

Phenotypical differentiation of tremor using time series feature extraction and machine learning

Verena Häring¹, Veronika Selzam¹, Juan Francisco Martin-Rodriguez^{2, 3, 4}, Petra Schwingenschuh⁵, Gertrúd Tamás⁶, Jan Raethjen⁷, Steffen Paschen⁷, Franziska Goltz⁸, Eoin Mulroy⁹, Anna Latorre⁹, Pablo Mir^{2, 3, 10}, Rick C. Helmich⁸, Kailash P. Bhatia⁹, Jens Volkmann¹, Robert Peach^{1, 11}, Sebastian R. Schreglmann^{1*}

*Word limit**: 5969/6000

Abstract word limit: 400/400

References: No limit

Display items: 8

Abstract

The reliable differentiation of tremor disorders poses a significant challenge, largely depending on the subjective interpretation of subtle signs and symptoms. Given the absence of a universally accepted bio-marker, diagnostic differentiation between the most prevalent tremor disorders, Essential Tremor (ET) and tremor-dominant Parkinson's Disease (PD), frequently proves to be a non-trivial task. To address this, we employed massive time series feature extraction, a powerful tool to examine the entirety of mathematical descriptors of oscillating biological signals without imposing bias, in combination with machine-learning (ML). We applied this approach to accelerometer recordings from tremor patients to identify

the optimal recording conditions, processing, and analysis settings, to differentiate ET and PD.

We utilized hand accelerometer recordings from 370 patients (167 ET, 203 PD), clinically diagnosed at five academic centres specialising in movement disorders, comprising an exploratory (158 ET, 172 PD from London, Graz, Budapest, Kiel) and a validation dataset (9 ET, 31 PD from Nijmegen). Using 15 second recording segments from the more affected hand, we first extracted established, standardized tremor characteristics and assessed their cross-centre accuracy and validity. Second, we applied supervised ML to massive higher-order feature extraction of the same recordings to achieve optimal stratification and mechanistic exploration.

While classic tremor characteristics were unable to consistently differentiate between conditions across centres, the resulting best classifying feature combination validated successfully. In comparison to tremor-stability index (TSI), the best performing classic tremor characteristic, feature-based analysis provided better classification accuracy (81.8% vs. 70.4%), sensitivity (86.4% vs. 70.8%) and specificity (76.6% vs. 70.2%), substantially improving stratification between ET and PD tremor. Similarly, this approach allowed the differentiation of rest from posture recordings independent of tremor diagnosis, again outperforming TSI (classification accuracy 99.6% vs. 49.2%). The interpretation of identified features indicates fundamentally different dynamics in tremor generating circuits: while there is an interaction between several central oscillators in the generation of PD rest tremor, resulting in several discrete but stable signal states, signal characteristics point towards a singular pacemaker in ET.

This study highlights the limitations of current, established tremor metrics and establishes the use of feature-based machine learning as a powerful method to explore accelerometry-

derived movement characteristics. More importantly, it showcases the strength of the combination of hypothesis-free, data-driven analyses and a large, multi-centre dataset. The results generated are thus resistant to device-, centre- and clinician-dependent bias and establish a generalizable differentiation method, representing a relevant step towards big data analysis in tremor disorders.

Author affiliations:

1 Department of Neurology, University Hospital Würzburg, 97080 Würzburg, Germany

2 Unidad de Trastornos del Movimiento, Servicio de Neurología y Neurofisiología Clínica, Instituto de Biomedicina de Sevilla, Hospital Universitario Virgen del Rocío / CSIC / Universidad de Sevilla, 41013 Seville, Spain

3 Centro de Investigación Biomédica en Red sobre Enfermedades Neurodegenerativas (CIBERNED), Instituto de Salud Carlos III, Madrid, Spain.

4 Departamento de Psicología Experimental, Facultad de Psicología, Universidad de Sevilla, Seville, Spain.

5 Department of Neurology, Medical University of Graz, 8036 Graz, Austria

6 Department of Neurology, Semmelweis University, 1083 Budapest, Hungary

7 Department of Neurology, Christian Albrechts University, 24118 Kiel, Germany

8 Department of Neurology, Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Centre, Nijmegen, the Netherlands

9 Department of Clinical and Movement Neurosciences, UCL Queen Square Institute of Neurology, London WC1N 3BG, UK

10 Departamento de Medicina, Facultad de Medicina, Universidad de Sevilla, Seville, Spain

11 Department of Brain Sciences, Imperial College London, London SW7 2AZ, UK

Correspondence to: Sebastian Schreglmann, MD PhD FEBN

Full address: Klinik und Poliklinik Neurologie, Univeristätsklinikum Würzburg, 97080

Würzburg, Germany

E-mail: sebastian.schreglmann@gmail.com

Running title: Feature-based Tremor Differentiation

Keywords: tremor analysis; Essential Tremor; Parkinson's disease; big data;

Abbreviations: AUC = area under the curve; EMG = electromyography; ET = Essential Tremor; FWHM = full-width half maximum; hctsa = highly comparative time-series analysis; HWP = Half width power; kNN = k-Nearest Neighbors; ML = machine learning; NaN = not a number; PCA = principal component analysis; PD = Parkinson's Disease; SVM = linear Support Vector Machine; TSI = Tremor Stability Index; UPDRS = Unified Parkinson's disease Rating Scale;

Introduction

Tremor stands as the most frequent movement disorder ¹, its involuntary, rhythmic, and oscillating nature predominantly affecting the hands and significantly impairing daily activities and quality of life ². It is a frequent symptom of various neurological conditions, with Essential Tremor (ET) and Parkinson's Disease (PD) being the most common, underlying aetiologies. ET, primarily a disorder of the cerebello-thalamo-cortical network, is

largely believed to originate in the cerebellum³⁻⁵, and is typically characterized by an isolated, bilateral postural or kinetic hand tremor⁶. Furthermore, a considerable number of ET patients also have rest tremor^{7,8}. Conversely, PD is a progressive neurodegenerative disease characterized by bradykinesia, rigidity, and a tremor that typically occurs during rest, but in many patients also re-emerges during posturing or sometimes actions⁹⁻¹². The pathophysiology of PD tremor is thought to involve both the basal ganglia, where tremor may be generated, and the cerebello-thalamo-cortical circuit, where the tremor is amplified, summarized in the “dimmer-switch” model^{13,14}.

Historically, clinical descriptive studies have established classic tremor phenotypes and classification systems for both disorders, thereby progressively shaping diagnostic criteria¹⁵⁻¹⁹. However, despite these established text-book definitions, significant phenotypical and pathophysiological overlap exists²⁰, complicating clinical differentiation and leading to misdiagnoses and mistreatment in over 30% of cases^{21,22}. In response, inertial measurement units, particularly accelerometers due to their affordability, practicality, and wide availability^{23,24}, have been increasingly employed to introduce objective measures into tremor evaluation²³, and have become standard diagnostic practice in many centres^{25-27, 17,18}.

To date, two validated sensor-based methods to differentiate ET from PD tremor have emerged: (1) the tremor stability index (TSI) and (2) the mean harmonic power. The TSI, a metric of the relative range of frequency tolerance of a tremor, was established to compare PD *rest* tremor vs. ET *postural* tremor recordings, offering best differentiation at a TSI cut-off of 1.05²⁸. However, absolute TSI values have been found to differ in other cohorts^{29,30} casting doubt on its generalizability. The mean harmonic power³¹, a metric quantifying the power at higher harmonic frequencies of the main tremor frequency, was established based on postural recordings and validated in three independent cohorts^{28,32}. However, it has been limited in its wider application due to reliance on specifically calibrated accelerometers.

Despite these advancements, established tremor metrics are unable to consistently differentiate tremor disorders^{25,33} and a general scarcity of comparative multi-centre tremor studies further impacts the generalizability of findings. Concurrently, the use of inertial measurement devices for tremor quantification is on the rise³⁴, along with the growth in available data sets²⁶ emphasising the need for better approaches to differentiate tremors.

Machine learning (ML) describes the process of identifying patterns from data without human interference. This analytical concept is transforming medical research³⁵⁻³⁷, allowing the integration and interrogation of complex data-sets by automatic pattern-recognition²⁹⁻³¹. ML allows the detection of previously un-noticed, unifying, generalizable and consistent disease characteristics. In the context of accelerometric tremor measurements, numerous studies have employed ML, utilising diverse accelerometers, recording positions, algorithms, and clinical settings³³. However, only a minority have applied ML to the specific challenge of distinguishing between different tremor disorders³⁸⁻⁴². Moreover, issues such as limited sample sizes, monocentric study designs, and the constraints of hypothesis-driven approaches have hindered the identification of reliable, generalizable disease-specific tremor movement characteristics³³.

Our study seeks to harness the potential of ML to uncover generalisable tremor characteristics that differentiate PD and ET patients. To capture the full spectrum of phenotypic inter-individual variability and to mitigate biases related to specific centres or devices, we combined recordings of individuals from various centres into a comprehensive, multi-centre dataset. We first explored the generalizability of established tremor characteristics. Next, by combining massive feature extraction⁴³ with ML⁴⁴, we identified previously unrecognized disease- and state-specific movement characteristics. Altogether, this method is advancing the accurate classification between PD and ET and represents a novel tool for hypothesis-free, unbiased time-series analysis.

Materials and methods

Participants and Ethics Approval

We analysed accelerometry recordings from patients with clinically established diagnoses of PD (n=203) or ET (n=167). These patients were recruited at five academic neurology institutes, each with dedicated movement disorder expertise (Graz; Budapest; London; Kiel; Nijmegen), some of which had been published previously^{31,45,46}. Clinical and demographic details were available from all centres, with the exception of the historical cohort from Kiel, and are summarized in Table 1. All patients were diagnosed by local movement disorder specialists based on at least two independent clinical consultations according to the Consensus Statement on Classification of Tremor^{17,18}. The established clinical diagnosis was used as reference throughout this work. All participants provided written informed consent according to local ethics at each participating centre prior to inclusion. The overall analysis was approved by the local research ethics committee (IRB University of Würzburg, Nr. 20210209 03) in accordance with the Declaration of Helsinki.

Accelerometer examination

Across sites, accelerometer recordings were performed after an overnight withdrawal of anti-parkinsonian or anti-tremor medication. Patients were seated with their forearms fully supported to negate the effects of gravity. All patients were recorded during rest and posture bilaterally and although all recordings adhered to local practices and standard operating procedures, as described previously^{28,45,46}, slight variations existed between centres in recording positions and equipment (see Table 2). Accelerometers were attached to the

dorsum of the proximal phalanx of the index (Graz, Budapest), middle finger (London), and dorsum of the hand (Kiel, Nijmegen). The recordings were tri-axial in most centres (Graz, Budapest, London, Nijmegen) and mono-axial in Kiel. Recording positions for rest were done with the hand hanging freely from the arm rest (Graz, Budapest, Kiel, Nijmegen) or the hand resting on a stable surface on its ulnar edge (London). For postural recordings, patients stretched their arms/wrists at shoulder level (Graz, Budapest, London, Nijmegen) or extended their hand while having their arms rested on the armrest (Kiel). To eliminate the influence of enhanced physiological tremor, recordings exhibiting a change in tremor frequency $>1\text{Hz}$ upon loading with 500g/1000g attached to the wrist were excluded^{27,47}.

The resulting inhomogeneity in recording positions served our primary objective: to discern disease- and posture-specific tremor characteristics from the time-series data, independent of centre-, equipment-, and position-related variables. Due to the unavailability of concurrent electromyography (EMG) recordings from every centre, EMG data could not be included in this analysis.

Data Analysis

Raw accelerometer data preparation

Primary data was converted from their respective original format to a universal format for analysis in Matlab (R2021b; Mathworks, USA). All time-series raw data and its conversion in frequency and time domain (Figure 1A) were individually visualized and screened for artefacts. For tri-axial accelerometer data, the vector amplitude sum across the three available axes was calculated and used for all subsequent analyses.

$$A = \sqrt{x^2 + y^2 + z^2} \quad (1)$$

All data was normalized to a mean of 0 and standard deviation of 1, down sampled to 100Hz, standardized to a recording length of 15 seconds (at least five seconds after the onset and before the end of tremor activity to prevent artefacts) and bandpass-filtered between 2 to 30 Hz [*third-order Butterworth filter; butter and filtfilt routines in Matlab*].

