REVIEW

Machine Learning in Tremor Analysis: Critique and Directions

Anwesan De,[1,2] Kailash P. Bhatia, MD, FRCP,[3] Jens Volkmann, MD, PhD, FEAN,[1] Robert Peach, PhD,[1,4] and Sebastian R. Schreglmann, MD, PhD, FEBN[1*]

[1]*Department of Neurology, University Hospital Wuerzburg, Wuerzburg, Germany*
[2]*Department of Electronics and Instrumentation, Birla Institute of Technology and Science, Pilani, Hyderabad, India*
[3]*Department of Clinical and Movement Neurosciences, Institute of Neurology, UCL, London, United Kingdom*
[4]*Department of Brain Sciences, Imperial College London, London, United Kingdom*

ABSTRACT: Tremor is the most frequent human movement disorder, and its diagnosis is based on clinical assessment. Yet finding the accurate clinical diagnosis is not always straightforward. Fine-tuning of clinical diagnostic criteria over the past few decades, as well as device-based qualitative analysis, has resulted in incremental improvements to diagnostic accuracy. Accelerometric assessments are commonplace, enabling clinicians to capture high-resolution oscillatory properties of tremor, which recently have been the focus of various machine-learning (ML) studies. In this context, the application of ML models to accelerometric recordings provides the potential for less-biased classification and quantification of tremor disorders. However, if implemented incorrectly, ML can result in spurious or nongeneralizable results and misguided conclusions. This work summarizes and highlights recent developments in ML tools for tremor research, with a focus on supervised ML. We aim to highlight the opportunities and limitations of such approaches and provide future directions while simultaneously guiding the reader through the process of applying ML to analyze tremor data. We identify the need for the movement disorder community to take a more proactive role in the application of these novel analytical technologies, which so far have been predominantly pursued by the engineering and data analysis field. Ultimately, big-data approaches offer the possibility to identify generalizable patterns but warrant meaningful translation into clinical practice. © 2023 The Authors. *Movement Disorders* published by Wiley Periodicals LLC on behalf of International Parkinson and Movement Disorder Society.

Key Words: accelerometer; artificial intelligence; classification; feature based

Introduction

Tremor is the most common disturbance of movement in man, defined as an involuntary rhythmic, oscillating movement of a body part. Oscillatory movements are a function of a mechanical component, that is, the inherent mechanical propensity of an object to oscillate, and a central component, that is, the result of an activation of agonist–antagonist muscles by rhythmic central nervous system activity.[1] Clinically, tremor presents with a particular *body distribution* (affecting limbs, the head, neck, jaw, vocal cords, or palate), *activation condition* (postural, kinetic, intention, or task specific), and *frequency*. It can occur in isolation or along various additional symptoms as part of a clinical syndrome, as summarized in the most recent consensus criteria.[2]

To date, tremor disorders remain as clinical diagnoses, and their definition has evolved with the first consensus criteria formalized only in 1998[3] and most

***Correspondence to:** Dr. Sebastian R. Schreglmann, Department of Neurology, University Hospital Wuerzburg, Josef-Schneiderstr. 11, 97080 Wuerzburg, Germany; E-mail: sebastian.schreglmann@gmail.com; schreglman_s@ukw.de

Relevant conflicts of interest/financial disclosure: Authors report no conflict of interest.

Funding agency: This work has not received dedicated financial support.

Robert Peach and Sebastian R. Schreglmann contributed equally.

Received: 26 September 2022; **Revised:** 16 January 2023; **Accepted:** 17 February 2023

Published online 23 March 2023 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/mds.29376

recently updated in 2018.[2] Over time, changes particularly affected the concept of essential tremor (ET),[4–6] which led to the identification of specific tremor entities previously subsumed within ET. The importance of exact clinical phenotyping is documented, for example, by the fact that evidence-based treatment options greatly differ between ET[7,8] and dystonic tremor (DT),[9] rubrics previously lumped together.

Correct diagnosis remains a challenge, with misdiagnosis rates reported up to 37%[10] or even 50%.[11] This uncertainty documents the need for more objective measures for tremor classification. In contrast to the perception of clinical raters, measurement devices provide objective measures of tremor movement, from which features can be extracted. In the ML context, a feature can be any quantifiable signal characteristic that can be fed into an algorithm. Several techniques equally fulfill quality criteria for the quantification and characterization of tremor,[12] with accelerometry by far the most widely used method.

For differentiation, simple tremor metrics have consistently proven unreliable,[13] and only two methods are validated to differentiate between ET and Parkinson's disease (PD): the tremor stability index (TSI)[14] was developed based on an exploratory data set of 16 PD *rest* tremor and 20 ET *postural* tremor recordings. In a validation data set of 55 PD and ET patients, a TSI cutoff of 1.05 (below: PD, above: ET) achieved a sensitivity and a specificity of 95% each. However, absolute mean TSI values for ET postural tremor have been reported below 0.5 by other groups,[15,16] casting doubt on its generalizability. The mean harmonic power[17] was established based on 30-s *postural* accelerometric recordings in 41 ET and 39 PD patients, reaching a classification accuracy of 94%. In a validation data set of 41 tremor patients, it has been shown to differentiate the two entities with 90.1% sensitivity and 100% specificity.[18] However, it has not been used beyond a third data set, where it was marginally outperformed by the TSI,[14] most likely as it relies on carefully calibrated accelerometers,[17,18] limiting its general use.

Currently, the most reliable diagnostic tests for ET,[19] enhanced physiological tremor[20] and functional tremor,[20–23] are diagnostic tools, based on the combination of electrophysiological features and clinical scoring systems,[13] which again are not regularly used in clinical practice.

Meanwhile, machine-learning (ML) approaches, the cornerstone of the current artificial intelligence (AI) revolution, are combining the theoretical attraction of purely data-driven analyses and the statistical power of large data sets, revolutionizing medicine[24] and neurology[25] at a rapid pace. Based on the integration of multiple high-dimensional sources of data, ML can help to identify unifying, consistent, and generalizable disease characteristics.[26] The promise of better detection, prediction, and treatment of human disease has been showcased in numerous examples.[27–31]

The unifying principle of ML approaches is to "learn" patterns from data without human instruction. Given its analytical power and unbiased nature, ML therefore holds great potential to aid tremor research. This review summarizes the most recent developments in the application of supervised ML to tremor disorders, as well as challenges in applying and translating these exciting possibilities into clinical practice.

Literature Search Strategy and Results

PubMed searches were conducted on November 25, 2022 (date of final search), without restrictions of publication language or type, considering publications from January 1, 2009, onward and the following search terms ("tremor*" OR "tremor" OR "tremor disorder") AND ("accelero*" OR "accelerometry" OR "accelerometer") AND ("machine*" OR "machine learning"). Additional publications were identified from the reference lists of selected papers. Identified abstracts were screened and selected based on reporting human clinical data.

After removing duplicates, the described searches provided 36 articles, including data of n = 1558 participants. The majority of publications (29, n = 1059 patients) related to the detection and quantification of PD tremor, five (n = 187 patients) focused on ET, four on the differentiation between PD and ET, and one each on the quantification of physiological and fatigue-induced enhanced physiological tremor, respectively. The sample size ranged between n = 398[32] and n = 1.[33] The majority of studies were conducted under laboratory conditions based on "scripted protocols" with fixed recording length and positions, whereas nine studies were "unscripted," that is, recording patients during activities of daily living (ADL; see Table 1).

Data Collection

Clinical Aspects

It is known from the classical tremor analysis literature that tremor depends not only on cause but also on limb position, activation, vigilance, and treatment state, which by nature of the experiment are controlled only under scripted protocols. In addition to the exact definition of diagnostic criteria applied, documentation of recording conditions therefore is crucial and should follow standard practice.

