Visual perceptive deep learning for smartphone video-based tremor analysis: VIPER-Tremor

Maximilian Friedrich (mfriedrich1@bwh.harvard.edu)
Brigham and Women's Hospital
Anna-Julia Roenn
University Hospital Wuerzburg
Chiara Palmisano
University Hospital Wuerzburg
Jane Alty
University of Tasmania
Steffen Paschen
Christian-Albrechts-University
Guenther Deuschl
University Hospital Kiel
Chi Wang Ip
University Hospital of Würzburg https://orcid.org/0000-0003-0484-385X
Jens Volkmann
University Hospital Wuerzburg
Muthuraman Muthuraman
University Hospital Würzburg https://orcid.org/0000-0001-6158-2663
Robert Peach
University Hospital Wuerzburg
Martin Reich
University Hospital Würzburg

Article

Keywords: Tremor analysis, movement disorders, computer vision, machine learning, deep learning, digital medicine, essential tremor

Posted Date: December 2nd, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3692906/v1

Additional Declarations: (Not answered)

Abstract

Background: Tremor is one of the most common neurological symptoms. Its clinical and neurobiological complexity necessitates novel approaches for deep and granular phenotyping. Instrumented neurophysiological analyses have proven useful for clinical management, but are highly resource-intensive and lack broad accessibility. Simplified bedside scores, on the other hand, lack the granularity to capture subtle but relevant tremor features. Addressing this gap, we develop a deep learning framework for the quantitative assessment of limb tremor utilizing only standard clinical videos. Methods: We engineer a visual perceptive limb tremor analysis tool based on Mediapipe, a convolutional neural network architecture for marker-less hand tracking: VIPER-Tremor. We validate it against gold standard methods, including marker-based motion capture, wrist-worn accelerometery, and clinical scoring across two independent clinical cohorts encompassing a total of 66 patients diagnosed with essential tremor and recorded in different therapeutic states of deep brain stimulation. Results: Computer vision-derived tremor metrics exhibit high convergent clinical validity to scores (Spearman's rho= 0.55 – 0.86, p≤ .01) as well as an accuracy of up to 2.60mm and ≤0.21Hz for tremor amplitude and frequency measurements, matching gold-standard equipment. VIPER-Tremor is capable of extracting advanced tremor features relevant for differential diagnosis and enables therapeutic outcome prediction, a dimension which conventional tremor scores were unable to provide. Conclusion: VIPER-Tremor is an accurate, unbiased and highly accessible solution for smartphone video-based tremor analysis and yields comparable results to gold standard recordings. VIPER-Tremor presents a significant advancement in tremor analysis, combining accuracy and accessibility, and promises to be a pivotal tool in the emerging field of precision neurology, enhancing diagnostic and therapeutic approaches.

Introduction

Tremor syndromes are among the most common neurological disorders. Of these, essential tremor affects up to 4.6% of the global population ≥65 years old[1]. This disorder is characterized by a mixture of postural and kinetic tremors, which likely represent diverse facets of pathological oscillations in brain motor networks[2–5]. Tremor is often accompanied by additional neurological signs such as dystonia or ataxia; as such, tremor is also a common symptom in a range of acquired and genetic neurological disorders, posing a significant diagnostic challenge in clinical neurology. This translates into high rates of misdiagnosed tremor disorders[6], which has profound therapeutic implications for deep brain stimulation (DBS), a potent neural circuit intervention for tremor disorders. DBS outcomes largely hinge on accurate patient selection, which itself is influenced by accurate tremor assessment[7]. The considerable clinical and neurobiological heterogeneity of tremor syndromes has been a roadblock to pathogenetic and diagnostic research, which culminated in a call to partly redefine tremor classification by means of multidimensional, deep phenotyping[5].

To this end, instrumented tremor analysis offers an unbiased and detailed assessment of key tremor features, such as frequency and amplitude, which are crucial for phenotyping[8,9], therapeutic

monitoring[10,11], differential diagnosis[12–15] and closed-loop neuromodulation[16]. However, the reliance on complex and resource-intensive methods like multi-camera 3D motion capture and combined accelerometery restricts its practical use, especially in routine clinical settings.

In contemporary practice, the complex phenomenology of tremor syndromes is therefore condensed into low dimensional, ordinal rating scales[33]. These scales represent tremor items in a non-linear, logarithmic manner[17,18] and, despite their simplicity, suffer from considerable clinimetric limitations. One of these limitations is interrater reliability, reported to be as low as 0.1 (Cohen's kappa)[17,19–22].