Detection of “more affected hand”

To further standardize analyses across data sets, and consistently use tremor recordings from the same hand for each participant throughout all analyses, we established a procedure to detect the more-affected hand, i.e. with larger tremor amplitude. Using the Area under the Curve (AUC) as a proxy for tremor power in both rest and postural position, we calculated the relative side differences in AUC for each tremor evoking position. The position with the largest relative AUC difference between sides was then used to detect the more affected hand, which was defined as the side with the larger relative AUC difference:

$$AUC_{rel} = \frac{AUC_{max_right} - AUC_{max_left}}{AUC_{max_right} + AUC_{max_left}} \quad (2)$$

Standard Tremor Characteristics

Sensor-based tremor characterization so far has predominantly focused on the frequency-domain. We thus calculated the following descriptive standard tremor characteristics across recording positions, patient groups and institutes, as previously described^{28,30}: area under the curve (AUC), Tremor Stability Index (TSI), Half width power (HWP), peak frequency, full-width half maximum (FWHM), and peak power. As accelerometers and calibrations differed across centres, mean harmonic power could not be calculated/included in this work.

Feature-based machine-learning

To increase the scope of analytical power, we next applied an approach not limited by pre-determined characteristics, but explored the entirety of mathematical descriptors of the included $n=370$ time-series data, as previously established⁴⁴. In short, we used the highly comparative time-series analysis (hctsa) algorithm^{43,48} to compute all possible features including autocorrelations, power spectra, wavelet decompositions, distributions, time-series models (e.g. Gaussian Processes, Hidden Markov model, autoregressive models), information-theoretic quantities (e.g. Sample Entropy, permutation entropy), non-linear measures (e.g. fractal scaling properties, nonlinear prediction errors) etc. for each time-series. This resulted in a 370×7873 feature matrix. After removing features with infinity, not a number (NaN) values or zero variance across the dataset this resulted in a reduced feature matrix of 370×7729 . In order to remove the impact of variance in feature scale, the value of each feature was individually normalised to the interval $[0,1]$ before using the feature matrix for classification.

Model evaluation

To compare the classification accuracy of different supervised ML algorithms (TreeBagger, k-Nearest Neighbors (kNN), linear Support Vector Machine (SVM), Neural Network, and Linear Regression), data was split into train and test partitions and the balanced classification accuracy computed for each algorithm.

$$Accuracy_{classification} = \frac{\text{number of accurate classifications}}{\text{number of all classifications}} \quad (3)$$

This was validated by a tenfold cross-validation, ensuring that recordings from each patient were consistently assigned to the same fold. To identify the most relevant features, we first applied a univariate approach, where features were evaluated based on their individual

classification accuracies. Features exhibiting strong discriminatory power were selected for further analysis. To optimize any classification process, we next systematically tested for the influence of recording position, centre, and accelerometer axes, before testing combinations of features. After detecting relevant differences between mono- and tri-axial recordings, we optimized the classification process in all tri-axial recordings from an explorative dataset (Graz, Budapest, London) and applied this on a separate validation dataset (Nijmegen), without cross-contamination.

Feature Combinations

We next assessed in how far a combination of univariate features would further improve the classification accuracy. To this end, we identified features with $>75\%$ univariate classification accuracy from the exploration data set. Univariate classification accuracies were calculated across all recordings and features selected based on performance. Next, all possible combinations of two, three and four of the identified features with $>75\%$ univariate classification accuracy were assessed for their combined classification accuracy, and arranged by performance. This was done with normalized as well as non-normalized time-series data to find features dependent as well as independent of tremor amplitude.

Comparison of recording position

In order to explore if our analysis pipeline could investigate features independent of tremor diagnosis, we determined generalizable, position-specific tremor features, that can differentiate between rest and postural recordings irrespective of tremor condition. To this end, we analysed the exploratory data set irrespective of diagnosis for differences between rest and postural recordings. Features were calculated using hctsa and the univariate

classification accuracies determined. For visualization purposes the 10 best performing features were compressed into two dimensions using the Matlab implemented function *t-SNE*.

Statistical analysis

After determining data distribution using a Kolmogorov–Smirnov Test, clinical and accelerometry data were compared using Fisher's exact test and Wilcoxon rank sum test according to the normality distribution of data. An n-way ANOVA was applied to explore for significant differences in tremor characteristics across recording centres and diseases. All tests were two tailed, and the α level was set at $p < 0.05$. All p values were corrected according to Bonferroni.

We used the Matlab implemented functions [*kstest2()*, *fishertest()* and *ranksum()*] as well as the statistical package JASP (JASP team 2023) for ANOVA testings.

Visualisation of multidimensional feature-space: To facilitate visualisation, principal component analysis (PCA) was performed by computing a covariance matrix for the normalised set of features. A principal component was constructed from eigenvectors and eigenvalues and displayed as a 2D scatter plot.

Data availability

Raw data is available from individual contributing centres (according to data sharing arrangements governed by patient consent) by reasonable request to the corresponding author.

Results

Patient Cohorts

In this study, we analyzed data and accelerometer recordings from a total of 370 tremor patients, comprising 167 diagnosed with ET and 203 with PD. The recordings originated from different centres with detailed demographic information (Graz (17ET, 21PD)⁴⁶, Budapest (36ET, 47PD), London (6ET, 5PD), Nijmegen (9ET, 31PD) and a historical sample from Kiel (99ET, 99PD). None of the patients was clinically classified as ET plus, as included recordings had been collected prior to the publication of the definition of ET plus.

Among the 172 patients for whom demographic information was available, ET and PD patients were matched for age and sex. However, notable differences were observed in the age at tremor onset and mean disease duration. The onset of tremor occurred at a younger age in ET (42.72 ± 23.0 years) compared to PD patients (57.2 ± 12.0 ; $P=0.4 \times 10^{-6}$) and the mean disease duration was longer in ET (20.6 ± 19.3 years) than PD patients (6.1 ± 4.8 ; $P=2.1 \times 10^{-11}$).

Comparison of Standard Tremor Characteristics

We first assessed established tremor characteristics (AUC, TSI, HWP, peak frequency, FWHM, peak power) as baseline measures to benchmark differentiation accuracy. To do this, we compared tremor characteristics within and between diagnoses and across centres for rest and postural recordings (Supplementary Tables S1 & S2). Specifically, we evaluated the consistency and generalizability of these measures in differentiating ET and PD.

Our analysis revealed that rest recordings generally yielded more pronounced differences between ET and PD (Figure 2). The tremor AUC failed to consistently differentiate between

the conditions, with substantial variability in measures among ET patients from different centres and contradictory trends in the two centres that showed significant differences (Figure 2A). The TSI demonstrated significant differentiation across centres overall, yet this pattern was not uniformly observed in individual centre data. Additionally, considerable variations were observed in the TSI values among ET patients from different centres (Figure 2B). The HWP achieved significant differentiation on an overall basis, but this distinction was not evident when analysing data from individual centres (Figure 2C). Peak frequency exhibited significant variability among ET patients across centres, with disease differentiation only apparent in the Kiel cohort (Figure 2D). The FWHM did not show significant differences either per disease or across centres (Figure 2E). Similar to TSI, peak power achieved differentiation between diseases across centres as a whole, yet this was only observable in two out of five centres individually (Figure 2F).

Taken together, TSI and Peak Power provided overall differentiation between diseases, resulting in individual classification accuracies of 70.4% (TSI) and 68.9% (peak power). By combining both characteristics, classification accuracy marginally improved to 70.5%, resulting in a specificity of 80.8% and a sensitivity of 57.6% to detect a difference between ET and PD (Table 3). However, it is crucial to note that none of the standard characteristics provided a reliable and generalizable means of differentiation within this dataset.

Differentiation using feature-based machine-learning

Optimizing Recording Settings

To overcome the limitations inherent in targeted analyses, we performed a comprehensive exploration using the full spectrum of 7729 time-series features identified by hctsa using

machine learning. This approach aimed to determine the optimal parameters for supervised machine learning, including the choice of algorithms, recording positions, sensor dimensionality (number of sensor axes; Figure 3A-C), and the number of top-performing features combined for classification accuracy (Figure 3A-F). This evaluation was based on one mono-centric dataset with triaxial recordings (Graz) to mitigate any centre-specific biases in recording SOPs.

We found that linear classifiers (SVM up to 86.1%, logistic regression classifier up to 82.8% classification accuracy) outperformed tree-based approaches such as Random Forest, particularly when integrating multiple features (Figure 3A). Furthermore, as above with classic tremor characteristics, we observed that recordings taken at rest consistently yielded higher classification accuracies (above 84%) compared to postural or combined recordings, irrespective of the number of features (Figure 3B).

In assessing the influence of the number of accelerometer axes on classification accuracy, we contrasted mono-axial recordings (Kiel) against both isolated single/z-axis time-series and vector amplitude sums from the same tri-axial recordings (Graz). Our findings indicated that the vector amplitude sum of triaxial accelerometer data outperformed single-axis data by approximately 13% (max. accuracy of 86.1% vs. 73.8%) and mono-axial recordings by 16% (86.1% vs. 70.0%).

Finally, we found that the combination of several best-performing features improved classification accuracy, while there was no substantial improvement beyond ten features (Supplementary Fig. 1). Taken together, we have selected SVM and vector amplitude sum generated from three axes as the ideal recording combination and tested for combinations of up to ten features in all subsequent analyses.

Comparison between tremor conditions

We next compared tremor recordings between ET and PD within the exploratory cohort, with a focus on how recording position influenced classification accuracy at each centre. Adopting the approach utilized in establishing the TSI²⁸, we first compared ET *postural* vs. PD *rest* recordings (Figure 3D). This approach, despite its inherent bias of contrasting different aetiology-specific postures, yielded classification accuracies as high as 90.4% for London, 89.0% for Graz, 81.1% for Budapest and 77.1% for Kiel. To address this bias, we subsequently compared ET versus PD postural recordings (Figure 3E), as well as ET versus PD rest recordings (Figure 3F). The latter comparison provided the most consistent results and was unaffected by the number of features included in the analysis.

Moreover, our findings indicated notable centre-specific variations in these analyses. For instance, recordings from Graz and London consistently demonstrated higher classification accuracies. Specifically, differentiation based on rest recordings achieved accuracies of up to 86.3% (Graz) and 84.6% (London), compared to 77.6% (Budapest) and 75.6% (Kiel). Overall, it was evident that rest tremor recordings provided more effective differentiation than postural recordings.

Identification of general classifiers to differentiate ET from PD tremor

We then asked whether we could identify features that reliably distinguish between ET and PD tremor signals across different centres. To do this, we analysed all available triaxial rest recordings (exploratory cohort consisting of 59 ET, 73PD) for both non-normalized as well as normalized time-series. This resulted in nine amplitude-dependent and four amplitude-independent features (Supplementary Table S3), each achieving a univariate classification accuracy of at least 75% for differentiating ET from PD tremor signals. Among these, four features were overlapping between the two categories (i.e., they appeared as top features for

both the non-normalized and normalized time series) suggesting that they are independent from amplitude. Of particular note, *MF_GARCHfit* emerged as the individual feature with the highest univariate classification accuracy.

To determine the best performing combination of features across centres, we next tested all possible combinations of two, three, and four of the 13 features identified for their combined classification accuracy on all exploratory rest recordings. The paired combination of *MF_GARCHfit* and *MF_hmm_CompareNStates* emerged as the optimal choice, featuring one non-amplitude dependent and one amplitude-dependent feature. This duo yielded a combined classification accuracy of 81.8%. Importantly, the addition of more than two features did not further improve overall performance (Fig. 4A). We evaluated the performance of these two features individually and in combination across each centre - the combined use of both features consistently outperformed the use of each feature alone (Fig. 4B). Notably, *MF_hmm_CompareNStates* exhibited lower accuracy in the London dataset, but this did not significantly affect the overall combined accuracy.

We also observed that these features, which were identified from tri-axial recordings across three institutes, showed more than 25% lower classification accuracy when applied to the mono-axial recordings from Kiel. This observation is consistent with our earlier findings from the standard tremor characteristic analyses, which similarly indicated diminished accuracy in mono-axial recordings.