Recordings should therefore be performed with both forearms fully supported on an armrest to isolate the limb movement from external factors.[68] As governed by clinical phenomenology,[2] tremor movements are recorded in several defined positions. Although the main tremor positions are generally accepted, their

TABLE 1 *Summary of tremor machine-learning studies*

Author	Sample size	Diagnostic criteria, clinical details given (y/n)	Sensors	Sensor position	Accelerometer type	Sampling frequency	Analysis method
Scripted protocols							
Balachandar et al[34]	78 (19 PD, 42 ET, 17 DT	y, y	Smartphone	Wrist	Triaxial	50 Hz	RF, UMAP
Xing et al[32]	398 (257 PD, 141 ET)	n, n	#317A Noraxon	Distal phalanx	Triaxial	12 kHz	NN, CNN, logistic regression, XGBoost, RF, SVM
Ali et al[35]	35 (17 ET, 18 Ctrl)	y, y	IMU sensors	Dorsum of hand, posterior forearm, posterior upper arm	Triaxial	148.1 Hz	SVM, regression
Channa et al[36]	40 (20 PD, 20 Ctrl)	n, y	AWEAR bracelet	Wrist	Triaxial	100 Hz	Neural net clustering, kNN
Sigcha et al[37]	18 PD	y, y	Custom-built smartwatch	Wrist	Triaxial	50 Hz	Multitask classification model (CNN)
Shahtalebi et al[38]	81 (47 PD, 34 ET)	n, y	#317A Noraxon	Dorsal middle metatarsal	Triaxial	1.5 kHz	Pathological hand tremor net: gated recurrent unit–based algorithm
Shahtalebi et al[38]	81 (47 PD, 34 ET)	n, y	#317A Noraxon	Dorsal middle metatarsal	Triaxial	1.5 kHz	NeurDNet (ensemble architecture introduced in the paper)
Sajal et al[39]	52 PD	n, n	Smartphone	Dorsal metatarsals	na	na	kNN, SVM, Naive Bayes
Zhang et al[40]	12 PD	n, n	Activity AX3	Wrist	Triaxial	100 Hz	Linear SVM
de Araújo et al (2020)[41]	50 (18 PD, 32 Ctrl)	y, y	MetaMotionC	Dorsal middle metatarsal	Triaxial	100 Hz	kNN, logistic regression, SVM, linear discriminant, RF, decision tree, Gaussian naive Bayes

(Continues)

TABLE 1 *Continued*

Author	Sample size	Diagnostic criteria, clinical details given (y/n)	Sensors	Sensor position	Accelerometer type	Sampling frequency	Analysis method
Aljihmani et al[42]	40 (20 type 1 DM, 20 Ctrl)	n, n	ADXL355	Dorsal middle phalanx	Triaxial	45 Hz	Decision tree, kNN, SVM, ensemble classifier
Shawen et al[43]	13 PD	n, y	BiostampRC + Apple watch	Wrist + dorsal middle metatarsal	Triaxial	62.5 + 50 Hz	RF
Mahadevan et al[44]	81 (31 PD, 50 Ctrl)	n, y	Opal and BioStamp watch	Wrist	Triaxial	128 Hz	RF
Ribeiro et al[45]	35 (14 PD, 21 Ctrl)	n, n	na	Hand	Triaxial	na	Bidirectional gated recurrent units—RNN
McNames et al[46]	17 (10 PD, 7 Ctrl)	n, n	Opal	Wrist	Triaxial	128 Hz	Proprietary two-stage algorithm
Hssayeni et al[47]	24 PD	n, n	Great Lakes Neurotechnologies	Wrist and ankle	Triaxial	64 Hz	Gradient tree boosting and LSTM
Pulliam et al (2017)[48]	13 PD	y, y	Accelerometer + gyroscope	Both wrists and ankles	Triaxial	na	Regression models
Sanchez-Perez et al[49]	57 PD	n, y	na	Middle of lower arm	Triaxial	50 Hz	Fuzzy inference system
Kim et al[50]	92 PD	n, y	Custom-developed devise	Finger + wrist	Triaxial	20 Hz	CNN
Jeon et al[51]	85 PD	y, n	Wristwatch	Finger + wrist	Triaxial	125 Hz	PCA, kNN, SVM, discriminant analysis, decision tree
Ghassemi et al[52]	24 (13 PD, 11 ET)	y, y	na	Dorsal metatarsal	na	na	SVM
Koçer et al (2016)[53]	55 (35 PD, 20 Ctrl)	y, y	Nintendo Wii: ADCL330 accelerometer	Hand	Triaxial	10 Hz	SVM
Kostikis et al[54]	45 (25 PD, 20 Ctrl)	n, y	Smartphone iPhone	Dorsal middle metatarsal	Triaxial	20 Hz	Tree Bagger
LeMoyne et al[33]	1 ET (deep brain stimulation On vs. Off)	n, n	Smartphone iPhone	Dorsal middle metatarsal	na	100 Hz	SVM

(Continues)

TABLE 1 *Continued*

Author	Sample size	Diagnostic criteria, clinical details given (y/n)	Sensors	Sensor position	Accelerometer type	Sampling frequency	Analysis method
Ahlrichs and Samà[55]	76 PD	y, n	Accelerometer + gyroscopes	Wrist	Triaxial	40 Hz	SVM with different kernels
Veluvolu et al (2011)[56]	10 (5 surgeons, 5 Ctrl)	n, n	ADXL 203	na	Mono-axial	500 Hz	Multistep prediction with BMFLC and MS-AR model
Rigas et al (2012)[57]	23 (18 PD, 5 Ctrl)	y, y	na	Wrist, ankle, sternum, waist	Triaxial	62.5 Hz	HMM
Patel et al[58]	12 PD	n, y	Shimmer sensor	Middle of lower arm	Mono-axial	100 Hz	SVM
Unscripted protocols							
San-Segundo et al[59]	12 PD	n, y	Smartwatch	Wrist	Triaxial	800 Hz	CNN
Battista et al (2020)[60]	20 PD	y, y	Custom-made	Wrist	Triaxial	na	Threshold method
Papadopoulos et al[61]	37 PD	n, n	Inertial measurement units and smartphone	Hand	Triaxial	100 Hz	MIL
Zheng et al[62]	8 ET (26 h per patient)	y, y	Pebble smartwatch	Wrist	Triaxial	25 Hz	FFT and clustering
Garcia-Magarino et al (2016)[63]	21 (11 PD, 10 Ctrl)	n, y	Smartphone	Hand	Triaxial	100 Hz	"Shake detection"
Cole et al[64]	23 (19 PD, 4 Ctrl)	n, y	na	Arms, legs, and sternum	Triaxial	1 kHz	Dynamic neural network, DSVM, HMM
Tzallas et al (2014)[65]	48 PD	y, n	Custom-made	Wrists, ankles, and waist	Triaxial	na	HMM, SVM, RF
Cole et al (2012)[66]	na (29 hours PD, 15h Ctrl)	n, n	na	One sensor per limb	Triaxial	1 kHz	SVM, dynamic neural network
Roy et al[67]	23 (19 PD, 4 Ctrl)	n, y	Trigno Delsys	One sensor per limb	Triaxial	na	Multiple dynamic neural network classifier

Abbreviations: PD, Parkinson's disease; ET, essential tremor; DT, dystonic tremor; RF, Random Forest; UMAP-XGBoost, Extreme Gradient Boosting; NN, neural networks; CNN, convolutional neural network; SVM, support vector machine; RNN, recurrent neural networks; DM, diabetes mellitus; kNN, k nearest neighbors; na, not available; LSTM, long short-term memory; PCA, principal component analysis; BMFLC, band-limited multiple Fourier linear combiner; MS-AR multistep-autoregressive; HMM, hidden Markov models; MIL, multiple-instance learning; FFT, fast Fourier transform; DSVM, dynamic support vector machine.

exact execution differs from center to center and throughout the tremor literature: this pertains predominantly to *rest* (hands hanging freely without active holding[68] or resting flat[54] or on the ulnar side of the hand on a desk[38] or own lap,[52,69] or not specifically defined[32]).

For *posture* recordings many authors adopted the classic clinical position (extending both arms and hands in parallel in front of the chest[32,52]), as previously done in several seminal analytical tremor publications,[14,22,70–72] over the more lab-oriented routine of extending the hands while the arms are still resting on an armrest.[68,73,74] Others again used certain activation paradigms, such as spreading fingers,[75,76] arithmetic stress[77] to induce tremor, or even measured patients lying supine to exclude all possible interfering body movements.[77,78]

It is very likely that although clinically the activation pattern between, for example, slightly different *posture* positions might be fundamentally the same, the physical signal characteristics might be very different. Although the effects of slightly different accelerometer positions have been shown to influence classic signal characteristics such as amplitude, frequency, and total power,[73,79] different *rest* and *posture* recording positions have not been systematically compared so far. One argument in favor of the clinical measurement position is the more direct correlation with routine clinical diagnosis and scoring. Future studies should compare such positions, identifying the ideal tremor recording position.

It is generally accepted that recordings should be done after sufficient (ideally >12 hours for levodopa, >30 hours for dopamine agonists, >24 hours for β-blockers, and >36 hours for primidone) withdrawal of tremor-influencing medication and other substances such as nicotine and caffeine or tremor-inducing medication.[80,81]

Our search identified several disadvantages in the tremor ML literature so far, as only 12 studies (of 36, representing 455 participants) stated the diagnostic criteria applied, and only 22 stated the summary clinical/demographic details. The patient treatment state was reported in some papers[33,43,47,53–55,58,64,67]: whereas some patients were recorded in their medication *on* state,[37,62] other reports included patients on various medication states,[32,49] or throughout a medication cycle.[64,67] These aspects greatly limit the generalizability of results, as the clinical context is paramount for interpretation.