While mobile technologies, such as smartphone accelerometers, have emerged as promising tools for tremor frequency assessment[3,8,23–25], they have critical limitations, such as their reliance on calibration, sensor weight, and placement[9]. Additionally, they cannot measure associated neurological signs such as dystonia and allow only indirect approximations of tremor amplitude, which, in contrast to frequency, is the key kinematic determinant of patient life quality[26].

Novel visual perceptive methods based on convolutional neural networks (CNNs) for marker-less pose tracking hold a transformative potential for neurological phenomenology and especially, movement disorders[27–31]. These methods are compatible with consumer-grade hardware, like smartphones or web cameras, and can unlock valuable data from legacy medical videography[27,32]. Yet, the clinical applicability of video pose tracking remains to be explored, particularly in medical settings where visual confounders can adversely affect performance[27,28,33,34]. While small-scale pilot studies have shown the feasibility of computer vision-based (CV) analysis of neurological symptoms[22,27,35,36], rigorous validation and clinical application in larger patient populations and clinical settings are scarce[37,38].

This study introduces a pioneering visual perceptive deep learning framework, designed to address these challenges by enabling comprehensive tremor analysis using through advanced computer vision techniques using only clinical standard videos. To this end, we systematically evaluate the capability of a visual perceptive deep learning framework to track hands and extract established tremor features from clinical video data showcasing postural and kinetic tremor assessments. Our first objective is to benchmark this framework against gold standard instrumented methods in a cohort of patients diagnosed with essential tremor. Subsequently, we apply the framework to a retrospective dataset of unstandardized, real-world videos sourced from two clinical sites, examining its convergent clinical validity and capability to characterize therapeutic effects of deep brain stimulation on tremor. Finally, we assess the framework's utility to inform differential diagnostic and prognostic challenges in two use case scenarios inspired by clinical tremor management.

Methods

Ethics approval

This study was conducted in accordance with the Declaration of Helsinki. Ethics approval was obtained from the Julius-Maximilians University Wuerzburg's ethics committee (#283/14 and 163/14_MP).

Study design and cohorts

The study consists of two independent phases with independent cohorts. This design was chosen to reflect best practices in machine learning, aiming to ensure validity, generalizability and reproducibility (Figure 1). The prospective cohort consisted of n= 8 patients (age 54-83, mean, 4 males) and the retrospective cohort of n= 58 patients (age at surgery 29 – 83 years, 31 males) with essential tremor diagnosed according to the Movement Disorder Society's consensus criteria[5] (detailed cohort characterization in supplementary methods).

Prospective experimental design

Experiments were conducted at the department of Neurology, University Hospital Wuerzburg. Participants underwent standardized assessment of postural (>30 seconds holding arms in front of chest, "wing-beating position", fingertips facing each other but not touching) and kinetic tremor (five repetitions of finger-to-nose pointing per side, each starting from a resting position of the laterally outstretched arm). Tremor assessments were recorded in a 2x2 block design with DBS (on/off) and method (video/motion capture & accelerometery) as intraindividual factors. Minimal DBS washout period after impulse generator deactivation was conservatively set to 45 minutes to exclude stimulation carry over effects[39,40]. DBS ON trials were conducted using the individual's best clinical stimulation settings. Experimental blocks were pseudorandomized to reduce systematic biases. Based on the video material, the corresponding items of the Fahn-Tolosa-Marin tremor rating scale (postural and kinetic tremor amplitudes) were annotated by a clinician expert in movement disorders blinded to the experimental condition (MMR).

Tremor recording

Hand tracking setup

Participants were seated on a chair in front of a neutral background. Tremor assessment was videotaped using a standard smartphone camera (Samsung Galaxy S20, Samsung, Seoul, South Korea), operating at a spatiotemporal resolution of 1920x1080 px and 60 Hz. The camera was mounted on a standard tripod in landscape mode at a viewing distance of 3 meters to cover the full body of the participants centrally in the video frame throughout the recording time. To avoid obscuring anatomical landmarks, participants were asked to wear sleeveless tops exposing the shoulders and arms. Watches or other jewellery were removed or covered with tape to prevent any interference with the limb tracking, e.g., through aberrant reflections. For videos, pixel-to-metric conversion was derived using a "ChArUco" board (a checkerboard with additional geometric shapes of known metric dimensions for calibration), which was presented

before each new video run, as previously described[27]. Motion capture markers and accelerometers significantly change the visual appearance of hands, which impacts computer vision tracking performance and reduces external validity in non-instrumented settings. Hence, motion capture combined with accelerometery and computer vision recordings were taken separately. Details of motion capture and accelerometery setup are outlined in supplementary methods.