To independently validate our model, we used the validation data set from Nijmegen (9 ET, 31 PD) and tested it with the two identified features: individually, *MF_GARCHfit* achieved 70% accuracy and *MF_hmm_CompareNStates* achieved 65% accuracy, which resulted in a combined classification accuracy of 72.5%.

Detection of Position-Specific Tremor Features

Next, we investigated whether our feature-based ML approach could identify position-specific tremor features irrespective of tremor diagnosis. For this we compared 132 rest and 132 postural recordings from all triaxial ET and PD recordings from our exploratory cohort and calculated test metrics for established tremor characteristics as baseline. In comparison to the found poor accuracies (Supplementary Table S4), ML-based feature differentiation identified several features with excellent classification accuracy (Supplementary Table S5). Remarkably, three of these identified features achieved a perfect accuracy of 100%. Further underscoring the robustness of our approach, the combination of the ten best individually performing features yielded a classification accuracy of 99.6%, which successfully validated in the validation dataset (97.4%; Figure 4C).

Interpretation of Disease-defining Tremor Features

To gain a more mechanistic understanding of the defining differences between ET and PD tremor characteristics, we aimed to interpret the two identified features: *MF_GARCHfit* and *MF_hmm_CompareNState* by assessing the underlying mathematics and relating it to more intuitive accelerometry data and feature distributions (Fig. 5A-D).

We found that *MF_hmm_CompareNStates* performed well in the raw data only (not normalized time-series), suggesting that it is amplitude-dependent. This feature assesses the dynamic temporal evolution of a time-series signal in a multi-dimensional feature-space. Over time signal characteristics can remain relatively stable, i.e. within a stable ‘signal state’, or change between several ‘signal states’ (Fig. 5E). The calculation of this feature relies on a

hidden Markov model as measured using log-likelihood, that is fitted to the position of signal states, with the number of identified signal states $<$ or ≥ 2 states defining the feature value. The feature values of ET signals show a poor fit to ≥ 2 states, indicating a higher probability that the signal remains within a singular signal state throughout the recording. On the other hand, in PD the data falls more clearly into several, separable but stable, generative states.

In contrast, for *MF_GARCHfit*, its presence in both normalized and raw data suggests an independence from tremor amplitude. This feature correlated with the skewness of the data, as illustrated by histogram plots (Fig. 5A, B). Specifically, this feature's value increases with deviations from a Gaussian distribution, i.e. if the relative contributions of certain acceleration signals are asymmetrically distributed (Fig 5A). Such cases of higher signal skewness are particularly evident in PD tremor time-series (Fig. 5D). Theoretically, this asymmetry relates to signals formed by a combination of oscillations, one or more of which are characterized by a direction-specific acceleration, creating an asymmetric signal distribution (Fig. 5F). With higher feature values present in PD signals, we interpret this as a likely correlate of the more asymmetrical accelerations evident in PD rest tremor, e.g. either the pill-rolling phenomenon, or well-known sudden fluctuations in hand tremor movements.

Stability of Features

To assess in how far our approach is influenced by recording length, we examined the stability of the identified features and their dependence on the duration of the underlying time-series. For this, we calculated the classification accuracy of the best performing combined features *MF_GARCHfit* and *MF_hmm_CompareNStates* on increasingly shortened segments of the same time-series from our exploratory cohort (59ET, 73PD; Figure 5). Our results revealed that the classification accuracy was consistently higher for feature combinations with longer time-series data. Nonetheless, even with time-series as short as two seconds, a combined accuracy of 68% was achieved, and an accuracy above 70% was achieved from a recording length of ten seconds onwards.

Discussion

In this study, we investigated and compared established tremor characteristics, and applied massive time-series feature extraction to optimise and generalise the differentiation of ET and PD tremor accelerometer recordings. Leveraging a large, multi-centre dataset, and systematically selecting the ideal recording and analysis settings, we identified movement characteristics from time-series data that reliably replicated across centres and resulted in higher sensitivity and specificity than the best performing standard metrics TSI and peak power. Furthermore, we identified posture-specific tremor features independent of tremor diagnosis. Mechanistically, the interpretation of the identified features indicates that ET tremor accelerometer signals remains stable within an individual, singular signal state with a

symmetrical signal distribution, whereas PD tremor signals can adopt ≥ 2 stable signal states with a more asymmetrical signal distribution.

Taken together, this work establishes multi-centre feature-based machine learning as a powerful novel method for unbiased tremor analysis. The proposed method allows reliable, valid, and generalizable differentiation of the two most prevalent tremor disorders, as well as their mechanistic exploration.

Standard tremor characteristics provide no reliable differentiation between ET and PD tremor

This study represents the largest multi-centre tremor accelerometry investigation to date²⁸, surpassing the scale of previous research which predominantly comprised mono-centre studies^{39,41}. Our approach, combining data curation and harmonization, facilitates a nuanced interrogation of tremor disorder signals, overcoming the effects of site- or device-specific influences.

Our findings contribute significantly to understanding cross-centre variations in tremor characteristics, revealing notable differences in tremor movements among patients from different centres. While our study design does not allow us to pinpoint the exact causes of this heterogeneity – be it population, recruitment, or diagnostic bias - it arguably provides the most comprehensive overview of signal variability in ET and PD currently available. It is well-acknowledged that patient presentations at tertiary referral centres differ from those in general neurology practices, highlighting the need for future studies to encompass a broader spectrum of clinical settings.

A core observation from our data is that stratification attempts perform better with rest than with postural recordings. In fact, both TSI and peak power - the only standard metrics showing significant differences across combined data from different centres (but not consistently in each centre individually) - were found to be significant only at rest. Although not surprising – clinically ET typically does not manifest with rest tremor, except in the case of ET plus as per the latest diagnostic criteria¹⁸ – this observation was consistent across classic and ML-based classification approaches. Given that ET plus was not included as a clinical diagnosis, this study cannot provide insights into the signal characteristics of this recently established clinical diagnostic category beyond published first evidence of barely discernible signal differences⁴⁹.

However, while both TSI and peak power demonstrated significance in combined cohort analysis, this was not uniformly observed across all individual centres, a result that could not be attributed to data from a singular centre unduly influencing the analyses. Our results hence highlight that none of the previously established tremor characteristics generalize across populations and centres – necessitating the discovery of novel metrics.

Triaxial accelerometry is superior to mono-axial measurements

Historically, electrophysiological assessments of tremor have combined accelerometry with EMG to differentiate between the mechanical, mechanical-reflex and central component of tremor²⁷. However, over the past decade, significant advances in tremor analysis, including the development of TSI and mean harmonic peak power^{28,31}, have only utilized accelerometer data^{33,34}. While earlier studies predominantly relied on mono-axial sensors^{26,50,51}, reflecting the constraints of technology availability and cost, tri-axial accelerometers are increasingly employed^{28,33,45,46,52} and recommended²⁷.

Our results on the effect of accelerometer dimensionality show that tri-axial accelerometer data is superior to mono-axial data in classifying tremor disorders. Mono-axial data consistently performed worse across our analyses, which cannot be attributed to sample size or centre, as exemplified in the data from Kiel and Graz. Mono-axial data hence inherently reduces the information content of the three-dimensional tremor signal, which substantially affects down-stream signal analysis. It is worth noting that while mono- and tri-axial accelerometers have been reported to be equally effective in detecting physical activity levels in adolescents and children^{53,54}, studies suggest that tri-axial sensors perform better in adults and the elderly⁵⁵. In line with our results, which are based on 15 second time-series data, the differentiation between types of activity over shorter time periods have been shown to rely on more fine-grained tri-axial information⁵⁶.

Feature-based machine learning – an un-biased method for tremor analysis

The basic principle of ML is to “learn” patterns in complex data without human interference, offering a means to overcome limitations of classical studies, which often limit their focus to a narrow set of variables or confine their exploration to predefined, hypothesis-driven characteristics.

Our study firmly establishes feature-based ML as an unbiased, generalizable method for tremor analysis. Based on the entirety of mathematical features extracted from time-series data⁴⁴, this methodology incorporates the maximum available information content from high-fidelity movement recordings to achieve pattern identification and stratification. As previous efforts^{28,31}, feature-based ML is based on movement characteristics only, not taking additional clinical information into account, which is often of paramount importance in daily practice. We demonstrate that this technique nevertheless can effectively categorize tremor

data by disease entity, outperforming traditional metrics. Similarly, the excellent differentiation by recording position irrespective of the clinical diagnosis confirms the analytical power of this novel approach. Our findings on ET and PD, the most prevalent tremor disorders, are derived from a multi-centre dataset incorporating phenotypic variability, limiting the vulnerability towards centre-, device- and clinician-specific bias. Results show robust performance of the identified feature-set, outperforming standard tremor characteristics.

The ongoing revolution in artificial intelligence is transforming medicine at a rapid pace and ML has been applied in several smaller tremor studies before, as reviewed by De *et al.*³³. The true strength of ML in this context is however only realised when paired with large, ideally multi-centre datasets, representing the physiological signal variability within a condition. By doing so in this work, we have confidently minimised *signal bias* – by taking the entirety of possible signal features into account –, *data set bias* – by combining recordings from several academic centres –, and *spectrum bias* – by aiming to sample the feature space evenly by including recordings from patients of different disease severities.

While it is anticipated that the precision of ML-based analyses will increase with larger and more clinically diverse sample sets, future international collaboration in this area is highly recommended. Nonetheless, our work demonstrates that ML-based differentiation between tremor disorders is not only feasible and superior to previous metrics but can also aid to advance our understanding of the underlying conditions.

Deductions on Pathophysiology

The robust identification of two features describing the signal state stability (*MF_hmm_CompareNStates*) and asymmetry of signal distribution (*MF_GARCHfit*) as the

defining characteristics discerning ET from PD tremor are instructive of the underlying dynamics and origin of the two tremor disorders.

Parkinsonian rest tremor is known to show spontaneous fluctuations in frequency over time and tremor amplitude is generally not affected by these subtle variations^{57,58}. This relatively broad frequency-amplitude tolerance, contrasting with a much narrower tolerance in ET, is summarized in the concept of TSI²⁸, and has been deduced from the spontaneous evolution of frequency fluctuations as well as the observation of different amplitude modulation effects despite significant frequency entrainment by brain stimulation⁵⁷. At the same time, tremor amplitude in PD shows a “waxing and waning” quality, and spontaneous amplitude fluctuations are preceded by increased cerebral integration several seconds before tremor onset⁵⁹. The detection of \geq two relatively stable signal states in PD tremor time-series is strong indication of either several discrete oscillatory pacemakers, which lead the rest tremor circuitry in turn, or several dynamic activity states of the involved structures. The observation that the signal fluctuates between a discrete number of stable states points towards a dynamic interplay between involved nodes. Theoretically this could be due to the pacemaker activity undulating forth and back between structures, or alternatively effects of drifting phase alignments in the rhythmicity between core oscillatory pacemakers.

Similarly, the combination of several oscillatory signals resulting in an asymmetric signal distribution in PD tremor is incompatible with a singular pacemaker. It rather supports the current view of at least two central tremor oscillation pacemakers in PD, likely situated in the basal ganglia and the cerebello-thalamo-cortical motor loop⁶⁰. The currently prevailing theory on the origin of parkinsonian rest tremor, the *dimmer-switch model*, posits that the origin of tremor-related oscillations stems from the basal ganglia (globus pallidus and subthalamic nucleus), while the cerebello-thalamo-cortical loop modulates tremor amplitude¹⁴. Accordingly, oscillatory activity at tremor-frequency has been detected by microelectrode

recordings not only in the ventral intermediate thalamic (VIM) nucleus, but also subthalamic nucleus (STN) and the internal pallidum (GPi)⁶¹⁻⁶⁴. Further, different frequency bands within the same anatomical structure (STN) have been shown to have opposing effects on PD rest tremor, representing additional, functionally discrete central oscillators^{64,65}. Our data hence provides additional evidence to support an interaction between several central oscillators in the generation of PD rest tremor, resulting in several discrete but stable signal states.