Experimental Setup

The exact placement of sensors does influence accelerometer signal characteristics.[73,79] Sensor placement in the aforementioned studies was not uniform, ranging from, for example, wrist,[34,36,37,40,44] dorsal middle metacarpal,[38,39,52,54,82] middle phalangeal,[42] distal phalangeal,[32] and middle of the lower arm[49,58] to combinations of, for example, wrist and finger[43,51] or wrist and ankle.[47] Traditionally, sensors have been—with some center-to-center differences—placed on the back of the hand.[68,73,74] Studies assessing the ideal sensor placement for tremor analysis showed contrasting results.[35,83]

Recording Device

Today capacitive microelectromechanical accelerometers—measuring translational acceleration—are most widely used.[84,85] To adequately capture the entire dynamics of the tremor signal, the device sampling rate, frequency range, and sensitivity have to be optimized for the respective signal. Most modern digital transducers have their own built-in A/D (analog-to-digital) converter or use pulse-width modulation to estimate the width of the pulse generated proportional to the physical quantity being measured. As a rule, the A/D converter must have a sampling rate of at least twice—ideally four times—the highest frequency of interest (Nyquist frequency) to avoid aliasing artifacts.[12,68,85] In the majority of studies the sampling frequency was 100 Hz, for example,[33,40,41,58] whereas some authors recorded at frequencies far above four times,[32,38,82] respectively, just about two-times the Nyquist frequency.[54]

Gravitational artifacts could be overcome by the application of multiple accelerometers on the same limb.[86] In practice, this has been applied very rarely[87] and is not commonly recommended for clinical or research recordings.[68]

It is encouraging that most of the aforementioned studies used triaxial recordings, which, by the nature of providing a vector sum, are independent of the main movement axis. It, however, remains to be proven if recording tremor along three axes is superior to mono-axial recordings.[68] Adding gyroscope data to triaxial recordings, however, has been shown not to increase tremor scoring accuracy.[43]

The accelerometer sensor range, that is, the maximal acceleration that can be recorded, ranged from 2 g[37] to 16 g.[51] Another sensor characteristic not commonly reported in the literature is the sensor sensitivity, measured in mV/g, governing the resolution at which a sensor can measure acceleration. This usually ranges between 20 and 100 mV/g, with exceptional devices providing a resolution of up to 800 mV/G. As the natural frequency of a limb depends on its weight,[1] the ideal sensor should also be as small and light as possible,[85] with heavier devices (eg, smartphones) presumably interfering more.

Signal Length and Preprocessing

As tremor intensity physiologically fluctuates with time, the recording length governs which temporal aspects of the signal are included in the analysis. For most studies, the length per tremor signal recorded ranged from 20[38,82] to 30[32] to 60 seconds,[51] representing common clinical examination durations. Whereas some studies suggested that differentiation accuracy plateaus with recording lengths from 5 seconds upward,[58] other studies for tremor

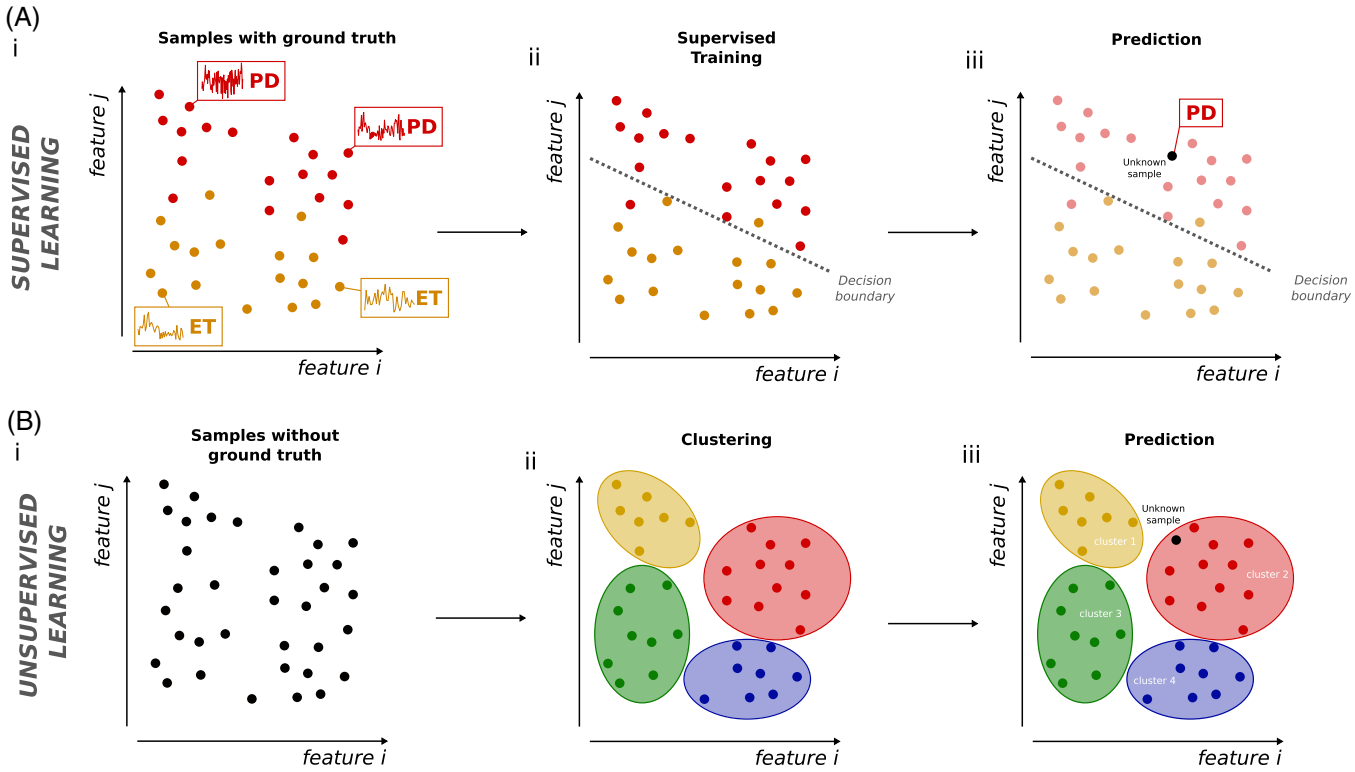

FIG. 1. (**A**) Supervised learning requires (i) samples with a known ground truth such as the disease (discrete classes) for classification or a clinical score (continuous value) for regression. Example here shows a classification problem. (ii) The supervised machine-learning (ML) algorithm learns a decision boundary that optimally separates the training samples based on the ground truth classes. (iii) Given a new sample with, for example, an unknown disease, the algorithm can assign a class based on its position relative to the decision boundary. (**B**) Unsupervised learning looks to uncover structure in the data when (i) ground truth samples are not known. (ii) Using a clustering algorithm, groups of samples that are naturally close to each other, but separated from the other clusters, can be uncovered. (iii) The clusters do not necessarily correspond to ground truth diseases but will reflect the structure of the underlying data and may, for example, find subgroups of diseases. [Color figure can be viewed at wileyonlinelibrary.com]

detection "in-the-wild" analyzed data collected over up to 26 hours per patient.[62]

To avoid biasing the analysis and learning features dependent on time-series length, equal lengths/equal number of data points of signal from all groups/conditions should be entered into the analytical pipeline. For in-the-wild experiments, it is necessary to segment the unequal prolonged time periods of experimentation into equal-length segments for effective classification. One study used a bag-of-features classification, where an overlapping window of fixed size was used to break the data into segments, and windows with the highest energy in the tremor frequency range were selected.[61]

ML Analysis

ML methods can generally be separated into supervised learning, meaning that an algorithm is trained to detect patterns in data according to ground truth labels or "gold standard" samples fed during training (Fig. 1A), and unsupervised learning, meaning that an algorithm detects similarities and differences between samples without ground truth labels or training on standard samples (Fig. 1B). The identified studies ultimately implemented supervised learning (although unsupervised ML, such as dimensionality reduction, was sometimes used as a precursor step). We therefore have not specifically included unsupervised learning in this review, which can provide additional insights into tremor data.[88]

Data Preparation

Data preparation involves the transformation of raw data into an appropriate format for modeling.