Computer vision tremor analysis workflow

Mediapipe hands

For video-based hand tracking, we utilized a powerful and widely used computer vision and pose tracking framework, Mediapipe[30,41] (MP). To this end, the Mediapipe PyPI package was executed in Python Version 3.9 and the respective hand landmark detection model applied to the video dataset which loaded using OpenCV[42]. Based on Mediapipe's internal computation of "world referenced landmarks", no further calibration step was needed and the coordinate time series of the 21 landmarks per hand were exported for subsequent calculation of tremor characteristics.

Tuning a tremor-specific convolutional neural network

Additionally, a residual convolutional neural network was fine-tuned using DeepLabCut[43,44] to track 29 upper body landmarks from diverse clinical videos (henceforth DLC-RCNN). 1489 frames were extracted from 202 retrospective clinical videos and labelled, with a subset reserved for out-of-sample validation. Model performance was evaluated in a multi-faceted approach as previously described[27,29] (Supplementary Figure 1).

Calculation of tremor characteristics

Coordinate time series outputs from computer vision frameworks, marker tracks and accelerations were extracted and analyzed using a harmonized analytical pipeline set up in Python 3.9. Briefly, preprocessing involved applying bandpass filters to isolate tremor frequencies, with the middle and index fingers serving as focal tracking points for postural and kinetic tremor, respectively. Individual scaling factors for pixel-to-metric conversion were derived from facial ground truth dimensions derived from individual cranial MR-images or ChArUco boards. Tremor characteristics were quantified through spectrogram analysis and power spectral density calculations, with results aggregated into mean and peak values for various experimental conditions (see supplementary materials for further details).

Statistical methods

Normality of datasets was examined using the Shapiro Wilk test and additional inspection of quartile ("Q-Q") plots to inform the appropriate display of data distributions and the selection of subsequent contrast tests. In case of a significant deviation of the (log-)normality assumption, non-parametric tests, i.e., Wilcoxon rank-sum test and matched rank biserial correlation were used. Linear relationships were examined using and Pearson or Spearman's rank correlations. When appropriate, outliers were removed using the robust regression and outlier removal (ROUT) method with balanced coefficient of Q = 1%[45]. Leave-one-out cross-validation (LOOCV) was performed using Python and scikit-learn's LogisticRegression class. Matplotlib and seaborn were utilized for visualization of performance and feature importance evaluation, as outlined in supplementary materials.

Equivalence testing

To systematically compare the accuracy and precision of different computer vision approaches for tremor amplitude quantification, relative deviations to technical gold standard measurements were computed and compared using the "two one-sided t-test" (TOST) method[46,47], as previously described[27]. The derivation of equivalence boundaries is further detailed in the supplementary materials.

Statistical computations were conducted using Python 3, JAMOVI Version 2.2.5[48], R Studio[49] and GraphPad Prism Version 9[50].

Results

Validation of the visual perceptive framework

To assess the visual perceptive framework's technical and clinical validity, we first applied it to video data from a prospectively recruited cohort of patients with a diagnosis of essential tremor and treated with thalamic DBS. For reference, we recorded Fahn-Tolosa Marin tremor scores as well as ground truth tremor amplitudes and frequencies using laboratory gold standard technologies: marker-based motion capture and simultaneous wrist-mounted accelerometery.

Amplitude analyses – postural tremor

Computer vision-derived peak postural tremor amplitudes showed strong correlation with clinical scores, similarly to gold standard motion capture (MP: rho> 0.86, MC: rho= 0.90, p< .001, Figure 1a-b). Excellent agreement of computer vision was found with motion capture (rho= 0.89, p< .001, Figure 1c). In comparison to motion capture, computer vision had a mean absolute error of 10 mm (95% CI [5.65, 14.4]) and no systematic relationship between measurement and error magnitudes was observed (Figure 1d). Computer vision-derived tremor amplitudes fell within equivalence boundaries of motion capture tracking (±10mm, Supplementary Figure 2a) and were comparably responsive to DBS effect (d > 0.94, all p < .001, Figure 1e), overall suggestive of equivalent accuracy. Median precision, measured by the standard

deviation of each amplitude measurement, was 1.29 mm for motion capture and 0.54 mm for Mediapipe; precision values reached equivalence to motion capture within gold-standard derived boundaries of ±3.63 mm (Supplementary Figure 2b). Reducing the 90% CI margins to ±1.5 and ±1.0 mm did not substantially change these results, indicating robustness beyond the defined boundaries.