Conversely, the presence of a singular stable signal space with a symmetrical signal distribution in our data is compatible with either a singular oscillator or a much more tightly connected net of oscillators in ET. Previously, the narrower TSI has been interpreted as a result of stronger coupling between the central elements involved in tremor generation^{28,57}, and although our data cannot provide the anatomical origin of such activity, it provides a strong and direct link between tremor phenomenology and underlying pathophysiology.

Context and Implications

Tremor disorders are diagnosed based on clinical assessment. Historically, classification systems have developed based on the clinician's interpretation of phenotypes^{16-18,66}. This classification framework is deeply rooted in the historical routine in neurology of deducing clinico-pathological correlation from the astute observation and interpretation of subtle movement characteristics⁶⁷. However, the diagnostic yield of this approach has been suboptimal, particularly regarding the diagnosis of ET, which has historically been applied in a less stringent way^{68,69}.

Recent efforts to refine these classifications have led to the re-conceptualization of ET as a clinical syndrome and the introduction of ET-plus¹⁸, acknowledging the diagnostic uncertainty prevalent in many tremor cases⁷⁰. Whilst this has been criticised for not being

based on a known pathology ⁷¹, this new classification, which is likely yet another intermediate step in the continuing evolution of the clinical concept ET ⁷², is taking clinical soft signs into consideration. As clinical soft signs in movement disorder research are known to have low clinimetric accuracy and poor interrater reliability ^{70,73–75}, however, there is a relevant clinical need for diagnostic improvement.

Despite these efforts, there remains no universally accepted biomarker to aid objective tremor classification, shrouding results from decades of research into the epidemiology, pathophysiology and genetics particularly of ET, in a cloud of ambiguity ^{34,76–79}. This uncertainty must be considered when interpreting the results of studies like ours that use the current clinical diagnostic gold standard as reference.

Notwithstanding the value of clinical phenotyping, our study lays the foundation for a more objective, generalizable approach to describing and characterizing tremor. This is particularly relevant, as stratification by movement characteristics has proven to have relevant therapeutic implications, exemplified by the results from phase-locked invasive and non-invasive brain stimulation in several tremor disorders that have shown that rhythmic tremor following a sinusoidal activity is more amenable to phase-specific intervention ^{44,80,81}. There is hence emerging evidence to support the notion that not the clinical diagnosis per se, but the presence of particular movement characteristics may predict the response to certain therapeutic interventions ^{69,82} – an emerging concept possibly not limited to the effect of electrical stimulation.

There is an interesting, recent debate, on whether the classic interpretation of phenotype in movement disorders should be used in a way that acknowledges the degeneracy of human movement control, i.e. the propensity of various pathophysiological entities/diseases to “break” in a limited number of phenotypes ⁶⁷. From this, authors argued that only the reliable

dynamic quantitative monitoring of movement is able to connect any specific network dysfunction with a particular phenotype. Feature-based (tremor) time-series analysis likely provides the descriptive power for any such attempts.

Several points for further improvement remain: ML approaches are known to be susceptible to incomplete sampling of the features space, e.g. by including not enough samples of common/rare variants of a disease or by failing to include populations of various genetic backgrounds. Certainly, models including larger data sets will achieve a more accurate performance and we acknowledge that the included data represent a population of predominantly European-Caucasian descent only. Outliers are another problem, as they tend to disproportionately influence ML results³³, and carefully curated datasets are paramount.

Conclusion

In summary, this study establishes feature-based ML as a novel analytical tool for unbiased tremor analysis, outperforming previous electrophysiology metrics for tremor stratification, and offering a means for hypothesis-free exploration of complex time-series signals. Future studies should aim to incorporate large, diverse samples from multiple centres and various genetic backgrounds in order to further optimize the generalizability and validity of feature-based ML tremor classification.

Acknowledgements

Verena Häring orcid: 0009-0006-7543-7540

Veronika Selzam orcid: 0009-0009-3209-9360

Juan Francisco Martin-Rodriguez orcid: 0000-0003-1392-8775

Petra Schwingenschuh orcid: 0000-0002-3294-6824

Gertrúd Tamás orcid: 0000-0001-8054-3678

Jan Raethjen orcid: 0000-0001-9812-1115

Steffen Paschen orcid: 0000-0002-9306-8332

Franziska Goltz orcid: [0000-0002-7146-2857](https://orcid.org/0000-0002-7146-2857)

Eoin Mulroy orcid: 0000-0003-3057-0592

Anna Latorre orcid : 0000-0001-7103-5209

Pablo Mir orcid: [0000-0003-1656-302X](https://orcid.org/0000-0003-1656-302X)

Rick Helmich orcid: 0000-0003-4035-6573

Kailash P. Bhatia, orcid: 0000-0001-8185-286X

Jens Volkmann orcid: 0000-0002-9570-593X

Robert Peach orcid: 0000-0002-8738-5825

Sebastian R. Schreglmann, orcid: 0000-0002-4129-5808

Funding

VH and VS received support from the Graduate School of Life Sciences at the University Wuerzburg. JFM and PMR are supported by Grants of the Spanish Ministry of Science and Innovation (RTC2019-007150-1), the Instituto de Salud Carlos III-Fondo Europeo de Desarrollo Regional (ISCIII-FEDER) (PI16/01575, PI18/01898, PI19/01576, PI22/01704), PID2021-127034OA-I00 funded by MCIN/AEI/ 10.13039/501100011033 by “ERDF A way

of making Europe”, the Fundación Progreso y Salud (PI-0055-2014), the Consejería de Economía, Innovación, Ciencia y Empleo de la Junta de Andalucía (CVI-02526, CTS-7685), the Consejería de Salud y Bienestar Social de la Junta de Andalucía (PI-0471-2013, PE-0210-2018, PI-0459-2018, PE-0186-2019), the Consejería de Transformacion Economica, Industria, Conocimiento y Universidades de la Junta de Andalucía (PY20_00903), and the Fundacion Alicia Koplowitz. PMR is a member of the COST Action IMMUPARKNET (CA21117). RP and SRS acknowledge the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) Project-ID 424778381-TRR 295 – SRS is a Fellow of the Thiemann Foundation. Open Access funding enabled and organized by Projekt DEAL.

Competing interests

The authors report no competing interests.

Supplementary material

Supplementary material is available at *Brain* online.

Appendix

References

1. Louis ED, Ferreira JJ. How common is the most common adult movement disorder? Update on the worldwide prevalence of essential tremor. *Mov Disord*. 2010;25(5):534-541. doi:10.1002/mds.22838

2. Chandran V, Pal PK. Quality of life and its determinants in essential tremor. *Park Relat Disord*. 2013;19(1):62-65. doi:10.1016/j.parkreldis.2012.06.011
3. Berg KRE van den, Helmich RC. The Role of the Cerebellum in Tremor – Evidence from Neuroimaging. *Tremor Other Hyperkinetic Movements*. 2021;11(1):49. doi:10.5334/tohm.660
4. Gallea C, Popa T, García-Lorenzo D, et al. Intrinsic signature of essential tremor in the cerebello-frontal network. *Brain*. 2015;138(Pt 10):2920-2933. doi:10.1093/brain/awv171
5. Muthuraman M, Raethjen J, Koirala N, et al. Cerebello-cortical network fingerprints differ between essential, Parkinson’s and mimicked tremors. *Brain*. 2018;141(6):1770-1781. doi:10.1093/brain/awy098
6. Welton T, Cardoso F, Carr JA, et al. Essential tremor. *Nat Rev Dis Primers*. 2021;7(1):83. doi:10.1038/s41572-021-00314-w
7. Erro R, Pilotto A, Magistrelli L, et al. A Bayesian approach to Essential Tremor plus: A preliminary analysis of the TITAN cohort. *Park Relat Disord*. 2022;103:73-76. doi:10.1016/j.parkreldis.2022.08.030
8. Louis ED, Hernandez N, Michalec M. Prevalence and correlates of rest tremor in essential tremor: cross-sectional survey of 831 patients across four distinct cohorts. *Eur J Neurol*. 2015;22(6):927-932. doi:10.1111/ene.12683
9. Hallett M. Parkinson’s disease tremor: pathophysiology. *Park Relat Disord*. 2012;18:S85-S86. doi:10.1016/s1353-8020(11)70027-x
10. Jankovic J, Schwartz KS, Ondo W. Re-emergent tremor of Parkinson’s disease. *J Neurology Neurosurg Psychiatry*. 1999;67(5):646. doi:10.1136/jnnp.67.5.646
11. Swinnen BEKS, Bie RMA de, Hallett M, Helmich RC, Buijink AWG. Reconstructing Re-emergent Tremor. *Mov Disord Clin Pr*. 2023;10(9):1293-1296. doi:10.1002/mdc3.13806
12. Dirkx MF, Zach H, Bloem BR, Hallett M, Helmich RC. The nature of postural tremor in Parkinson disease. *Neurology*. 2018;90(13):e1095-e1103. doi:10.1212/wnl.0000000000005215
13. Lauro PM, Lee S, Akbar U, Asaad WF. Subthalamic-Cortical Network Reorganization during Parkinson’s Tremor. *J Neurosci*: Off J Soc Neurosci. 2021;41(47):9844-9858. doi:10.1523/jneurosci.0854-21.2021
14. Helmich RC, Hallett M, Deuschl G, Toni I, Bloem BR. Cerebral causes and consequences of parkinsonian resting tremor: a tale of two circuits? *Brain*. 2012;135(11):3206-3226. doi:10.1093/brain/aws023
15. Marsden, Obeso, Rothwell. Benign essential tremor is not a single entity. In: Yahr, ed. *Current Concepts of Parkinson Disease and Related Disorders*. Excerpta Medica; 1983:31–46.

16. Findley, Koller. Handbook of Tremor Disorders. In: . Marcel Dekker; 1995.
17. Deuschl G, Bain P, Brin M. Consensus Statement of the Movement Disorder Society on Tremor. *Movement Disord*. 1998;13(S3):2-23. doi:10.1002/mds.870131303
18. Bhatia KP, Bain P, Bajaj N, et al. Consensus Statement on the classification of tremors. from the task force on tremor of the International Parkinson and Movement Disorder Society. *Mov Disord*. 2018;33(1):75-87. doi:10.1002/mds.27121
19. Elble R, Deuschl G. Milestones in tremor research. *Mov Disord*. 2011;26(6):1096-1105. doi:10.1002/mds.23579
20. Nieuwhof F, Panyakaew P, Warrenburg BP van de, Gallea C, Helmich RC. The patchy tremor landscape: recent advances in pathophysiology. *Curr Opin Neurol*. 2018;31(4):455-461. doi:10.1097/wco.0000000000000582
21. Jain S, Lo SE, Louis ED. Common Misdiagnosis of a Common Neurological Disorder: How Are We Misdiagnosing Essential Tremor? *Arch Neurol-chicago*. 2006;63(8):1100-1104. doi:10.1001/archneur.63.8.1100
22. Schrag A, Münchau A, Bhatia KP, Quinn NP, Marsden CD. Essential tremor: an overdiagnosed condition? *J Neurol*. 2000;247(12):955-959. doi:10.1007/s004150070053
23. Haubenberger D, Abbruzzese G, Bain PG, et al. Transducer-based evaluation of tremor. *Movement Disord*. 2016;31(9):1327-1336. doi:10.1002/mds.26671
24. Vescio B, Quattrone A, Nisticò R, Crasà M, Quattrone A. Wearable Devices for Assessment of Tremor. *Front Neurol*. 2021;12:680011. doi:10.3389/fneur.2021.680011
25. Veen S van der, Klamer MR, Elting JWJ, Koelman JHTM, Stouwe AMMV der, Tijssen MAJ. The diagnostic value of clinical neurophysiology in hyperkinetic movement disorders: a systematic review. *Parkinsonism Relat D*. Published online 2021. doi:10.1016/j.parkreldis.2021.07.033
26. Gironell A, Kulisevsky J, Pascual-Sedano B, Barbanoj M. Routine Neurophysiologic Tremor Analysis as a Diagnostic Tool for Essential Tremor: A Prospective Study. *J Clin Neurophysiol*. 2004;21(6):446-450. doi:10.1097/00004691-200411000-00009
27. Vial F, Kassavetis P, Merchant S, Haubenberger D, Hallett M. How to do an electrophysiological study of tremor. *Clin Neurophysiology Pract*. 2019;4:134-142. doi:10.1016/j.cnp.2019.06.002
28. Biase L di, Brittain JS, Shah SA, et al. Tremor stability index: a new tool for differential diagnosis in tremor syndromes. *Brain*. 2017;140(7):1977-1986. doi:10.1093/brain/awx104
29. Luft F, Sharifi S, Mugge W, et al. A Power Spectral Density-Based Method to Detect Tremor and Tremor Intermittency in Movement Disorders. *Sensors Basel Switz*. 2019;19(19):4301. doi:10.3390/s19194301