First, recordings (Fig. 2A) need to be "cleaned," removing mistakes, artifacts, or recording errors in the raw data when visually inspected in the time and frequency domain (Fig. 2B). In the absence of time stamps, as a reference for subselecting parts of the full recordings, the time window can be treated as a hyperparameter that needs to be optimized during the validation step of the ML process. However, we emphasize that biased subselecting of data, akin to "cherry picking," may omit the physiological fluctuations inherent to tremor. Some authors clip the first and last 10 seconds of the recordings to exclude artifacts and instability associated with starting the experiment[51]

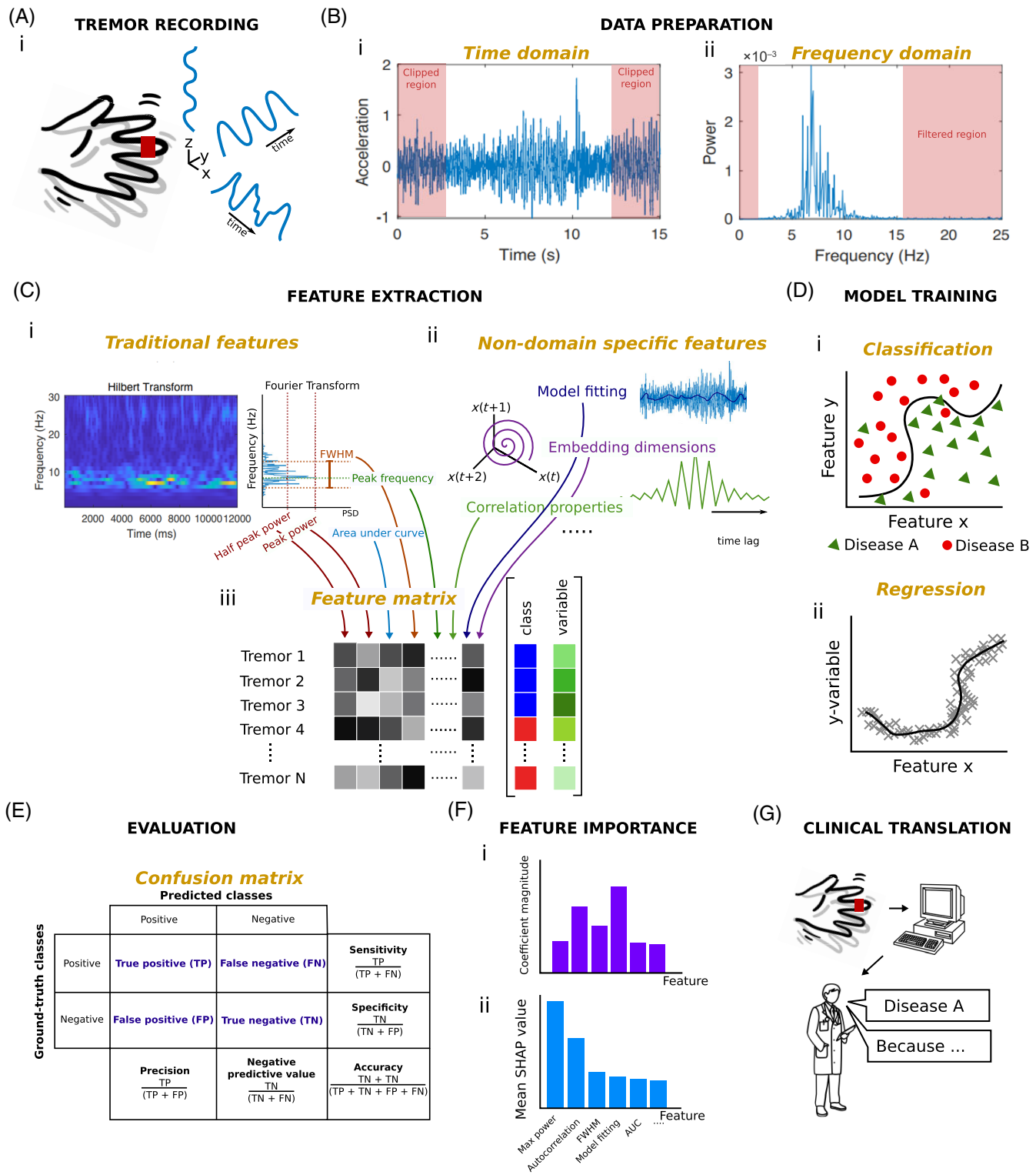

FIG. 2. Machine-learning workflow for analyzing tremor accelerometry recordings. (**A**) Raw tremor accelerometry recordings are (**B**) first preprocessed to clip periods of time with artifacts in both time (i) and frequency domain (ii) and then band-pass filtered to remove high- and low-frequency components. (**C**) Time-series features are extracted from the recordings, including (i) traditional hypothesis-driven features that are often derived from Fourier or Hilbert transforms and (ii) nondomain specific data-driven features, and which are then entered into a feature matrix (iii). (**D**) The ML model is trained for (i) classification or (ii) regression against the dependent variable. (**E**) The trained model is evaluated on a test set where the predicted classes of each sample are compared against the ground-truth classes using, for example, a confusion matrix. (**F**) The importance of each feature in the ML model is measured using different approaches, such as the (i) magnitude of the coefficient in a linear model or (ii) the Shapley additive explanation (SHAP) value in a nonlinear model so that (**G**) a clinician can make informed diagnoses and decisions. [Color figure can be viewed at wileyonlinelibrary.com]

(Fig. 2Bi). Nonphysiological artifacts due to ADL, volitional movement, and spurious artifacts are further removed by band-pass filtering (Fig. 2Bii), classically in the range of 1 to 16,[51] 0.5 to 15,[47] or 2 to 30 Hz.[68]

Next, the features for best characterization of tremor time series need to be selected (Fig. 2C). Here, one can take either (1) a hypothesis-driven approach using engineering features that one expects to be predictive or (2) a data-driven approach by utilizing nondomain specific feature extraction tools or directly learning the features using a deep-learning (DL) algorithm. The former allows to input prior knowledge into the modeling process and can often provide more interpretable insights, whereas the latter is less biased and can often identify novel, unexpected, and (potentially nonlinear) relationships.

Early attempts to analyze tremors used hypothesis-driven approaches and defined features, including position-dependent tremor peak frequency (Fig. 2Ci), patient demographics, and derived metrics to further quantify the dynamic nonlinear oscillatory characteristics of tremor time series.[74] Cross-correlation and autocorrelation features are generally accepted as good representatives of tremor characteristics.[89] Most of the reviewed accelerometer publications engineered 1 to 10 features from both time and frequency domains,[34,39,42,47] representing amplitude and regularity,[32,51] spectral power,[51] fast Fourier transform coefficients,[32] and spectrograms.[59]

Transitioning from a hypothesis-driven toward a data-driven approach, a selection of studies increased the number of features that they engineered (Fig. 2Cii), including 40 features,[32] 74 features,[43] or up to 290.[37,41] Any actively selected feature might however introduce bias and limit the scope for data exploration. Recently, an entirely data-driven approach employing highly comparative time-series analysis for massive feature extraction was applied to tremor signals and was able to accurately predict between pre- and poststimulation,[90] highlighting the potential of training predictive models of tremor signals without any domain knowledge. Finally, DL methods circumvent feature extraction steps by learning hierarchical features directly from time series.[91] Whereas classical ML approaches allow the input of human intuition and domain knowledge, DL methods can identify complex nonlinear patterns that cannot be captured by individually engineered features.

After a feature matrix has been constructed from engineered features (Fig. 2Ciii), a second round of data cleaning can be performed on the features, for example, mean imputation to fill in missing values.[32] Commonly, features are scaled by either normalization (bounding values between 0 and 1) or standardization (zero mean and unit variance). This is to make each feature comparatively similar in magnitude, allowing ML algorithms based on gradient descent to iterate/converge more smoothly. Some authors used standardization to scale the features to ensure comparability between ML model performance,[32] whereas others normalized, minimum to maximum, their power-spectrum density (PSD)–based features before implementing a two-stage algorithm.[46]

Models

The choice of supervised ML model is partially dependent on the research question and the accompanying data. Generally, supervised ML models can be separated into regression and classification models (Fig. 2D), whereby the former predict continuous values (eg, a tremor score) and the latter predict discrete values (eg, disease). A wide range of linear, ensemble, and DL methods have been applied for ML-based classification, and most of the reviewed publications applied all or a combination of these.