Amplitude analyses – kinetic tremor

Mediapipe's peak kinetic tremor amplitude estimates were strongly correlated to the clinical scores, again comparable to motion capture derived values (rho= 0.55, p< .01, Figure 1f-g). Mediapipe reached substantial agreement with motion capture (rho= 0.72, p< .001, Figure 1h). Mean absolute error was -2.60 mm (95% CI [-3.13, 8.23], Figure 1i). Mediapipe's accuracy in kinetic tremor amplitude measurement was equivalent to motion capture (Supplementary Figure 2a). Mediapipe and motion capture were again comparably responsive to DBS effects on kinetic tremor amplitude (d= 0.69 and 0.60, figure 1j). Median precision of kinetic tremor amplitude measurement was calculated to be 0.31 mm for motion capture and 0.49 mm for Mediapipe. Mediapipe's precision fell within the equivalence boundaries of ±2.1 mm (Supplementary Figure 2c). Repeating the equivalence tests with empirically reduced 90% CI margins of ±1.5 and ±1.0 mm did not substantially change these results.

Notably, the aforementioned results were similar when using *mean* instead of *peak* amplitude measurements (Supplementary Figures 3 & 4).

Frequency measurements – postural tremor

Computer vision-derived tremor frequency measurements were validated against wrist-worn accelerometery, a clinical and laboratory gold standard for tremor analysis. The correspondence of tremor frequencies from computer vision and motion capture to accelerometery was found to be similarly strong (r> 0.40, Figure 3a). The mean dominant frequency of postural tremor was measured to be 5.7 ± 0.72 Hz with accelerometery, 6.04 ± 0.65 Hz with motion capture, and 5.9 ± 0.58 Hz with Mediapipe, resulting in mean absolute errors of -0.34 Hz [95% CI -0.08, 0.60] for motion capture and -0.21 Hz [95% CI -0.05, 0.46] for Mediapipe (Figure 3b).

Within the predefined margins of ±0.5 Hz, Mediapipe-derived frequency measurements achieved equivalent accuracy to accelerometery, while motion capture exceeded the equivalence bounds (Figure 3c and supplementary figure 2d). Median precision of tremor frequency measurements was 0.58 Hz for accelerometery, 1.15 Hz for motion capture and 1.12 Hz for Mediapipe. Precision values from motion capture and Mediapipe were equivalent to accelerometer within gold standard derived margins of ±2 Hz (Supplementary Figure 2e). Again, reducing the 90% CI margins to ±1.5 and ±1.0 Hz did not substantially alter these results.

Frequency measurements – kinetic tremor

Both motion capture and Mediapipe-derived kinetic tremor frequencies demonstrated moderate agreement with accelerometric measurements (motion capture: rho= 0.38, p= .034; Mediapipe: rho= 0.37, p= 0.033, Figure 3c). The mean dominant frequency of kinetic tremor was 5.25 ± 1.06 Hz using accelerometery, 5.48 ± 0.41 Hz using motion capture, and 5.31 ± 0.34 Hz using Mediapipe, with mean absolute errors of 0.22 Hz (95% CI [-0.15, 0.59]) for motion capture and 0.06 Hz (95% CI [-0.30, 0.41]) for Mediapipe. Bland-Altman plots for motion capture and Mediapipe suggested a systematic relationship between error and measurement magnitudes (Figure 3d).

Within the predefined boundaries of ±0.5 Hz, Mediapipe's accuracy in frequency measurements was equivalent to accelerometery (Figure 3e and supplementary figure 2d). In contrast, motion capture's accuracy was significantly lower than Mediapipe (T(31)= 2.98, 95% CI of difference [0.05, 0.28], p= .006, not shown). Median precision was 0.58 Hz for accelerometry, 0.14 Hz for motion capture, and 0.1 Hz for Mediapipe. Both motion capture and Mediapipe precision values fell within equivalence boundaries derived from the minimal precision achieved by accelerometery, ±2 Hz (Supplementary Figure 2f). Reducing the margins to ±1.5 and ±1.25 Hz in equivalence tests did not substantially alter these results.