30. Panyakaew P, Cho HJ, Lee SW, Wu T, Hallett M. The Pathophysiology of Dystonic Tremors and Comparison With Essential Tremor. *J Neurosci Official J Soc Neurosci*. 2020;40(48):9317-9326. doi:10.1523/jneurosci.1181-20.2020
31. Muthuraman M, Hossen A, Heute U, Deuschl G, Raethjen J. A new diagnostic test to distinguish tremulous Parkinson's disease from advanced essential tremor. *Mov Disord*. 2011;26(8):1548-1552. doi:10.1002/mds.23672
32. Wile DJ, Ranawaya R, Kiss ZHT. Smart watch accelerometry for analysis and diagnosis of tremor. *Journal of neuroscience methods*. 2014;230:1-4. doi:10.1016/j.jneumeth.2014.04.021
33. De A, Bhatia KP, Volkmann J, Peach R, Schreglmann SR. Machine Learning in Tremor Analysis: Critique and Directions. *Mov Disord*. 2023;38(5):717-731. doi:10.1002/mds.29376
34. Erro R, Fasano A, Barone P, Bhatia KP. Milestones in Tremor Research: 10 Years Later. *Mov Disord Clin Pr*. 2022;9(4):429-435. doi:10.1002/mdc3.13418
35. Rajkomar A, Dean J, Kohane I. Machine Learning in Medicine. *New Engl J Medicine*. 2019;380(26):2589-2590. doi:10.1056/nejmc1906060
36. Auger SD, Jacobs BM, Dobson R, Marshall CR, Noyce AJ. Big data, machine learning and artificial intelligence: a neurologist's guide. *Pract Neurology*. 2021;21(1):4-11. doi:10.1136/practneurol-2020-002688
37. Aggarwal R, Sounderajah V, Martin G, et al. Diagnostic accuracy of deep learning in medical imaging: a systematic review and meta-analysis. *Npj Digital Medicine*. 2021;4(1):65. doi:10.1038/s41746-021-00438-z
38. Balachandar A, Algarni M, Oliveira L, et al. Are smartphones and machine learning enough to diagnose tremor? *J Neurol*. Published online 2022:1-12. doi:10.1007/s00415-022-11293-7
39. Xing X, Luo N, Li S, Zhou L, Song C, Liu J. Identification and Classification of Parkinsonian and Essential Tremors for Diagnosis Using Machine Learning Algorithms. *Front Neurosci-switz*. 2022;16:701632. doi:10.3389/fnins.2022.701632
40. Shahtalebi S, Atashzar SF, Patel RV, Jog MS, Mohammadi A. A deep explainable artificial intelligent framework for neurological disorders discrimination. *Sci Rep-uk*. 2021;11(1):9630. doi:10.1038/s41598-021-88919-9
41. Shahtalebi S, Atashzar SF, Samotus O, Patel RV, Jog MS, Mohammadi A. PHTNet: Characterization and Deep Mining of Involuntary Pathological Hand Tremor using Recurrent Neural Network Models. *Scientific reports*. 2020;10(1):2195. doi:10.1038/s41598-020-58912-9
42. Ghassemi NH, Marxreiter F, Pasluosta CF, et al. Combined Accelerometer and EMG Analysis to Differentiate Essential Tremor from Parkinson's Disease. *2016 38th Annu Int Conf Ieee Eng Medicine Biology Soc Embc*. 2016;2016:672-675. doi:10.1109/embc.2016.7590791

43. Fulcher BD, Jones NS. hctsa: A Computational Framework for Automated Time-Series Phenotyping Using Massive Feature Extraction. *Cell systems*. 2017;5(5):527-531.e3. doi:10.1016/j.cels.2017.10.001
44. Schreglmann SR, Wang D, Peach RL, et al. Non-invasive suppression of essential tremor via phase-locked disruption of its temporal coherence. *Nat Commun*. 2021;12(1):363. doi:10.1038/s41467-020-20581-7
45. Schwingenschuh P, Ruge D, Edwards MJ, et al. Distinguishing SWEDDs patients with asymmetric resting tremor from Parkinson's disease: A clinical and electrophysiological study. *Mov Disord*. 2010;25(5):560-569. doi:10.1002/mds.23019
46. Schwingenschuh P, Saifee TA, Katschnig P, Winter P, et al. Validation of "laboratory-supported" criteria for functional (psychogenic) tremor. *Movement Disord*. 2016;31(4):555-562. doi:10.1002/mds.26525
47. Stouwe AMM van der, Elting JW, Hoeven JH van der, et al. How typical are 'typical' tremor characteristics? Sensitivity and specificity of five tremor phenomena. *Parkinsonism Relat D*. 2016;30:23-28. doi:10.1016/j.parkreldis.2016.06.008
48. Fulcher BD, Little MA, Jones NS. Highly comparative time-series analysis: the empirical structure of time series and their methods. *Journal of the Royal Society, Interface*. 2013;10(83):20130048. doi:10.1098/rsif.2013.0048
49. Rajan R, Anandapadmanabhan R, Vishnoi A, et al. Essential tremor and essential tremor plus are essentially similar electrophysiologically. *Mov Disord Clin Pr*. Published online 2023. doi:10.1002/mdc3.13941
50. Deuschl G, Krack P, Lauk M, Timmer J. Clinical neurophysiology of tremor. *Journal of clinical neurophysiology*: official publication of the American Electroencephalographic Society. 1996;13(2):110-121.
51. Raethjen J, Lauk M, Köster B, et al. Tremor analysis in two normal cohorts. *Clin Neurophysiol*. 2004;115(9):2151-2156. doi:10.1016/j.clinph.2004.04.006
52. Schwingenschuh P, Katschnig P, Seiler S, et al. Moving toward "laboratory-supported" criteria for psychogenic tremor. *Movement Disord*. 2011;26(14):2509-2515. doi:10.1002/mds.23922
53. Vanhelst J, Béghin L, Duhamel A, Bergman P, Sjöström M, Gottrand F. Comparison of uniaxial and triaxial accelerometry in the assessment of physical activity among adolescents under free-living conditions: the HELENA study. *BMC Méd Res Methodol*. 2012;12(1):26. doi:10.1186/1471-2288-12-26
54. Hislop JF, Bulley C, Mercer TH, Reilly JJ. Comparison of epoch and uniaxial versus triaxial accelerometers in the measurement of physical activity in preschool children: a validation study. *Pediatr Exerc Sci*. 2012;24(3):450-460. doi:10.1123/pes.24.3.450

55. Sagelv EH, Ekelund U, Pedersen S, et al. Physical activity levels in adults and elderly from triaxial and uniaxial accelerometry. The Tromsø Study. *PLoS ONE*. 2019;14(12):e0225670. doi:10.1371/journal.pone.0225670
56. Smith MP, Horsch A, Standl M, Heinrich J, Schulz H. Uni- and triaxial accelerometric signals agree during daily routine, but show differences between sports. *Sci Rep*. 2018;8(1):15055. doi:10.1038/s41598-018-33288-z
57. Cagnan H, Little S, Foltynie T, et al. The nature of tremor circuits in parkinsonian and essential tremor. *Brain*. 2014;137(12):3223-3234. doi:10.1093/brain/awu250
58. Thanawattano C, Pongthornseri R, Anan C, Dumnin S, Bhidayasiri R. Temporal fluctuations of tremor signals from inertial sensor: a preliminary study in differentiating Parkinson's disease from essential tremor. *Biomed Eng OnLine*. 2015;14(1):101. doi:10.1186/s12938-015-0098-1
59. Dirkx MF, Shine JM, Helmich RC. Integrative Brain States Facilitate the Expression of Parkinson's Tremor. *Mov Disord*. 2023;38(9):1615-1624. doi:10.1002/mds.29506
60. Dirkx MF, Bologna M. The pathophysiology of Parkinson's disease tremor. *J Neurol Sci*. 2022;435:120196. doi:10.1016/j.jns.2022.120196
61. Magnin M, Morel A, Jeanmonod D. Single-unit analysis of the pallidum, thalamus and subthalamic nucleus in parkinsonian patients. *Neuroscience*. 2000;96(3):549-564. doi:10.1016/s0306-4522(99)00583-7
62. Moran A, Bergman H, Israel Z, Bar-Gad I. Subthalamic nucleus functional organization revealed by parkinsonian neuronal oscillations and synchrony. *Brain*. 2008;131(12):3395-3409. doi:10.1093/brain/awn270
63. Amtage F, Henschel K, Schelter B, et al. Tremor-correlated neuronal activity in the subthalamic nucleus of Parkinsonian patients. *Neurosci Lett*. 2008;442(3):195-199. doi:10.1016/j.neulet.2008.06.087
64. Asch N, Herschman Y, Maoz R, et al. Independently together: subthalamic theta and beta opposite roles in predicting Parkinson's tremor. *Brain Commun*. 2020;2(2):fcaa074. doi:10.1093/braincomms/fcaa074
65. Hirschmann J, Butz M, Hartmann CJ, et al. Parkinsonian Rest Tremor Is Associated With Modulations of Subthalamic High-Frequency Oscillations. *Mov Disord*. 2016;31(10):1551-1559. doi:10.1002/mds.26663
66. Chouinard S, Louis ED, Fahn S. Agreement among movement disorder specialists on the clinical diagnosis of essential tremor. *Movement Disord*. 1997;12(6):973-976. doi:10.1002/mds.870120621
67. Sadnicka A, Edwards MJ. Between Nothing and Everything: Phenomenology in Movement Disorders. *Mov Disord*. 2023;38(10):1767-1773. doi:10.1002/mds.29584

68. Amlang CJ, Diaz DT, Louis ED. Essential Tremor as a “Waste Basket” Diagnosis: Diagnosing Essential Tremor Remains a Challenge. *Front Neurol*. 2020;11:172. doi:10.3389/fneur.2020.00172
69. Erro R, Picillo M, Pellecchia MT, Barone P. Diagnosis Versus Classification of Essential Tremor: A Research Perspective. *J Mov Disord*. 2023;16(2):152-157. doi:10.14802/jmd.23020
70. Elble RJ. Bayesian Interpretation of Essential Tremor Plus. *J Clin Neurol (Seoul, Korea)*. 2022;18(2):127-139. doi:10.3988/jcn.2022.18.2.127
71. Louis ED, Bares M, Benito-Leon J, et al. Essential tremor-plus: a controversial new concept. *Lancet Neurology*. 2019;19(3):266-270. doi:10.1016/s1474-4422(19)30398-9
72. Vidailhet M. Essential tremor-plus: a temporary label. *Lancet Neurol*. 2020;19(3):202-203. doi:10.1016/s1474-4422(19)30442-9
73. Fearon C, Espay AJ, Lang AE, et al. Soft signs in movement disorders: friends or foes? *J Neurol, Neurosurg Psychiatry*. 2019;90(8):961. doi:10.1136/jnnp-2018-318455
74. Eggink H, Kremer D, Brouwer OF, et al. Spasticity, dyskinesia and ataxia in cerebral palsy: Are we sure we can differentiate them? *Eur J Paediatr Neurol*. 2017;21(5):703-706. doi:10.1016/j.ejpn.2017.04.1333
75. Salm SMA van der, Haan RJ de, Cath DC, Rootselaar AF van, Tijssen MAJ. The eye of the beholder: inter-rater agreement among experts on psychogenic jerky movement disorders. *J Neurol, Neurosurg Psychiatry*. 2013;84(7):742. doi:10.1136/jnnp-2012-304113
76. Espay AJ, Lang AE, Erro R, et al. Essential pitfalls in “essential” tremor. *Movement Disord*. 2017;32(3):325-331. doi:10.1002/mds.26919
77. Algarni M, Fasano A. The overlap between Essential tremor and Parkinson disease. *Park Relat Disord*. 2018;46:S101-S104. doi:10.1016/j.parkreldis.2017.07.006
78. Clark LN, Louis ED. Challenges in essential tremor genetics. *Rev Neurol*. 2015;171(6-7):466-474. doi:10.1016/j.neurol.2015.02.015
79. Quinn NP, Schneider SA, Schwingenschuh P, Bhatia KP. Tremor--some controversial aspects. *Mov Disord*. 2011;26(1):18-23. doi:10.1002/mds.23289
80. Nieuwhof F, Toni I, Buijink AWG, Rootselaar AF van, Warrenburg BPC van de, Helmich RC. Phase-locked transcranial electrical brain stimulation for tremor suppression in dystonic tremor syndromes. *Clin Neurophysiol*. Published online 2022. doi:10.1016/j.clinph.2022.03.020
81. Cagnan H, Pedrosa D, Little S, et al. Stimulating at the right time: phase-specific deep brain stimulation. *Brain*. 2017;140(1):132-145. doi:10.1093/brain/aww286

82. Schreglmann S, Cagnan H. Towards phenotype-specific, non-invasive therapeutic interventions for tremor. *Clin Neurophysiol.* 2022;140:169-170.
doi:10.1016/j.clinph.2022.04.019

Figure legends

Figure 1 Analytical pipeline from tremulous movement and raw accelerometry signal to supervised statistical learning. (A) Tremor time-series data includes a wealth of information in its raw format, which can be conceptualized in frequency (power density) and time domain (Hilbert transform/wavelet). (B) Scheme illustrating the steps from recording tremulous hand movements via an inertial measurement unit, extraction of higher mathematical features and machine-learning to validation.