Linear discriminant analysis (LDA)[92,93] and support vector machines (SVM)[52,58] are frequently used algorithms for tremor classification. LDA inherently provides dimensionality reduction while preserving the interclass variance and ensuring maximum class separability, whereas SVMs try to estimate the best hyperplane that would serve as a boundary between classes by mapping the input to a higher dimension. Other popular ML algorithms for tremor classification include Random Forest, an ensemble algorithm used for both regression and classification tasks with an additional layer of randomness for bootstrap aggregation, and naive-Bayes classifier, based on Bayes' theorem, that assumes features are independent and each feature is learned separately, simplifying the learning in comparison to other algorithms.[38,94]

A DL model employs the use of artificial neural networks (ANN) several layers deep, which provides a very powerful nonlinear architecture to analyze the input. The characteristics of the hidden layers can be easily modified, making this architecture very flexible. The neurons in the hidden layer help recognize the features of time-series data, whereas the individual weights associated with the neurons represent the feature. DL models are completely independent of user-defined features, and time-series data with minimal preprocessing can be input directly into the model.

Convolutional neural networks are a modification of ANNs that learn relational information between spatially close data, for example, across the three dimensions of triaxial data or through time.[50,91,95] Alternatives for continuous time-series data analysis include sequence models like recurrent neural networks (RNN), which learn features across sequential timesteps, pushing the most relevant features from each timestep forward for a better representation of the

sequential data. RNN[82] and its modified versions, long short-term memory (LSTM)[96] and gated recurrent unit, proved very useful in sequential tremor time-series analysis.[45,97]

DL methods, however, have several disadvantages. They are very data intensive, and proper training requires large data sets and expensive computations. ANNs are often described as black boxes due to the long series of complex operations that are difficult to disentangle, although explainable AI is emerging as a possible route to interpretability.[98] The absence of handcrafted custom features lends this approach more objectiveness but leads to a rather opaque view of features and the inner workings of the algorithms. In this context, careful clinical phenotyping of the training set becomes an even more pressing necessity.

A common approach therefore is to use a variety of models, including linear, ensemble, and DL-based algorithms.[32,99] This has similarly been tried in unscripted experiments.[59] An interesting study used multiple-instance learning, whereby a bag (sequence of signal segments associated with a single class label) was provided as input and the bag (ie, not singular data points) mapped to a label, overcoming noise in the data set.[61]

Training and Evaluation

Once an appropriate ML model has been chosen, it must be optimally trained and evaluated. Training involves optimizing the model parameters and hyperparameters on a training set and a validation set, respectively, using cross-validation, whereas evaluation usually requires testing the optimized model on unseen test sets.

Cross-validation is a technique of splitting the data into a specified number of folds and permuting the training and validation sets among them. This helps to tune hyperparameters, overcome instability in sampling, and test the model performance on unseen data. K-fold cross-validation (splitting the training data into, eg, 10-folds) is one of the most common forms,[32,36,38,47] and leave-one-out-cross-validation implies the number of folds equals the number of samples.[51–53,61] To tune the hyperparameters of a model, a popularly adopted method is grid search, testing all possible combinations of hyperparameters within a defined range of values to find the optimal combination. Cross-validation is usually carried out on the training data within the grid search loops to evaluate the optimal set of hyperparameters.[58]

Given a trained model, its ability to generalize to unseen data must be evaluated. Deciding the best metrics for evaluating model performance is a critical step in the process. A poor metric choice might lead to a distorted representation of the model capabilities and characteristics; for example, a model that simply classifies every sample into the same class in the presence of a large class imbalance will inevitably lead to many positively classified samples but has little clinical use. For tremor classification, model evaluation is frequently done using a confusion matrix (Fig. 2E), which allows the calculation of precision, recall, specificity, and F1 scores.[32,61,100] The F1 score appears to be a favorable metric especially in unbalanced class cases.[100] For binary classification models, some studies used the receiver operating characteristic curve.[32,36,38] Regression models must be evaluated using some measure of distance between the ground truth and predicted scores; metrics such as mean absolute error or root mean square have been used in this context.[35,51]

The computational cost to train ML models varies considerably. We emphasize that standard ML models (linear models, SVMs, Random Forest) with a few hundred samples can easily be trained on modern laptops in a matter of seconds to minutes. The computational cost increases with larger data sets and algorithms, with increasing numbers of parameters to optimize, that is, ANN, which can take hours to days to train. For tremor measurements, the computational cost of such models will not yet be a limiting factor.

Interpretation

Adequate interpretation of ML results in the context of each experiment is paramount for insights into disease mechanisms and clinical translation. As the tremor field transitions from hypothesis-driven analyses toward data-driven analyses with high-dimensional data sets and complex DL models, it is necessary to identify the features that contributed the most toward classification (Fig. 2F).

Model comparison allows to train multiple models on different features, and subsequently compare their performance, identifying the feature sets providing the strongest predictive power. In this way, several ML studies identified, for example, models training on power-spectrum engineered features[101] or SVM models[52] to perform best. The coefficients in linear ML models, for example, linear SVM or logistic regression, are often directly interpretable (Fig. 2Fi). For example, Ali et al[35] found a linear relationship between Fahn–Tolosa–Marin tremor score and PSD features using a least-squares linear regression, with which they could define a model equation relating the contributions of each feature to the clinical score.[35]

Interpreting results of nonlinear ML models is more difficult: model gain analysis can be used to identify the parts of the signal most relevant for classification, for example, after XGBoost (Extreme Gradient Boosting) and LSTM analysis,[47] or the use of Shapley additive explanation (SHAP) values (Fig. 2Fii). Another approach generated visual explanations of tremor

spectrograms using gradient-weighted class activation mapping to highlight the regions of the spectrogram most relevant for classification.[38]

Discussion

The results from the studies summarized earlier document, in principle, the great potential of ML for the study of tremor disorders, as time-series analysis has evolved from comparing relatively simple metrics to more sophisticated feature-based analysis. Concurrently, methods for measuring tremor signals are well established, and the strengths and weaknesses of ML analytical approaches are acknowledged. As evidenced by the strong presence of authors from the engineering/signal analysis field in the aforementioned publications, ML technology is now sufficiently developed and widely available.

This contrasts with the lack of clinically well-defined tremor cohorts in our search. As in other fields of medicine,[102] the absence of clinical and demographic details from most of the papers summarized earlier limits their clinical value: similar to training a clinician's eye to detect clinical patterns, ML-based analyses can only be as good as the data fed to the algorithm. This pertains to the meticulous clinical description to determine potential bias. Further, most analyses are limited to historically established, well-known features, potentially introducing *signal bias* and limiting the scope for data exploration. Together with relatively small sample sizes, the single-center design by its nature is limiting the ability to identify generalizable, disease-specific characteristics.

What Is Ground Truth in Tremor Research?

Accurate tremor classification has proven to be a challenge and an ongoing struggle for the movement disorder community,[5,6,11,103] and several concepts continue to be debated. As tremor remains a clinical diagnosis without clear biomarker profile beyond DAT SPECT (dopamine transporter-single-photon emission computed tomography) scans to quantify dopaminergic neurodegeneration,[104] it is important to recognize that it is the combination of *diagnostic criteria* and their *clinical interpretation* that governs how individual symptoms are categorized. As there is no doubt this already affects the differentiation of well-established concepts such as ET and PD, this is even more likely to be the case in, for example, comparably recent additions to the diagnostic spectrum, such as DT and tremor associated with dystonia.[2,105–107] In the absence of consensus biomarkers, the clinical diagnosis remains the gold standard for comparison for now. This in turn implies that, by its nature, absolute ground truth in tremor disorders is veiled by uncertainty.