Retrospective application

Postural tremor

In order to clinically validate the visual perceptive framework in an independent sample, we applied it to clinical videos of 43 individuals undergoing clinical tremor assessment before and after thalamic DBS implantation. The peak postural tremor amplitudes derived from Mediapipe were strongly correlated with the corresponding tremor scores (Figure 4a). Wilcoxon testing further revealed that the computer vision framework’s peak amplitude measurements were highly responsive to the effect of DBS (Figure 3b-c). Repeating the analyses using *mean* instead of *peak* tremor amplitudes yielded similar results with respect to score correlation (Supplementary Figure 5b-c). Mean dominant frequency of postural tremor was calculated to be 5.96 ± 0.76 Hz (Supplementary Figure 5d).

Kinetic tremor

In the 25 available individuals, a moderate correlation was found between the measured peak amplitudes and the corresponding tremor scores (Figure 4d). Wilcoxon testing revealed that peak kinetic tremor amplitude measurements were highly sensitive to the effect of VIM-DBS (Figure 3e-f). Repeating the analyses using mean instead of peak amplitudes yielded similar results (Supplementary Figure 5e-f). The mean dominant frequency of kinetic tremor was calculated to be 5.75 ± 0.58 Hz.

VIPER-Tremor in clinical use cases

Advanced diagnostic tremor features

Instrumented tremor analysis can provide valuable differential diagnostic clues for tremor syndromes. Beyond the basic tremor characteristics like amplitude and frequency, advanced features such as harmonics or inter-limb tremor coherence have previously been established to support differential diagnosis of tremor syndromes[11]. To this end, we investigated whether the visual perceptive framework is capable of extracting advanced diagnostic tremor features, which usually require electromyography or other sensors.

Indeed, the Mediapipe-derived tremor signal displayed a harmonic peak which was located at twice the mean dominant frequency (Figure 5a-b), a feature previously reported to differentiate essential from parkinsonian tremor[51]. Moreover, no significant inter-limb tremor coherence was detected, a feature reported to discern essential tremor from orthostatic tremor[11,52] (Figure 5b-c).

Predictive modelling

Albeit efficacious in the majority of cases, thalamic DBS outcomes vary[10]. Lack of tremor improvement or even paradoxical *increases* in kinetic tremor amplitude signify a poor DBS outcome[7]. Patient-specific factors such as baseline clinical tremor scores have been shown to aid DBS outcome prognostication across tremor disorders[53], which facilitate patient counselling.

Therefore, we aimed to assess the utility of computer vision-derived metrics in predicting DBS outcomes from preoperative kinematics and clinical score information. Since persisting kinetic tremor is a key driver of functional disability in essential tremor and among the main reasons for failed DBS interventions[54,55], we binarized our patient cohort into good and poor responders based on post-operative kinetic tremor. We chose a threshold of $\geq$2 cm residual kinetic tremor amplitude and $\leq$30% relative kinetic tremor reduction in DBS ON, so as to identify cases with clinically relevant disability[7,54,55]. Applying this threshold, we found that baseline kinetic tremor was more frequently associated with a poor outcome (55% vs. 21% fraction of poor responders, $p < .001$, Figure 5b-c).

To identify determinants of suboptimal DBS outcomes, which might assist in preoperative patient counselling, we conducted a logistic regression analysis. Using binarized response group as the outcome variable and preoperative limb kinematic features as covariates, we detected a strong and significant association of preoperative tremor measurements to DBS outcomes ($\chi 2 = 58.4$, $p < .001$, McFadden $R^2 = 0.65$). Among all covariates, baseline kinetic tremor amplitude emerged as a significant and independent predictor of DBS response ($p = .002$, OR 0.89, 95% CI [0.82, 0.96]). Implementing a rigorous leave-one-out cross-validation to evaluate the model's performance yielded an area under the receiver operator curve of 0.88 and a F1-score of 0.89 (Figure 5d). Moreover, baseline kinetic tremor amplitude emerged as an independent predictor of DBS-associated improvement of kinetic tremor amplitude in a linear regression model ($R^2 = 0.18$, $p < .001$; baseline kinetic tremor: $p = .021$, Figure 5e). Of note, preoperative tremor scores were neither a significant predictor of binary outcome nor tremor amplitude change.

