Park, H.J. Lee, S.K. Kim, H.B. Kim, B. Jeon, K.S. Park, Automatic Classification of Tremor Severity in Parkinson's Disease Using a Wearable Device, Sensors 17 (2017) 2067.
[9] C. Pulliam, S. Eichenseer, C. Goetz, O. Waln, C. Hunter, J. Jankovic, D. Vaillancourt, J. Giuffrida, D. Heldman, Continuous in-home monitoring of essential tremor, Parkinsonism Rel. Disord. 20 (2014) 37–40.
[10] J.D. Loaiza Duque, A.M. González-Vargas, A.J. Sánchez Egea, H.A. González Rojas, Using Machine Learning and Accelerometry Data for Differential Diagnosis of Parkinson's Disease and Essential Tremor, in: Workshop On Engineering Applications. Springer, Cham, Springer International Publishing, Cham, 2019, pp. 368–378.
[11] L.A. Sanchez-Perez, L.P. Sanchez-Fernandez, A. Shaout, J.M. Martinez-Hernandez, M.J. Alvarez-Noriega, Rest tremor quantification based on fuzzy inference systems and wearable sensors, Int. J. Med. Inform. 114 (2018) 6–17.
[12] F. Luft, S. Sharifi, W. Mugge, A.C. Schouten, L.J. Bour, A.-.F. van Rootselaar, P.H. Veltink, T. Heida, Deficits in tapping accuracy and variability in tremor patients, J. Neuroeng. Rehabil. 16 (2019) 54-54.
[13] B. Boroojerdi, R. Ghaffari, N. Mahadevan, M. Markowitz, K. Melton, B. Morey, C. Otoul, S. Patel, J. Phillips, E. Sen-Gupta, O. Stumpp, D. Tatla, D. Terricabras, K. Claes, J.A. Wright, N. Sheth, Clinical feasibility of a wearable, conformable sensor patch to monitor motor symptoms in Parkinson's disease, Parkinsonism Relat. Disord. 61 (2019) 70–76.
[14] M. Heijmans, J.G. Habets, C. Herff, J. Aarts, A. Stevens, M.L. Kuijf, P.L. Kubben, Monitoring Parkinson's disease symptoms during daily life: a feasibility study, NPJ Parkinson's Dis. 5 (2019) 1–6.
[15] R. San-Segundo, A. Zhang, A. Cebulla, S. Panev, G. Tabor, K. Stebbins, R.E. Massa, A. Whitford, F. de la Torre, J. Hodgins, Parkinson's Disease Tremor Detection in the Wild Using Wearable Accelerometers, Sensors 20 (2020) 5817.
[16] J.M. Fisher, N.Y. Hammerla, T. Ploetz, P. Andras, L. Rochester, R.W. Walker, Unsupervised home monitoring of Parkinson's disease motor symptoms using body-worn accelerometers, Parkinsonism Rel. Disord. 33 (2016) 44–50.
[17] A. Zhang, A. Cebulla, S. Panev, J. Hodgins, F. De la Torre, Weakly-supervised learning for Parkinson's disease tremor detection, in: 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, 2017, pp. 143–147.
[18] A. Zhang, F. De la Torre, J. Hodgins, Comparing laboratory and in-the-wild data for continuous Parkinson's Disease tremor detection, in: 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), IEEE, 2020, pp. 5436–5441.
[19] A. Zhan, M.A. Little, D.A. Harris, S.O. Abiola, E. Dorsey, S. Saria, A.a. Terzis, High frequency remote monitoring of Parkinson's disease via smartphone: platform overview and medication response detection, arXiv preprint, (2016).
[20] F. Lipsmeier, K.I. Taylor, T. Kilchenmann, D. Wolf, A. Scotland, J. Schjodt-Eriksen, W.Y. Cheng, I. Fernandez-Garcia, J. Siebourg-Polster, L. Jin, Evaluation of smartphone-based testing to generate exploratory outcome measures in a phase 1 Parkinson's disease clinical trial, Mov. Disord. 33 (2018) 1287–1297.
[21] A. Papadopoulos, K. Kyritsis, S. Bostanjopoulou, L. Klingelhoefer, R.K. Chaudhuri, A. Delopoulos, Multiple-instance learning for in-the-wild Parkinsonian tremor detection, in: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, 2019, pp. 6188–6191.
[22] A. Papadopoulos, K. Kyritsis, L. Klingelhoefer, S. Bostanjopoulou, K.R. Chaudhuri, A. Delopoulos, Detecting parkinsonian tremor from IMU data collected in-the-wild using deep multiple-instance learning, IEEE J. Biomed. Health Inf. 24 (2019) 2559–2569.