What to Expect from ML-Based Tremor Studies

It is therefore important to correctly set the expectations for ML tremor studies. They can only learn/identify patterns laid out by the clinical diagnoses fed in the training data set. Therefore, it is neither realistic nor desirable to replicate prediction accuracies 100% in mono-centric data sets but rather to identify truly disease- and not center-, device-, or clinician-specific characteristics. The larger the pool of recordings from several centers and populations, the more representative will be the results. Purely data-driven attempts,[34] aiming at clustering patient recordings without the influence of clinical diagnoses, might provide additional insights, if applied to clinically well-characterized multicenter data sets of sufficient size. First, clustering exercises based on a very limited number of manually chosen variables allowed to identify patient subgroups based on medication response.[108] The power of such approaches becomes evident when all possible movement characteristics are included via unbiased feature extraction—ML-based analyses therefore bear the realistic potential to predict the effect of an intervention.[90]

Lessons from Other Domains

Despite the large number of research papers applying ML across various domains of medicine,[109] models so far only rarely transition into routine clinical practice due to various conceptual and methodological issues[28,30]: first, research papers that use ML to target a medical problem often focus on the design and development of an ML model, followed by evaluation on a limited data set, but leave validation, diffusion, and scaling of the model into clinical care untouched. The validity of an ML model should, however, be assessed in different settings and across time periods to optimize its validity.[110] Second, translation necessarily requires continued monitoring and maintenance as factors such as data quality and population characteristics change over time. The continued training of ML models to ensure robust predictions in medical care is an active field of research.[111] Moreover, changes in regulatory frameworks or definitions of diseases require the evolution or overhaul of existing models.[112] Third, the predictions made by ML models can sometimes be difficult to interpret and explain, posing relevant medicolegal implications in potentially high-risk clinical decisions, as the parties involved need to understand the reasons for a health-care decision. Depending on the training data set used, ML models can also exacerbate existing racial and socioeconomic health inequalities.[113]

After a first wave of ML publications, several efforts focused on identifying and overcoming relevant disadvantages[114]: for example, a meta-analysis on ML approaches in imaging research identified the frequently unmet need to compare the performance of clinical and

ML diagnostic accuracy on the same data set, as well as external validation.[29] Despite the translational difficulties of ML models, they still provide unprecedented diagnostic and prognostic opportunities,[28] driving a shift toward precision and personalized health care likely to accelerate further.

Like traditional statistics, ML methods are prone to being biased if not correctly implemented. Over-fitting algorithms to the specific nuances of a data set can produce a model that cannot generalize to new data. Commonly, over-fitting is overcome by splitting the data set into a training set (trains the ML model), a validation set (optimizes ML hyperparameters), and a test set (evaluates final model performance). Other routes to prevent over-fitting include early stopping criteria (DL models usually include criteria to stop the training process when the model performance is no longer improving), expanding the data set (with a larger variety of samples), and regularization (penalizing the parameters with larger coefficients to limit variance in the ML model). Similarly, underlying issues with the data set can bias the trained ML model. For example, outliers, which are samples that are unlike the rest of the data set, can heavily bias ML models. The definition of an outlier is not fixed and depends on the data set; nonetheless, outliers can be detected using visualization or statistical methods,[115] and training multiple ML algorithms can help avoid models that are more robust to outliers.

Future Directions for Tremor Research

ML predominantly holds great promises for improved generalizability for tremor research (Fig. 2G). From the methodological principles and advantages summarized earlier, we identify several core points to aid future analyses.

Learning from the successful application of ML in other domains, it is clear that only collaborative analyses of recordings from different centers, including as many samples as possible and sampling the whole spectrum of presentations, will provide the opportunity to detect not patient-, center-, or population- but truly disease-specific characteristics, improving diagnosis and prognosis. Attempts in this direction failed so far, possibly due to a mixture of the aforementioned limitations.[34]

Increasing sample size alone, however, is insufficient to improve accuracies, as *data set* and *spectrum bias*, relating to the coverage of the disease/control spectrum and its distribution within the data, are further relevant factors influencing ML performance.[114] First, this relates to well-documented clinical inclusion criteria and excluding treatment effects. Second, this should be accounted for by ideally selecting a range of participating centers, covering general neurology outpatient clinics as well as centers with dedicated specialist movement disorder expertise.

It still remains to be seen which tremor characteristics/features are ideal to be compared across centers, as different sensor positions, recording devices, and protocols influence established metrics. Thus, scripted protocols will remain superior in addressing the pressing questions in tremor research, such as improving diagnosis, prognosis; monitoring treatment effect; predicting treatment response; and exploring disease mechanism simply by reducing the amount of noise on top of layers of physiological inter- and intraindividual variability. The combination of multicenter data sets and the use of extensive, unbiased, and automatically extracted features[116] appears a realistic strategy[90] to overcome the problem of multiple known and unknown confounders introduced by the aforementioned factors.

Simultaneously, the interpretability of ML results will remain key for the translation of such attempts into clinical practice. It is essential that derived results be compared against known metrics and clinically evaluated so that they remain interpretable and intuitive to clinicians.[117]

To make a translational impact on clinical care, features identified through ML analyses should be made available to the community, so they can be applied in routine clinical accelerometer assessments.

Acknowledgments: S.R.S. received support from the Advanced Clinician Scientist program of the Interdisciplinary Centre for Clinical Research, Wuerzburg, Germany. He is a fellow of the Thiemann-Foundation. R. P. acknowledges funding from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) Project-ID 424778381-TRR 295. Open Access funding enabled and organized by Projekt DEAL.

Data Availability Statement

Data sharing not applicable - no new data generated

References

1. Hallett M. Overview of human tremor physiology. Mov Disord 1998;13(S3):43–48.
2. 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.
3. Deuschl G, Bain P, Brin M. Consensus statement of the Movement Disorder Society on Tremor. Mov Disord 1998;13(S3):2–23.
4. Chouinard S, Louis ED, Fahn S. Agreement among movement disorder specialists on the clinical diagnosis of essential tremor. Mov Disord 1997;12(6):973–976.
5. Espay AJ, Lang AE, Erro R, et al. Essential pitfalls in "essential" tremor. Mov Disord 2017;32(3):325–331.
6. Quinn NP, Schneider SA, Schwingenschuh P, Bhatia KP. Tremor--some controversial aspects. Mov Disord 2011;26(1):18–23.
7. Ferreira JJ, Mestre TA, Lyons KE, et al. MDS evidence-based review of treatments for essential tremor. Mov Disord 2019;34(7):950–958.
8. Zappia M, Albanese A, Bruno E, et al. Treatment of essential tremor: a systematic review of evidence and recommendations from

the Italian Movement Disorders Association. J Neurol 2013; 260(3):714–740.