Figure 2 Consistency and generalizability of standard tremor characteristics in their ability to differentiate ET and PD tremor between and across centres. For the characterization of patient cohorts, established tremor characteristics including (A) area under the curve (AUC), (B) tremor stability index (TSI), (C) half width power (HWP), (D) peak frequency, (E) full width half maximum (FWHM) and (F) peak power from rest recordings were calculated and compared between centres and diseases. The comparison of half width

power, peak power and tremor-stability index resulted in group differences between disorders. Notably however, while the same metrics did not detect differences between diseases in any individual centre for half width power (C) and for several centres for tremor stability index (B) and peak power (F), there were substantial differences within diseases for the latter both. Hence, standard characteristics did not reliably differentiate tremor diseases in a generalizable manner.

Figure 3 Accuracy of unbiased feature-based supervised tremor classification depends on machine-learning algorithm, recording position, sensor-dimensionality and number of features. To explore which fundamental analytical settings are best suited to differentiate ET and PD tremor, we systematically examined the effects of (A) machine-learning algorithm, (B) recording position, (C) sensor-dimensionality (uniaxial vs. vector amplitude sum) and number of combined best-performing features on classification accuracy. For the differentiation between PD and ET we used the comparison between (D) PD rest vs. ET postural, (E) PD vs. ET postural, as well as (F) PD vs. ET rest recordings. All data are based on hctsa features extracted from 15s segments of tremor accelerometer recordings (down-sampled to 100Hz) from the more severely affected hand, and support vector machine AL algorithm (B-F) with an ascending number of combined features on the x-axis.

Figure 4 Assessment of combined, disease- and position-specific tremor features. (A) To identify the ideal feature combination for disease differentiation, the performance of random combinations of two, three, or four of all top-performing univariate features were assessed and arranged in ascending order (worst to best classification) by their combined classification accuracy. Notably, combining more than two features did not improve overall classification

accuracy. **(B)** The two features best performing in combination (feature 1: 'MF_GARCHfit_ar_P1_Q1.stde_normksstat', feature 2: 'MF_hmm_CompareNStates_06_24.maxLLtrain') showed consistently good accuracy across tri-axial data-sets, despite feature two individually performing worse on the London data-set. Performance on mono-axial data was consistently lower. **(C)** Feature-based machine-learning also detected posture-specific tremor characteristics irrespective of clinical tremor diagnosis. The results of the ten best performing features differentiation between rest and postural recordings are combined in a multi-dimensional space and graphically displayed after dimensionality reduction (arbitrary units on both x- and y-axis). Each dot is colour-coded by dataset and represents a singular recording. The distribution of datapoints is illustrating an excellent differentiation accuracy in both the training (99.6%) and validation (97.4%) data-set, allowing reliable differentiation between rest and postural recordings irrespective of tremor diagnosis.

Figure 5 Interpretation of disease defining signal features. The two best performing features capture different aspects of the time-series signal: exemplary data plots from raw accelerometry, power histogram and power density of randomly selected **(A)** PD and **(B)** ET patient recordings graphically illustrate different signal characteristics. Violin plots capture the observed distributions and group-differences in calculated feature values between PD and ET recordings for features **(C)** 'MF_hmm_CompareNStates' and **(D)** 'MF_GARCHfit'. The mechanistic translation of identified features is based on the interpretation of the underlying mathematical feature function, and graphically illustrated **(E, F)**: MF_hmm_CompareNStates **(E)** assesses the position of consecutive time-series signal points (illustrated as colour-coded crosses) in a multi-dimensional feature space. Signal positions change over time, either within a 'signal state' (dotted circle), marking a period of relative

signal stability, or between signal states, marking a change to another period of relative signal stability. 'MF_hmm_CompareNStates' evaluates the quality of fitting ≥ 2 states to the data. Low quality fits of the Hidden Markov Model (HMM) suggest that the data cannot be well explained with ≥ 2 states, suggesting that the signal stays within a singular signal state (red). Higher quality fitting of the HMM suggests that the signal naturally varies between several relatively stable signal states (blue). From the examined tremor data, ET tremor signals stay within a singular signal state, whereas PD tremor signals have a higher likelihood to change between ≥ 2 stable signal states. 'MF_GARCHfit' (F) examines in how far histogram-converted regular oscillations of different frequencies follow a Gaussian distribution (red) or deviate from it (blue). The latter occurs, for example, if a regular/symmetrical signal at a certain frequency is combined with asymmetrical oscillations. Higher values in this skewness metric relate to more prevalent asymmetries in the frequency distribution, as evident in PD tremor signals, whereas ET tremor signals show much lower levels of asymmetry.

Table 1) Clinical and demographic details of participants.

Table 2) Centre-specific accelerometry specifications.

Table 3) Metrics of machine-learning based classification in comparison to best-performing established tremor characteristics.

Supplement:

Supplementary Fig. 1 Changes of accuracy of tremor classification with ever increasing number of features. As extension of feature exploration (Fig. 3), increasing number of feature combinations were explored in order to identify the best suited feature combinations to differentiate ET and PD tremor. Extending the feature count >10 did not overall improve differentiation accuracy for the most relevant settings algorithm, position and dimensionality. All data are based on hctsa features extracted from 15s segments of tremor accelerometer recordings (down-sampled to 100Hz) from the more severely affected hand, and support vector machine AL algorithm (**B-F**) with an ascending number of combined features on the x-axis.

Supplementary Fig. 2 Differentiation accuracy depends on signal length. The stability and duration-dependence of features were examined using progressively longer segments of the same time-series from the exploratory cohort (59 ET, 73 PD). Differentiation accuracy gradually increased with longer time-series.

Table 1) Clinical and demographic details of participants.

	ET					PD					All (mean ± SD)		
	Graz	Budapest	London	Kiel	Nijmegen	Graz	Budapest	London	Kiel	Nijmegen	ET	PD	stat
N=	17	36	6	99	9	21	47	5	99	31	68 (167 incl. Kiel)	104 (203 incl. Kiel)	/
Age	60.6 ± 19.7 (1 missing)	67.6 ± 13.0	55.0 ± 14.5	n.a.	59.0 ± 14.9	63.4 ± 11.9	63.7 ± 10.1	73.6 ± 10.6	n.a.	63.0 ± 7.9	63.6 ± 15.9 (1 missing)	63.8 ± 10.2	P=0.929
Sex (% f)	F=3, M=11 (3 missing)	F=21, M=15	F=3, M=3	n.a.	F=1, M=8	F= 9, M=11 (1 missing)	F=20, M=27	F=1, M=4	n.a.	F=4, M=27	F=28, M=37 (3 missing)	F=34, M=69 (1 missing)	P=0.19
Age at onset	37.8 ± 21.8 (3 missing)	51.5 ± 20.5	39.3 ± 15.9	n.a.	17.0 ± 15.3	56.9 ± 12.8 (1 missing)	57.1 ± 13.6	62.6 ± 7.4	n.a.	56.0 ± 8.9	42.74 ± 23.0 (3 missing)	57.2 ± 12.0 (1 missing)	P=0.4* 10 ⁻⁶ (P<0.05)
Tremor duration	20.8 ± 23.5 (3 missing)	16.0 ± 15.0	15.7 ± 9.4	n.a.	42.0 ± 17.9	6.7 ± 5.8 (1 missing)	5.3 ± 4.4	11 ± 8.3	n.a.	6 ± 3.0	20.6 ± 19.3 (3 missing)	6.1 ± 4.8	P=2.1*10 ⁻¹¹ P<0.05
CRST	24.4 ± 9.9 (3 missing, \$)	39.1 ± 15.9 (\$)	17.8 ± 10.1 (1 missing, \$)	n.a.	37.4 ± 14.7 (#)	/					\$ 32.9 ± 11.2 # 37.4 ± 14.7	/	/
MDS-UPDRS						29.3 ± 15.8 (10 missing)	20.5 ± 17.1	43.4 ± 13.7	n.a.	44.0 ± 11.7 (1 missing)	/	30.3 ± 14.1	/

CRST=Fahn-Tolosa-Marin scale; MDS-UPDRS = Movement Disorder Society-United Parkinson’s disease rating scale; / = not applicable; n.a. = not available; \$ = CRST items 1-21;
= CRST items 1-14 only;

Table 2) Centre-specific accelerometry specifications.

	Graz	Budapest	London	Kiel	Nijmegen
Accelerometer	Tri-axial accelerometer (Biometrics ACL300 + amplifier K800, sensitivity ± 100 mV/g)	Tri-axial accelerometer (Kinesia quantitative motor assessment system; Great Lakes Neurotechnologies, sensitivity ± 4.5 mV/g)	Tri-axial accelerometer (MMA7361, Freescale Semiconductor, Inc.; sensitivity ± 800 mV/g)	Mono-axial accelerometer (Jäger-Tönnies GmbH/ VIASYS Healthcare Inc., sensitivity ± 50 mV/g)	Tri-axial accelerometer (Brain Products; sensitivity ± 1450 mV/g)
Sampling frequency	1000Hz	128Hz	500Hz	800Hz	5000Hz
Sensor placement	Dorsum of middle phalanx, index finger	Dorsum of middle phalanx, index finger	Dorsum of proximal phalanx, middle finger	Dorsum of proximal phalanx, middle finger	Dorsum of the hand
Rest position	30s hands hanging freely from arm rest, supported	15s hands hanging freely from arm rest, supported	20s hands resting on <u>ulnar edge</u> on chair armrest / table	60s hands hanging freely from arm rest, supported	60s hands hanging freely from arm rest, supported

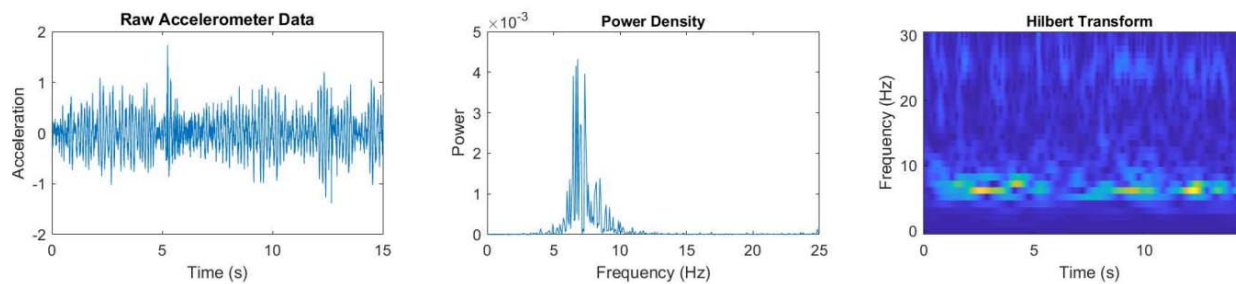
	lower arm	lower arm		lower arm	lower arm
Posture position	30s arms/wrists outstretched at shoulder level, fingers prone slightly parted	15s arms/wrists outstretched at shoulder level, fingers prone slightly parted	20s arms/wrists outstretched at shoulder level, fingers prone slightly parted	60s hands extended, prone, supported lower arm	60s arms/wrists outstretched at shoulder level, fingers prone slightly parted

Table 3) Metrics of machine-learning based classification in comparison to best-performing established tremor characteristics.