[23] N. Shawen, M.K. O'Brien, S. Venkatesan, L. Lonini, T. Simuni, J.L. Hamilton, R. Ghaffari, J.A. Rogers, A. Jayaraman, Role of data measurement characteristics in the accurate detection of Parkinson's disease symptoms using wearable sensors, J. Neuroeng. Rehabil. 17 (2020) 52.
[24] L. Lonini, A. Dai, N. Shawen, T. Simuni, C. Poon, L. Shimanovich, M. Daeschler, R. Ghaffari, J.A. Rogers, A. Jayaraman, Wearable sensors for Parkinson's disease: which data are worth collecting for training symptom detection models, npj Digital Med. 1 (2018) 64.
[25] L. Sigcha, I. Pavón, N. Costa, S. Costa, M. Gago, P. Arezes, J.M. López, G. De Arcas, Automatic Resting Tremor Assessment in Parkinson's Disease Using Smartwatches and Multitask Convolutional Neural Networks, Sensors 21 (2021).
[26] P. Locatelli, D. Alimonti, G. Traversi, V. Re, Classification of Essential Tremor and Parkinson's Tremor Based on a Low-Power Wearable Device, Electronics (Basel) 9 (2020).
[27] D. Lukšys, G. Jonaitis, J. Griškevičius, Quantitative Analysis of Parkinsonian Tremor in a Clinical Setting Using Inertial Measurement Units, Parkinsons Dis. (2018) 1683831 2018.
[28] E. Rovini, C. Maremmani, A. Moschetti, D. Esposito, F. Cavallo, Comparative Motor Pre-clinical Assessment in Parkinson's Disease Using Supervised Machine Learning Approaches, Ann. Biomed. Eng. 46 (2018) 2057–2068.
[29] A.H. Butt, E. Rovini, H. Fujita, C. Maremmani, F. Cavallo, Data-Driven Models for Objective Grading Improvement of Parkinson's Disease, Ann. Biomed. Eng. 48 (2020) 2976–2987.
[30] N. Mahadevan, C. Demanuele, H. Zhang, D. Volfson, B. Ho, M.K. Erb, S. Patel, Development of digital biomarkers for resting tremor and bradykinesia using a wrist-worn wearable device, NPJ Digital Med. 3 (2020) 1–12.
[31] A. Zhao, L. Qi, J. Li, J. Dong, H. Yu, A hybrid spatio-temporal model for detection and severity rating of Parkinson's disease from gait data, Neurocomputing 315 (2018) 1–8.
[32] M.D. Hssayeni, J. Jimenez-Shahed, M.A. Burack, B. Ghoraani, Symptom-based, Dual-channel LSTM Network for The Estimation of Unified Parkinson's Disease Rating Scale III, in: 2019 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI), 2019, pp. 1–4.
[33] S. Del Din, A. Godfrey, C. Mazzà, S. Lord, L. Rochester, Free-living monitoring of Parkinson's disease: lessons from the field, Mov. Disord. 31 (2016) 1293–1313.
[34] D.A. Heldman, D.A. Harris, T. Felong, K.L. Andrzejewski, E.R. Dorsey, J.P. Giuffrida, B. Goldberg, M.A. Burack, Telehealth Management of Parkinson's Disease Using Wearable Sensors: an Exploratory Study, Digital Biomarkers 1 (2017) 43–51.
[35] I. Cova, A. Priori, Diagnostic biomarkers for Parkinson's disease at a glance: where are we? J. Neural Transm. 125 (2018) 1417–1432.
[36] R. Bhidayasiri, Differential diagnosis of common tremor syndromes, Postgrad. Med. J. 81 (2005) 756–762.
[37] S. Moon, H.-J. Song, V.D. Sharma, K.E. Lyons, R. Pahwa, A.E. Akinwuntan, H. Devos, Classification of Parkinson's disease and essential tremor based on balance and gait characteristics from wearable motion sensors via machine learning techniques: a data-driven approach, J. Neuroeng. Rehabil. 17 (2020) 125.
[38] M. Delrobaei, S. Memar, M. Pieterman, T.W. Stratton, K. McIsaac, M. Jog, Towards remote monitoring of Parkinson's disease tremor using wearable motion capture systems, J. Neurol. Sci. 384 (2018) 38–45.
[39] Z. Galaz, J. Mekyska, V. Zvoncak, J. Mucha, T. Kiska, Z. Smekal, I. Eliasova, M. Mrackova, M. Kostalova, I. Rektorova, M. Faundez-Zanuy, J.B. Alonso-Hernandez, P. Gomez-Vilda, Changes in Phonation and Their Relations with Progress of Parkinson's Disease, Appl. Sci. 8 (2018) 2339.