9. Fasano A, Bove F, Lang AE. The treatment of dystonic tremor: a systematic review. J Neurol Neurosurg Psychiatry 2014;85(7): 759–769.
10. Jain S, Lo SE, Louis ED. Common misdiagnosis of a common neurological disorder: how are we misdiagnosing essential tremor? Arch Neurol 2006;63(8):1100–1104.
11. Schrag A, Münchau A, Bhatia KP, et al. Essential tremor: an overdiagnosed condition? J Neurol 2000;247(12):955–959.
12. Haubenberger D, Abbruzzese G, Bain PG, et al. Transducer-based evaluation of tremor. Mov Disord 2016;31(9):1327–1336.
13. van der Veen S, Klamer MR, Elting JWJ, et al. The diagnostic value of clinical neurophysiology in hyperkinetic movement disorders: a systematic review. Parkinsonism Relat Disord 2021;89:176–185.
14. di Biase L, Brittain J-S, Shah SA, et al. Tremor stability index: a new tool for differential diagnosis in tremor syndromes. Brain 2017;140(7):1977–1986.
15. Luft F, Sharifi S, Mugge W, et al. A power spectral density-based method to detect tremor and tremor intermittency in movement disorders. Sensors 2019;19(19):4301.
16. Panyakaew P, Cho HJ, Lee SW, Wu T, Hallett M. The pathophysiology of dystonic tremors and comparison with essential tremor. J Neurosci 2020;40(48):9317–9326.
17. 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.
18. Wile DJ, Ranawaya R, Kiss ZHT. Smart watch accelerometry for analysis and diagnosis of tremor. J Neurosci Methods 2014;230: 1–4.
19. 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.
20. van der Stouwe AMM, Elting JW, van der Hoeven JH, et al. How typical are 'typical' tremor characteristics? Sensitivity and specificity of five tremor phenomena. Parkinsonism Relat Disord 2016;30: 23–28.
21. Kramer G, der Stouwe AMMV, Maurits NM, Tijssen MAJ, Elting JWJ. Wavelet coherence analysis: a new approach to distinguish organic and functional tremor types. Clin Neurophysiol 2018;129(1):13–20.
22. Schwingenschuh P, Katschnig P, Seiler S, et al. Moving toward "laboratory-supported" criteria for psychogenic tremor. Mov Disord 2011;26(14):2509–2515.
23. Schwingenschuh P, Saifee TA, Katschnig-Winter P, et al. Validation of "laboratory-supported" criteria for functional (psychogenic) tremor. Mov Disord 2016;31(4):555–562.
24. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. New Engl J Med 2019;380(26):2589–2590.
25. Auger SD, Jacobs BM, Dobson R, Marshall CR, Noyce AJ. Big data, machine learning and artificial intelligence: a neurologist's guide. Pract Neurol 2021;21(1):4–11.
26. Savage N. How AI and neuroscience drive each other forwards. Nature 2019;571(7766):S15–S17.
27. Fleuren LM, Klausch TLT, Zwager CL, et al. Machine learning for the prediction of sepsis: a systematic review and meta-analysis of diagnostic test accuracy. Intensive Care Med 2020;46(3):383–400.
28. Aggarwal R, Sounderajah V, Martin G, et al. Diagnostic accuracy of deep learning in medical imaging: a systematic review and meta-analysis. NPJ Digit Med 2021;4(1):65.
29. Liu X, Faes L, Kale AU, et al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. Lancet Digit Health 2019;1(6):e271–e297.
30. Roberts M, Driggs D, Thorpe M, et al. Common pitfalls and recommendations for using machine learning to detect and prognosticate for COVID-19 using chest radiographs and CT scans. Nat Mach Intell 2021;3(3):199–217.
31. Cho SJ, Sunwoo L, Baik SH, Bae YJ, Choi BS, Kim JH. Brain metastasis detection using machine learning: a systematic review and meta-analysis. Neuro Oncol 2020;23(2):214–225.
32. 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 2022;16: 701632.
33. LeMoyne R, Tomycz N, Mastroianni T, McCandless C, Cozza M, Peduto D. Implementation of a smartphone wireless accelerometer platform for establishing deep brain stimulation treatment efficacy of essential tremor with machine learning. Annu Int Conf IEEE Eng Med Biol Soc 2015;2015:6772–6775.
34. Balachandar A, Algarni M, Oliveira L, et al. Are smartphones and machine learning enough to diagnose tremor? J Neurol 2022;269 (11):6104–6115.
35. Ali SM, Arjunan SP, Peters J, et al. Wearable sensors during drawing tasks to measure the severity of essential tremor. Sci Rep 2022; 12(1):5242.
36. Channa A, Ifrim R-C, Popescu D, Popescu N. A-WEAR bracelet for detection of hand tremor and bradykinesia in Parkinson's patients. Sensors 2021;21(3):981.
37. Sigcha L, Pavón I, Costa N, et al. Automatic resting tremor assessment in Parkinson's disease using smartwatches and multitask convolutional neural networks. Sensors 2021;21(1):291.
38. Shahtalebi S, Atashzar SF, Patel RV, Patel RV, Jog MS, Mohammadi A. A deep explainable artificial intelligent framework for neurological disorders discrimination. Sci Rep 2021;11(1): 9630.
39. Sajal MSR, Ehsan MT, Vaidyanathan R, et al. Telemonitoring Parkinson's disease using machine learning by combining tremor and voice analysis. Brain Inform 2020;7(1):12.
40. Zhang A, Torre FD la, Hodgins J. Comparing laboratory and in-the-wild data for continuous Parkinson's disease tremor detection. Paper presented at: 2020 42nd Annu Int Conf IEEE Eng Medicine Biology Soc EMBC 2020;5436–5441.
41. de Araújo ACA, Santos EG d R, de Sá KSG, et al. Hand resting tremor assessment of healthy and patients with Parkinson's disease: an exploratory machine learning study. Front Bioeng Biotechnol 2020;8:778.
42. Aljihmani L, Kerdjidj O, Zhu Y, et al. Classification of fatigue phases in healthy and diabetic adults using wearable sensor. Sensors 2020;20(23):6897.
43. Shawen N, O'Brien MK, Venkatesan S, et al. Role of data measurement characteristics in the accurate detection of Parkinson's disease symptoms using wearable sensors. J Neuroeng Rehabil 2020; 17(1):52.
44. Mahadevan N, Demanuele C, Zhang H, et al. Development of digital biomarkers for resting tremor and bradykinesia using a wrist-worn wearable device. NPJ Digit Med 2020;3(1):5.
45. Ribeiro LCF, Afonso LCS, Papa JP. Bag of samplings for computer-assisted Parkinson's disease diagnosis based on recurrent neural networks. Comput Biol Med 2019;115:103477.
46. McNames J, Shah VV, Mancini M, et al. A two-stage tremor detection algorithm for wearable inertial sensors during Normal daily activities. Annu Int Conf IEEE Eng Med Biol Soc 2019;2019:2535–2538.
47. Hssayeni MD, Jimenez-Shahed J, Burack MA, Ghoraani B. Wearable sensors for estimation of parkinsonian tremor severity during free body movements. Sensors 2019;19(19):4215.
48. Pulliam CL, Heldman DA, Brokaw EB, et al. Continuous assessment of levodopa response in Parkinson's disease using wearable motion sensors. IEEE Trans Biomed Eng 2017;65(1):159–164.
49. Sanchez-Perez LA, Sanchez-Fernandez LP, Shaout A, et al. Rest tremor quantification based on fuzzy inference systems and wearable sensors. Int J Med Inform 2018;114:6–17.
50. Kim HB, Lee WW, Kim A, et al. Wrist sensor-based tremor severity quantification in Parkinson's disease using convolutional neural network. Comput Biol Med 2018;95:140–146.