Assessment of a disease-specific convolutional network: DLC-RCNN

In real-world clinical contexts, generally trained pose tracking algorithms such as Mediapipe have been found to suffer from performance issues. This is ascribed to the large out-of-domain variance introduced in clinical settings, such as dressings, obscuration or clutter[27,34,56]. To gauge this effect's relevance in the context of tremor, we additionally developed a tremor-specific residual convolutional neural network using DeepLabCut[43]: DLC-RCNN[33,57]. Briefly, the RCNN was trained with >120,000 frames of clinical video material. Final performance evaluation showed a median Euclidean distance of 3.56 mm and 10.74 mm between user-annotated and predicted keypoints, demonstrating acceptable generalization and tracking accuracy related to fingertip size (occupying 10-20 pixels, corresponding to 10-20 mm on average[27,44], Supplementary Figure 1). The model's generalization to an out-of-sample validation dataset (>15,000 frames) showed high confidence in predicting postural tremor keypoints (median likelihood of 0.99, 4,884 predictions) but unacceptably low confidence for kinetic tremor keypoints (median likelihood of 0.22, 10,532 predictions, Supplementary Figure 1b). Therefore, DLC-RCNN could only be used for postural tremor analysis.

In the prospective cohort, DLC-RCNN-derived tremor amplitudes were strongly correlated to clinical scores (rho= 0.92, p< .001) and gold standard motion capture (rho= 0.88, p< .001). The mean absolute error was 2.55 mm. DLC-RCNN-derived mean dominant tremor frequencies were moderately correlated to accelerometer (rho= 0.44, p< .05), with a mean absolute error of -0.69 Hz. In the retrospective cohort, DLC-RCNN-derived postural tremor amplitudes were moderately to strongly correlated with assigned clinical scores (rho= 0.72, p= .001). DLC-RCNN's accuracy and precision (0.66 mm) for amplitudes was equivalent to motion capture. Mean dominant frequency was calculated to be 6.38 ± 0.54 Hz, but the DLC-RCNN's frequency accuracy was significantly lower than Mediapipe and motion capture, hence not equivalent (Supplementary Figure 2).

Discussion

Tremor disorders, owing to their intricate clinical and neurobiological nature, underscore the critical need for deep phenotyping in effective clinical management. While traditional instrumented methods provide valuable insights, their high resource demands significantly limit their widespread application in clinical settings. As a result, clinicians often rely on a more reductionist approach, employing semi-quantitative rating scales. Although quick and practical, this method offers only a rudimentary representation of the complex spectrum of tremor phenomenology. Such oversimplification translates to considerable clinimetric limitations, impacting the accuracy and reliability of tremor assessment and, consequently, patient care[17,19,20]. In response to these challenges, we have developed VIPER-Tremor, a visual perceptive framework leveraging robust pose tracking algorithms. This framework is specifically designed for the comprehensive evaluation of postural and kinetic limb tremor in clinical video recordings. VIPER-Tremor underwent extensive validation against state-of-the-art instrumented methods and clinical scoring

systems, showcasing its accuracy and robustness. VIPER-Tremor demonstrated its practical utility not only in characterizing the effects of deep brain stimulation but also in providing valuable insights into diagnostic and prognostic challenges – aspects that conventional scores failed to capture. Finally, our study elucidated the impact of different algorithmic architectures on clinical pose tracking, thereby providing a roadmap for future technical scalability.

A deep learning framework for hand tracking and tremor analysis

The results of our prospective validation underline the framework's accuracy, precision, and clinical validity, which largely match gold standard equipment. While prior studies have tapped into the potential of computer vision for tremor detection[58,59] and frequency extraction[23], amplitude quantification remained largely unexplored. Yet, tremor amplitude is pivotal in assessing patient disability and therapeutic outcomes[7,10]. Our findings indicate that smartphone videos, coupled with computer vision tracking tools, can gauge tremor amplitude with an accuracy of up to 2.6mm, a value that falls on the low end of reported pose tracking accuracies[31,33] and that is almost an order of magnitude smaller than the lowest anchor value provided in the tremor rating scale (20mm). Compared to gold standard accelerometery, computer vision-derived tremor frequency measurements demonstrated a mean absolute error between -0.06 and -0.21Hz, values falling well within, if not below modern vision-based frameworks[23]. More generally, large scale studies investigating clinical pose tracking in other movement disorders[37,38] report moderate to high score correlation strengths in the range of 0.6 - 0.8, which corresponds closely to our reported values of 0.55 - 0.86. Overall, this is strongly indicative that the visual perceptive framework effectively captured the clinically relevant target information. Notably, some correlation plots exhibit increasing residuals with higher scores, which is well in line with the notion of a logarithmic rather than linear relationship of tremor severity and ordinal scores[17]. Continuous digital biomarkers are not subject to such non-linearity, which often complicates both intra- and interindividual comparisons which are relevant for clinical studies and management.