	ML Model	TSI	Peak power	TSI + Peak power
Accuracy	81.1%	70.4%	68.9%	70.5%
Sensitivity	86.4%	70.8%	68.7%	57.6%
Specificity	76.7%	70.2%	69.0%	80.8%
Positive predictive value	75.0%	57.6%	55.9%	70.8%
Negative predictive value	87.5%	80.8%	79.4%	70.2%

Figure 1

A



B

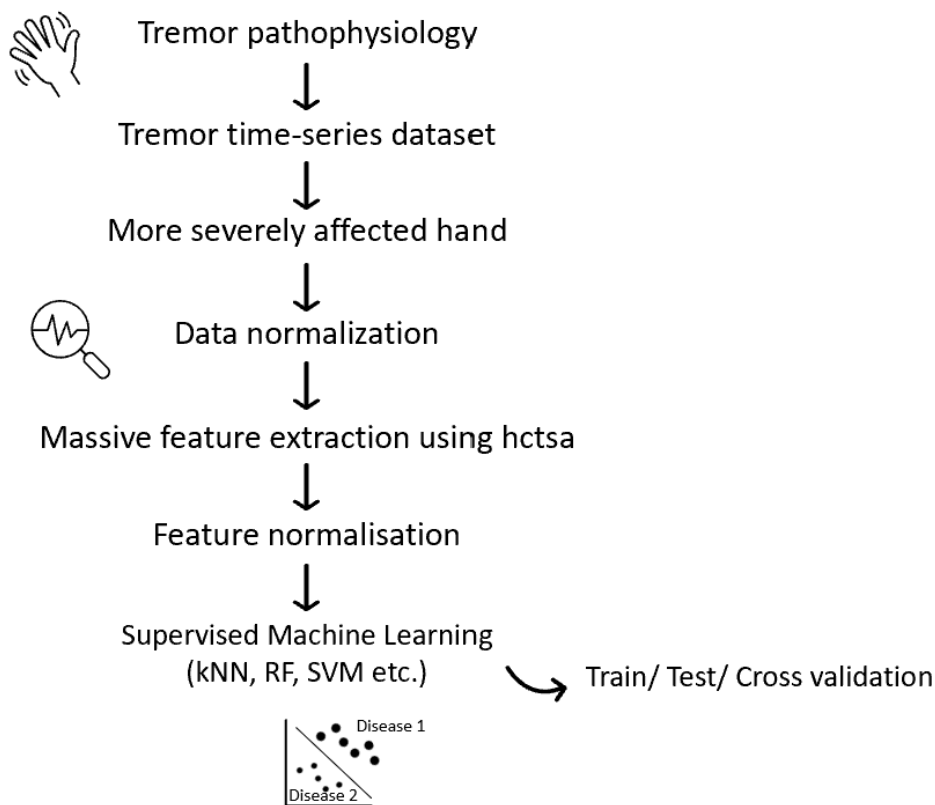


Figure 2

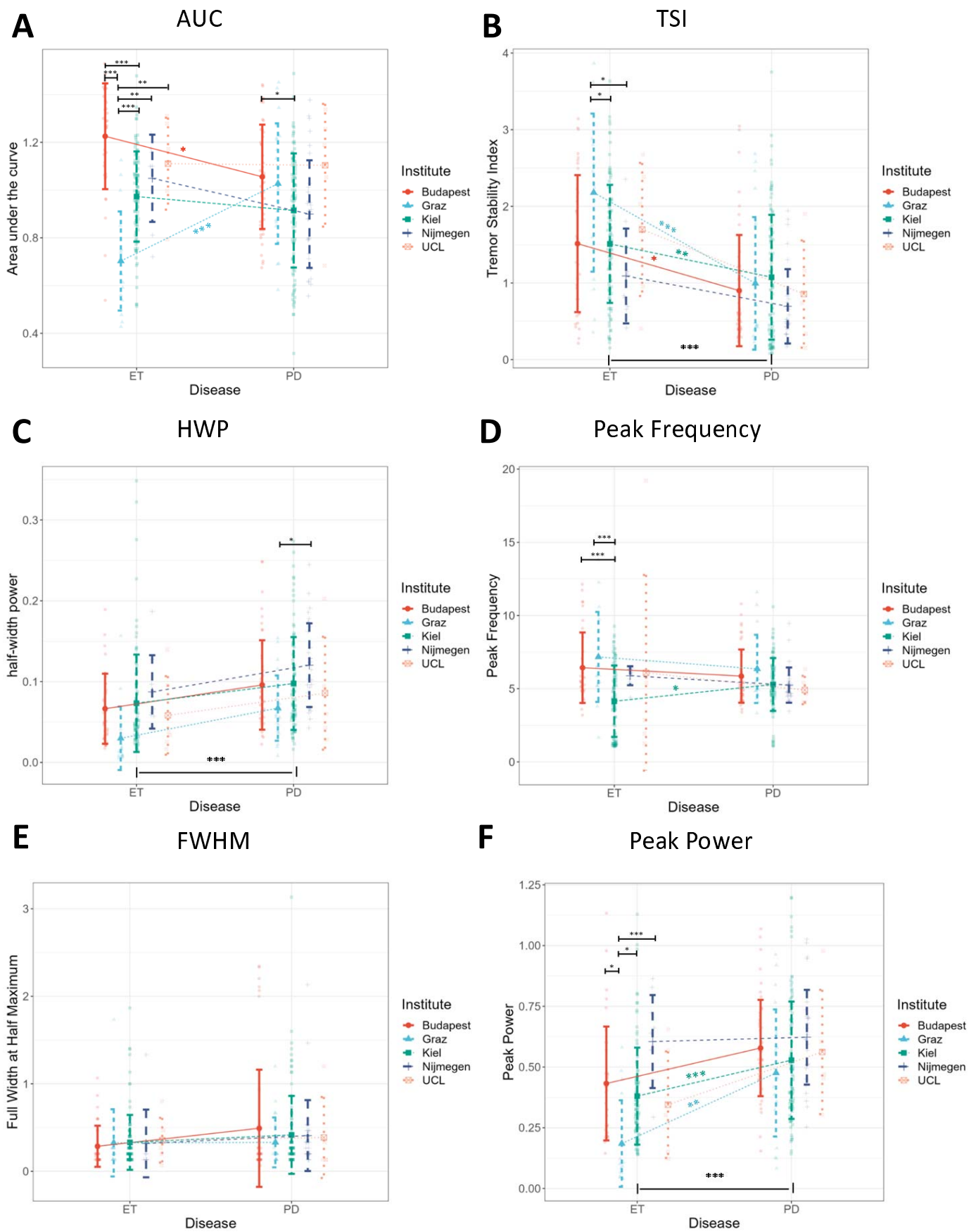


Figure 3

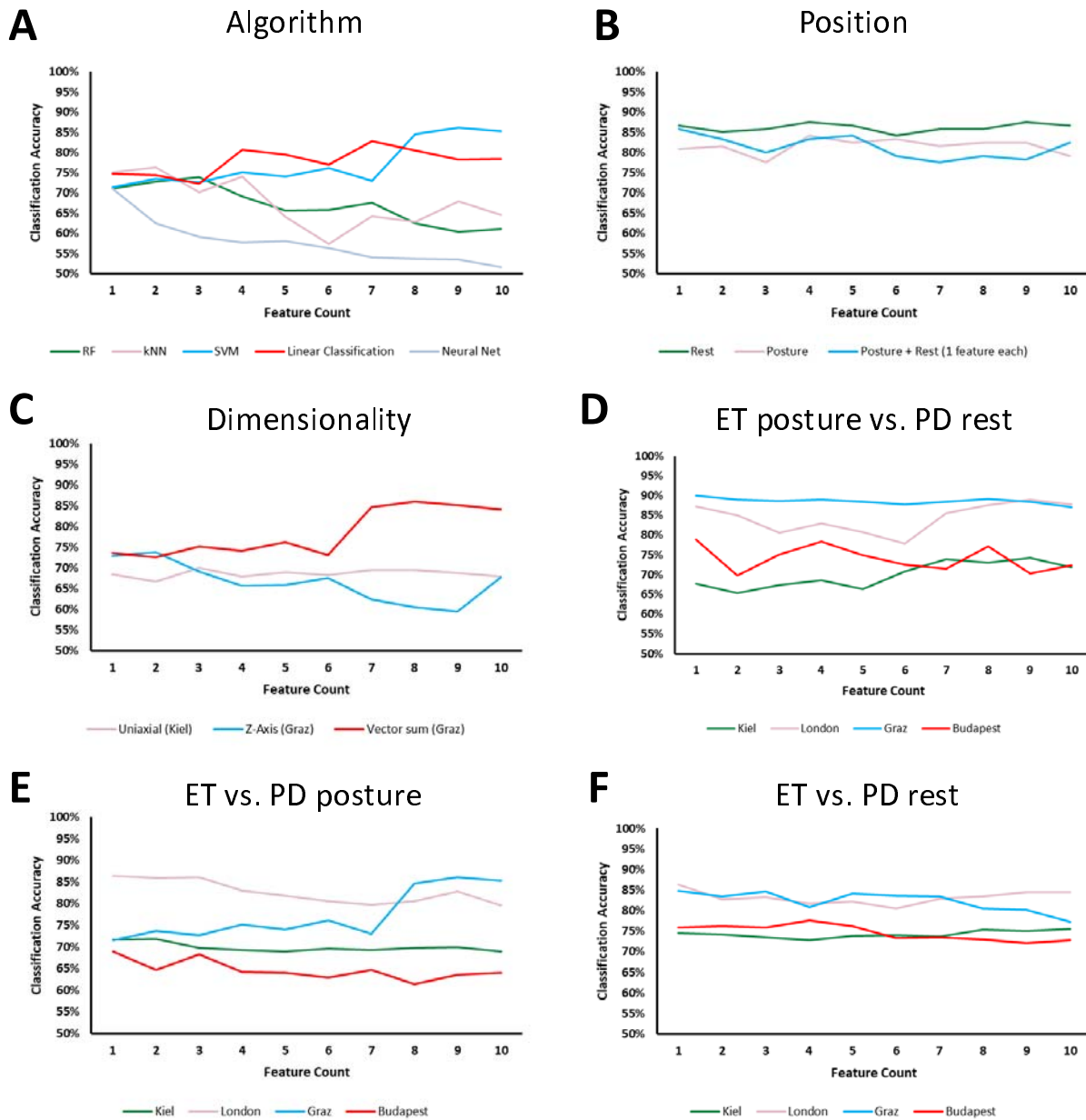


Figure 4

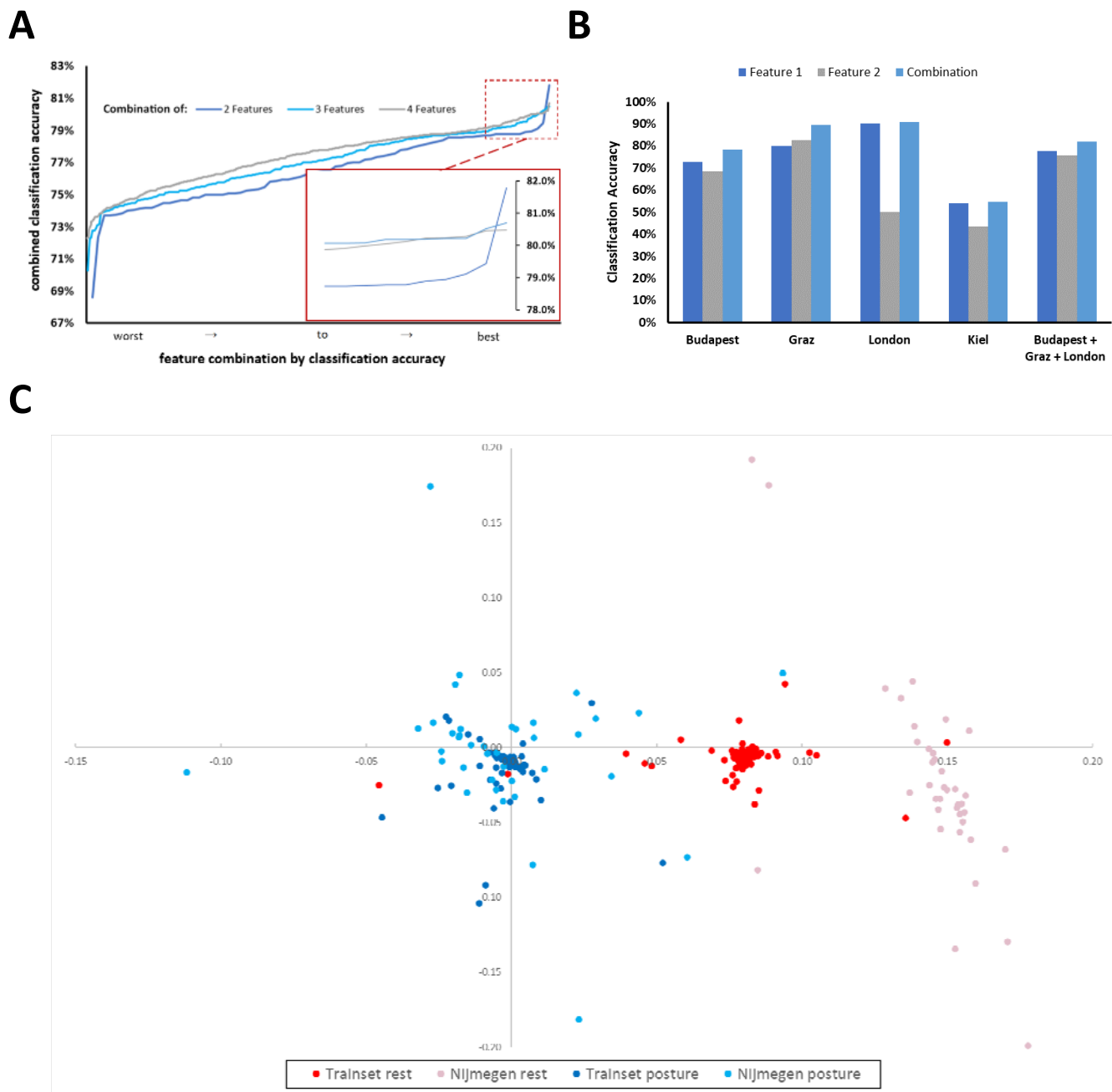
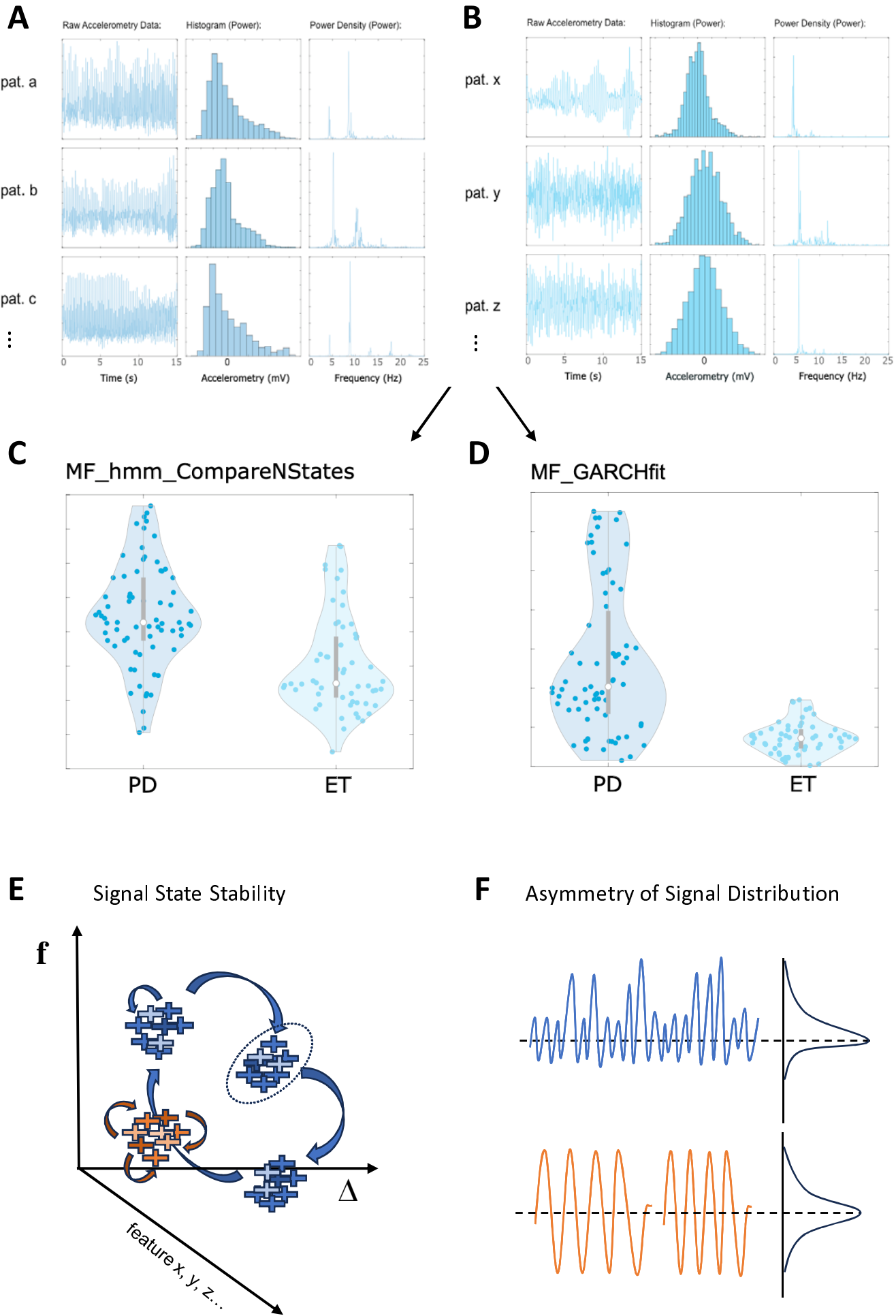