[40] J.P. Giuffrida, D.E. Riley, B.N. Maddux, D.A. Heldman, Clinically deployable KinesiaTM technology for automated tremor assessment, Movement Disord. 24 (2009) 723–730.
[41] E.D. Louis, Essential tremor, J. New Engl. J. Med. 345 (2001) 887–891.
[42] P. Pierleoni, L. Palma, A. Belli, L. Pernini, A real-time system to aid clinical classification and quantification of tremor in Parkinson's disease, in: IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI), 2014, pp. 113–116.
[43] R.J. Elble, S.L. Pullman, J.Y. Matsumoto, J. Raethjen, G. Deuschl, R. Tintner, Tremor amplitude is logarithmically related to 4- and 5-point tremor rating scales, Brain 129 (2006) 2660–2666.
[44] B. Abhinaya, D. Charanya, K.P. Thanaraj, Feature extraction and selection of a combination of entropy features for real-time epilepsy detection, Int. J. Eng. Comput. Sci. 5 (2016) 16073–16078 2016.
[45] M.T. Angelillo, D. Impedovo, G. Pirlo, G. Vessio, Performance-driven handwriting task selection for Parkinson's disease classification, in: International Conference of the Italian Association for Artificial Intelligence, Springer, 2019, pp. 281–293.
[46] G. Hripcsak, A.S. Rothschild, Agreement, the F-Measure, and Reliability in Information Retrieval, J. Am. Med. Inform. Assoc. 12 (2005) 296–298.
[47] N.L. Keijsers, M.W. Horstink, S.C. Gielen, Ambulatory motor assessment in Parkinson's disease, Movement Disord. 21 (2006) 34–44.
[48] D.G. Zwartjes, T. Heida, J.P. Van Vugt, J.A. Geelen, P.H. Veltink, Ambulatory monitoring of activities and motor symptoms in Parkinson's disease, IEEE Trans. Biomed. Eng. 57 (2010) 2778–2786.
[49] G. Rigas, A.T. Tzallas, M.G. Tsipouras, P. Bougia, E.E. Tripoliti, D. Baga, D.I. Fotiadis, S.G. Tsouli, S. Konitsiotis, Assessment of Tremor Activity in the Parkinson's Disease Using a Set of Wearable Sensors, IEEE Trans. Inf. Technol. Biomed. 16 (2012) 478–487.
[50] C.L. Pulliam, D.A. Heldman, E.B. Brokaw, T.O. Mera, Z.K. Mari, M.A. Burack, Continuous assessment of levodopa response in Parkinson's disease using wearable motion sensors, IEEE Trans. Biomed. Eng. 65 (2017) 159–164.
[51] H. Khodakarami, P. Farzanehfar, M. Horne, The use of data from the Parkinson's KinetiGraph to identify potential candidates for device assisted therapies, Sensors 19 (2019) 2241.
[52] A. Santiago, J.W. Langston, R. Gandhy, R. Dhall, S. Brillman, L. Rees, C. Barlow, Qualitative evaluation of the personal kinetigraph tm movement recording system in a parkinson's clinic, J. Parkinsons. Dis. 9 (2019) 207–219.
[53] J. Hoff, V. Van Der Meer, J. Van Hilten, Accuracy of objective ambulatory accelerometry in detecting motor complications in patients with Parkinson disease, Clin. Neuropharmacol. 27 (2004) 53–57.
[54] M. Giuberti, G. Ferrari, L. Contin, V. Cimolin, C. Azzaro, G. Albani, A. Mauro, 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. 19 (2015) 803–814.
[55] F. Parisi, G. Ferrari, M. Giuberti, L. Contin, V. Cimolin, C. Azzaro, G. Albani, A. Mauro, Body-Sensor-Network-Based Kinematic Characterization and Comparative Outlook of UPDRS Scoring in Leg Agility, Sit-to-Stand, and Gait Tasks in Parkinson's Disease, IEEE J. Biomed. Health Inform. 19 (2015) 1777–1793.
[56] A. Salarian, H. Russmann, C. Wider, P.R. Burkhard, F.J. Vingerhoets, K. Aminian, Quantification of tremor and bradykinesia in Parkinson's disease using a novel ambulatory monitoring system, IEEE Trans. Biomed. Eng. 54 (2007) 313–322.
[57] I. Teshuva, I. Hillel, E. Gazit, N. Giladi, A. Mirelman, J.M. Hausdorff, Using wearables to assess bradykinesia and rigidity in patients with Parkinson's disease: a focused, narrative review of the literature, J. Neural Transm. 126 (2019) 699–710.