Therefore, our framework dramatically simplifies tremor analysis by eliminating the need for multiple devices and sensors and even enabling the analysis of unstandardized legacy videos, underscoring its generalizability and versatility. The fully vision-based approach can be further scaled to additionally quantify tremor-associated neurological signs such as ataxia[60] or dystonia[37]. This capability aligns with the central goals of future deep phenotyping efforts in tremor disorders[5].

Clinical usability

Next, we applied the computer-vision framework in exemplary use cases that are directly inspired by clinical tremor management. First, VIPER-Tremor was capable of extracting advanced diagnostic tremor features, which offer additional insights relevant for the differential diagnosis of tremor

disorders[11,51,52,61]. For example, a harmonic peak at twice the dominant tremor frequency or a lack of inter-limb tremor coherence can be diagnostic clues differentiating essential tremor from other tremor syndromes[11,51,52]. While the scope of this study was not designed to facilitate comparisons across different tremor disorders, our results nonetheless demonstrate the feasibility of using our framework to derive diagnostically relevant tremor features, linking computer vision-derived biomarkers with sensor- or EMG-based findings reported in the neurophysiological literature[11,51,52].

Second, computer vision-derived features could aid in characterizing thalamic neurostimulation outcomes. Our predictive model, focusing on kinetic tremor reduction as the key determinant of disability and life quality after DBS implantation[4,7,10,54,62], identified baseline kinetic tremor amplitude as a predictor of DBS outcome. Interestingly, conventional tremor scores lacked this predictive power, emphasizing the advantages of sensitive and continuously encoded digital biomarkers in capturing such nuanced clinical relationships. This finding aligns with similar results for DBS outcome prediction based on scores in Parkinson's disease[53] as well as emerging evidence for the added value of digital phenotyping in neurological disorders which reaches far beyond conventional scores[27,37,38,63,64].

Deep learning models

Lastly, we found that algorithm selection is paramount for clinical pose tracking. While the tremor-specific model showed excellent performance in postural tremor tracking, it entirely failed to track kinetic tremor, drastically reducing its versatility. We hypothesize that Mediapipe outperformed the disease-specific model due to its 3D pose tracking capability, which is essential for tracking of complex configurational changes, such as hand rotations along the finger-to-nose trajectory[33,65,66]. These findings differ from those in the context of eye movement analysis[27], thereby highlighting the importance of future studies specifically investigating the customization of neural network architectures to meet the complex demands of clinical pose tracking.

Limitations

Several limitations should be acknowledged. First, the visual perceptive algorithm showed diminished performance in kinetic tremor assessments, likely due to hand configurational complexities leading to partial occlusions. Especially in presence of hands re-entering the video frame or excessive zooming and camera movement, significant prediction errors have been observed, which led to a substantial number of excluded videos. We therefore recommend the implementation of basic video quality criteria for prospective clinical pose tracking studies, as used here. Second, our sequential recording strategy, adopted to minimize marker interference which could lead to overly optimistic tracking results, might introduce biological variance in tremor amplitude measurements[67]. This however biases us against overestimation of the framework's accuracy and suggests that the technical agreement between VIPER-Tremor and gold standard might, in fact, be even higher. Lastly, the prospective validation cohort was

relatively small and recruited from a single center. However, application in a large and independent retrospective cohort yielded similarly robust results. Larger, multi-centric studies are needed to further establish the framework's applicability in diverse clinical settings.

In conclusion, we argue that computer vision-derived biomarkers are poised to transform neurological practice, boosting statistical power in clinical studies and heralding a new era of precision neurology[8,9,38,63]. The resulting granular biomarkers offer a powerful, yet highly accessible means to address contemporary challenges in tremor classification towards novel, neurobiologically grounded disease models[5]. With its compatibility for on-device operation, VIPER-Tremor meets the demands for accuracy and efficiency in neurological practice. We envision that digital neurophenotyping will be seamlessly integrated with emerging technological advancements in neuroimaging and neurotechnology[7,68,69], potentially guiding adaptive neurostimulation approaches[16,70] and ultimately, improving patient care in neurology.

Declarations

Conflicts of interest

The authors declare no conflicts of interest.