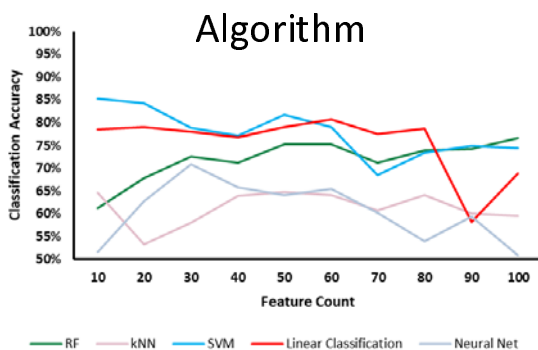


Figure 5

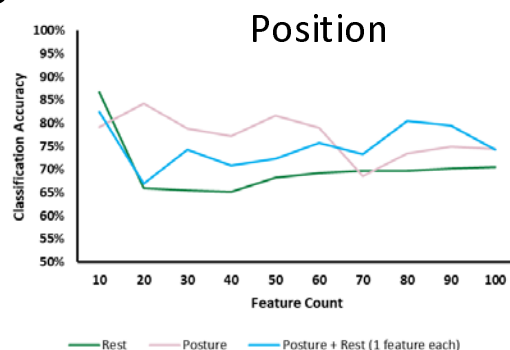


Supplementary Figure 1

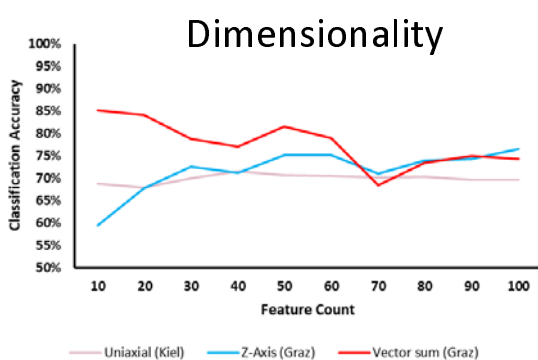
A



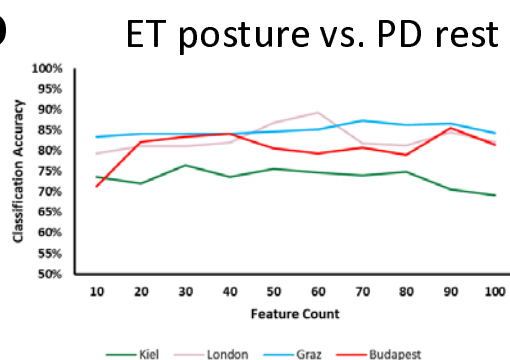
B



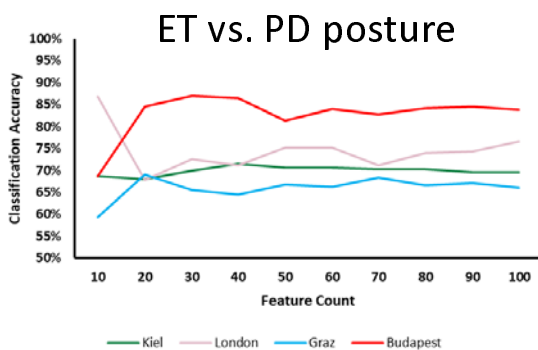
C



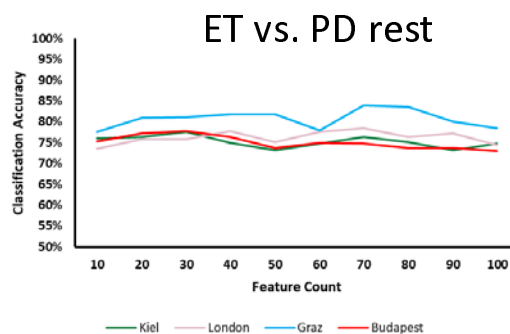
D



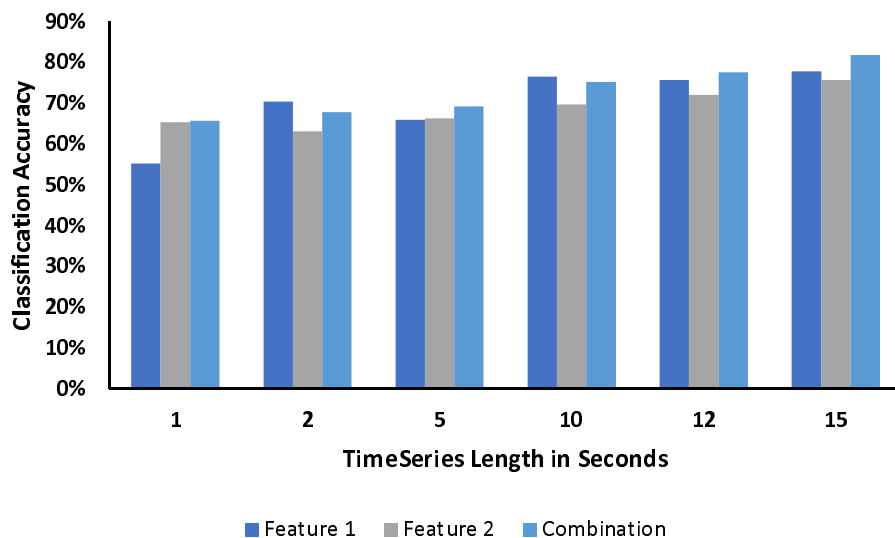
E



F



Supplementary Figure 2



Recording length (s)	1	2	5	10	12	15
Feature 1	55%	70%	66%	76%	76%	78%
Feature 2	65%	63%	66%	70%	72%	76%
Combination	66%	68%	69%	75%	77%	82%