51. Jeon H, Lee W, Park H, et al. High-accuracy automatic classification of Parkinsonian tremor severity using machine learning method. Physiol Meas 2017;38(11):1980–1999.
52. Ghassemi NH, Marxreiter F, Pasluosta CF, et al. Combined accelerometer and EMG analysis to differentiate essential tremor from Parkinson's disease. Annu Int Conf IEEE Eng Med Biol Soc 2016; 2016:672–675.
53. Koçer A, Oktay AB. Nintendo Wii assessment of Hoehn and Yahr score with Parkinson's disease tremor. Technol Health Care 2016; 24(2):185–191.
54. Kostikis N, Hristu-Varsakelis D, Arnaoutoglou M, Kotsavasiloglou C. A smartphone-based tool for assessing parkinsonian hand tremor. IEEE J Biomed Health 2015;19(6):1835–1842.
55. Ahlrichs C, Samà A. Is "frequency distribution" enough to detect tremor in PD patients using a wrist worn accelerometer? Paper presented at: Proc 8th Int Conf Pervasive Comput Technologies Healthc 2014;65–71.
56. Veluvolu KC, Ang WT. Estimation of physiological tremor from accelerometers for real-time applications. Sensors 2011;11(3): 3020–3036.
57. Rigas G, Tzallas AT, Tsipouras MG, et al. Assessment of tremor activity in the Parkinson's disease using a set of wearable sensors. IEEE Trans Inf Technol Biomed 2012;16(3):478–487.
58. Patel S, Lorincz K, Hughes R, et al. Monitoring motor fluctuations in patients with Parkinson's disease using wearable sensors. IEEE Trans Inf Technol Biomed 2009;13(6):864–873.
59. San-Segundo R, Zhang A, Cebulla A, et al. Parkinson's disease tremor detection in the wild using wearable accelerometers. Sensors 2020;20(20):5817.
60. Battista L, Romaniello A. A wearable tool for selective and continuous monitoring of tremor and dyskinesia in Parkinsonian patients. Parkinsonism Relat Disord 2020;77:43–47.
61. Papadopoulos A, Kyritsis K, Bostanjopoulou S, et al. Multiple-instance learning for in-the-wild Parkinsonian tremor detection. Annu Int Conf IEEE Eng Med Biol Soc 2019;2019:6188–6191. https://doi.org/10.1109/EMBC.2019.8856314
62. Zheng X, Campos AV, Ordieres-Meré J, et al. Continuous monitoring of essential tremor using a portable system based on smartwatch. Front Neurol 2017;8:96.
63. García-Magariño I, Medrano C, Plaza I, Oliván B. A smartphone-based system for detecting hand tremors in unconstrained environments. Pers Ubiquitous Comput 2016;20(6):959–971.
64. Cole BT, Roy SH, Luca CJD, Nawab SH. Dynamical learning and tracking of tremor and dyskinesia from wearable sensors. IEEE Trans Neural Syst Rehabil Eng 2014;22(5):982–991.
65. Tzallas AT, Tsipouras MG, Rigas G, et al. PERFORM: a system for monitoring, assessment and management of patients with Parkinson's disease. Sensors 2014;14(11):21329–21357.
66. Cole BT, Ozdemir P, Nawab SH. Dynamic SVM detection of tremor and dyskinesia during unscripted and unconstrained activities. 2012 Annu Int Conf IEEE Eng Medicine Biology Soc; 2012: 4927–4930.
67. Roy SH, Cole BT, Gilmore LD, et al. Resolving signal complexities for ambulatory monitoring of motor function in Parkinson's disease. Annu Int Conf IEEE Eng Med Biol Soc 2011;2011:4836–4839.
68. Vial F, Kassavetis P, Merchant S, Haubenberger D, Hallett M. How to do an electrophysiological study of tremor. Clin Neurophysiol Pract 2019;4:134–142.
69. Cagnan H, Little S, Foltynie T, et al. The nature of tremor circuits in parkinsonian and essential tremor. Brain 2014;137(12):3223–3234.
70. 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.
71. Brittain J-S, Cagnan H, Mehta AR, et al. Distinguishing the central drive to tremor in Parkinson's disease and essential tremor. J Neurosci 2015;35(2):795–806.
72. Reis C, Arruda BS, Pogosyan A, Brown P, Cagnan H. Essential tremor amplitude modulation by median nerve stimulation. Sci Rep 2021;11(1):17720.
73. Raethjen J, Lauk M, Köster B, et al. Tremor analysis in two normal cohorts. Clin Neurophysiol 2004;115(9):2151–2156.
74. Deuschl G, Lauk M, Timmer J. Tremor classification and tremor time series analysis. Chaos 1995;5(1):48–51.
75. Pedrosa DJ, Brown P, Cagnan H, Visser-Vandewalle V, Wirths J, Timmermann L, Brittain J-S. A functional micro-electrode mapping of ventral thalamus in essential tremor. Brain 2018;223(Pt 3): 50–11.
76. Schnitzler A, Münks C, Butz M, Timmermann L, Gross J. Synchronized brain network associated with essential tremor as revealed by magnetoencephalography. Mov Disord 2009;24(11):1629–1635.
77. Dirkx MF, Zach H, van Nuland AJ, Bloem BR, Toni I, Helmich RC. Cognitive load amplifies Parkinson's tremor through excitatory network influences onto the thalamus. Brain 2020; 143(5):1498–1511.
78. Dirkx MF, Zach H, Bloem BR, Hallett M, Helmich RC. The nature of postural tremor in Parkinson disease. Neurology 2018; 90(13):e1095–e1103.
79. van Brummelen EMJ, Ziagkos D, de Boon WMI, et al. Quantification of tremor using consumer product accelerometry is feasible in patients with essential tremor and Parkinson's disease: a comparative study. J Clin Mov Disord 2020;7(1):4.
80. Albanese A, Bonuccelli U, Brefel C, et al. Consensus statement on the role of acute dopaminergic challenge in Parkinson's disease. Mov Disord 2001;16(2):197–201.
81. Zach H, Dirkx MF, Pasman JW, et al. Cognitive stress reduces the effect of levodopa on Parkinson's resting tremor. CNS Neurosci Ther 2017;23(3):209–215.
82. Shahtalebi S, Atashzar SF, Samotus O, et al. PHTNet: characterization and deep Mining of Involuntary Pathological Hand Tremor using recurrent neural network models. Sci Rep 2020;10(1):2195.
83. Skaramagkas V, Andrikopoulos G, Kefalopoulou Z, Polychronopoulos P. A study on the essential and Parkinson's arm tremor classification †. Signals 2021;2(2):201–224.
84. Vescio B, Quattrone A, Nisticò R, et al. Wearable devices for assessment of tremor. Front Neurol 2021;12:680011.
85. Elble RJ, McNames J. Using portable transducers to measure tremor severity. Tremor Other Hyperkinet Mov 2016;6:375.
86. Elble RJ. Gravitational artifact in accelerometric measurements of tremor. Clin Neurophysiol 2005;116(7):1638–1643.
87. Sprdlik O, Hurak Z, Hoskovcova M, Ruzicka E. Tremor analysis by decomposition of acceleration into gravity and inertial acceleration using inertial measurement unit. Paper presented at: 2009 9th International Conference on Information Technology and Applications in Biomedicine 2009;1:1–4.
88. Nair P, Trisno R, Baghini MS, et al. Predicting early stage drug induced Parkinsonism using unsupervised and supervised machine learning. Annu Int Conf IEEE Eng Med Biol Soc 2020;00: 776–779.
89. Timmer J, Lauk M, Häußler S, et al. Cross-spectral analysis of tremor time-series. Int J Bifur Chaos 2000;10(11):2595–2610.
90. 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.
91. Eskofier BM, Lee SI, Daneault J-F, et al. Recent machine learning advancements in sensor-based mobility analysis: deep learning for Parkinson's disease assessment. Annu Int Conf IEEE Eng Med Biol Soc 2016;2016:655–658.
92. Barth J, Klucken J, Kugler P, et al. Biometric and Mobile gait analysis for early diagnosis and therapy monitoring in Parkinson's disease. Annu Int Conf IEEE Eng Med Biol Soc 2011;2011:868–871.
93. Perumal SV, Sankar R. Gait and tremor assessment for patients with Parkinson's disease using wearable sensors. ICT Express 2016;2(4):168–174.
94. Peres LB, Calil BC, da Silva APSPB, et al. Discrimination between healthy and patients with Parkinson's disease from hand resting

activity using inertial measurement unit. Biomed Eng Online 2021; 20(1):50.

95. Zheng X, Vieira A, Marcos SL, et al. Activity-aware essential tremor evaluation using deep learning method based on acceleration data. Parkinsonism Relat Disord 2019;58:17–22.

96. Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput 1997;9(8):1735–1780.

97. Cho K, Merrienboer B van, Bahdanau D, Bengio Y. On the properties of neural machine translation: encoder–decoder approaches. Paper presented at: Proc Ssst-8 Eighth Work Syntax Semant Struct Statistical Transl 2014;103–111. https://doi.org/10.48550/arXiv.1409.1259

98. Linardatos P, Papastefanopoulos V, Kotsiantis S. Explainable AI: a review of machine learning interpretability methods. Entropy 2020; 23(1):18.

99. Reinfelder S, Hauer R, Barth J, et al. Timed Up-and-Go phase segmentation in Parkinson's disease patients using unobtrusive inertial sensors. Annu Int Conf IEEE Eng Med Biol Soc 2015;2015:5171–5174.

100. Moon S, Song H-J, Sharma VD, et al. 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 2020;17(1):125.

101. Locatelli P, Alimonti D, Traversi G, Re V. Classification of essential tremor and Parkinson's tremor based on a low-power wearable device. Electronics 2020;9(10):1695.

102. Yusuf M, Atal I, Li J, et al. Reporting quality of studies using machine learning models for medical diagnosis: a systematic review. BMJ Open 2020;10(3):e034568.

103. Louis ED, Bain PG. Problems and controversies in tremor classification. J Neurol Sci 2022;435:120204.

104. Kägi G, Bhatia KP, Tolosa E. The role of DAT-SPECT in movement disorders. J Neurol Neurosurg Psychiatry 2010;81(1):5–12.

105. Hvizdošová L, Nevrlý M, Otruba P, et al. The prevalence of dystonic tremor and tremor associated with dystonia in patients with cervical dystonia. Sci Rep 2020;10(1):1436.

106. Albanese A, Sorbo FD. Dystonia and tremor: the clinical syndromes with isolated tremor. Tremor Other Hyperkinet Mov 2016;6:319.

107. Erro R, Rubio-Agusti I, Saifee TA, et al. Rest and other types of tremor in adult-onset primary dystonia. J Neurol Neurosurg Psychiatry 2014;85(9):965–968.

108. Zach H, Dirkx MF, Roth D, et al. Dopamine-responsive and dopamine-resistant resting tremor in Parkinson disease. Neurology 2020;95(11):e1461–e1470.

109. Egger J, Gsaxner C, Pepe A, et al. Medical deep learning – a systematic meta-review. Comput Meth Prog Bio 2022;221:106874.

110. Oh J, Makar M, Fusco C, et al. A generalizable, data-driven approach to predict daily risk of Clostridium difficile infection at two large academic health centers. Infect Control Hosp Epidemiol 2018;39(4):425–433.

111. Lenert MC, Matheny ME, Walsh CG. Prognostic models will be victims of their own success, unless…. J Am Med Inform Assoc 2019;26(12):1645–1650.

112. Wiens J, Saria S, Sendak M, et al. Do no harm: a roadmap for responsible machine learning for health care. Nat Med 2019;25(9): 1337–1340.

113. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science 2019;366(6464):447–453.

114. Varoquaux G, Cheplygina V. Machine learning for medical imaging: methodological failures and recommendations for the future. NPJ Digit Med 2022;5(1):48.

115. Thudumu S, Branch P, Jin J, Singh J. A comprehensive survey of anomaly detection techniques for high dimensional big data. J Big Data 2020;7(1):42.

116. Fulcher BD, Little MA, Jones NS. Highly comparative time-series analysis: the empirical structure of time series and their methods. J R Soc Interface 2013;10(83):20130048.

117. Kelly CJ, Karthikesalingam A, Suleyman M, et al. Key challenges for delivering clinical impact with artificial intelligence. BMC Med 2019;17(1):195.