Data and code availability

Non-identifiable patient data are available upon request to the corresponding author. Mediapipe hands model is openly available (https://developers.google.com/mediapipe/). DeepLabCut code is openly available (https://github.com/DeepLabCut/DeepLabCut). The tremor-specific DLC-RCNN model is available Harvard dataverse: https://doi.org/10.7910/DVN/CRJRJF.

Acknowledgements

The authors thank Prof. David Hogg for helpful remarks.

Funding

MF was supported by the Interdisciplinary Center for Clinical Research (IZKF) Z2-CSP13 at the University Hospital Wuerzburg. JV, MMR and CP are supported by the German Research Foundation (DFG, Project-ID 424778381, TRR 295). CP was supported by the Fondazione Grigioni per il Morbo di Parkinson and the Fondazione Europea di Ricerca Biomedica (FERB Onlus).

Author contributions:

MF: Conceptualization, data analysis, data interpretation, writing and revision of the manuscript, project supervision. AJR: Data collection, data analysis, revision of the manuscript. CP: Data collection, data analysis, revision of the manuscript. JA: data interpretation, revision of manuscript. SP and GD: data collection, data analysis, revision of the manuscript. CWI: computational resources. JV: data collection, data interpretation, revision of manuscript. MM: data analysis, revision of manuscript. RP: data analysis, writing and revision of the manuscript. MMR: Conceptualization, data collection, data interpretation, revision of the manuscript, project supervision.

References

1. Louis, E. D. & Ferreira, J. J. How common is the most common adult movement disorder? Update on the worldwide prevalence of essential tremor. *Mov. Disord.* **25**, 534–541 (2010).

2. Lenka, A. & Jankovic, J. Tremor Syndromes: An Updated Review. *Front. Neurol.* **12**, 684835 (2021).

3. Barrantes, S. *et al.* Differential diagnosis between Parkinson's disease and essential tremor using the smartphone's accelerometer. *PloS One* **12**, e0183843 (2017).

4. Deuschl, G., Raethjen, J., Lindemann, M. & Krack, P. The pathophysiology of tremor. *Muscle Nerve* **24**, 716–735 (2001).

5. Bhatia, K. P. *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. Off. J. Mov. Disord. Soc.* **33**, 75–87 (2018).

6. Jain, S., Lo, S. E. & Louis, E. D. Common Misdiagnosis of a Common Neurological Disorder: How Are We Misdiagnosing Essential Tremor? *Arch. Neurol.* **63**, 1100–1104 (2006).

7. Reich, M. M. *et al.* Progressive gait ataxia following deep brain stimulation for essential tremor: adverse effect or lack of efficacy? *Brain J. Neurol.* **139**, 2948–2956 (2016).

8. Balachandar, A. *et al.* Are smartphones and machine learning enough to diagnose tremor? *J. Neurol.* **269**, 6104–6115 (2022).

9. De, A., Bhatia, K. P., Volkmann, J., Peach, R. & Schreglmann, S. R. Machine Learning in Tremor Analysis: Critique and Directions. *Mov. Disord. Off. J. Mov. Disord. Soc.* (2023) doi:10.1002/mds.29376.

10. Welton, T. *et al.* Essential tremor. *Nat. Rev. Dis. Primer* **7**, 83 (2021).

11. Deuschl, G. *et al.* The clinical and electrophysiological investigation of tremor. *Clin. Neurophysiol. Off. J. Int. Fed. Clin. Neurophysiol.* **136**, 93–129 (2022).

12. Alusi, S. H., Macerollo, A., MacKinnon, C. D., Rothwell, J. C. & Bain, P. G. Tremor and Dysmetria in Multiple Sclerosis: A Neurophysiological Study. *Tremor Hyperkinetic Mov.* **11**, 30.

13. Casamento-Moran, A. *et al.* Quantitative Separation of Tremor and Ataxia in Essential Tremor. *Ann. Neurol.* **88**, 375 (2020).

14. Herzog, J. *et al.* Kinematic analysis of thalamic versus subthalamic neurostimulation in postural and intention tremor. *Brain* **130**, 1608–1625 (2007).

15. Vissani, M. *et al.* Impaired reach-to-grasp kinematics in parkinsonian patients relates to dopamine-dependent, subthalamic beta bursts. *Npj Park. Dis.* **7**, 1–10 (2021).

16. Schreglmann, S. R. *et al.* Non-invasive suppression of essential tremor via phase-locked disruption of its temporal coherence. *Nat. Commun.* **12**, 363 (2021).

17. Elble, R. J. *et al.* Tremor amplitude is logarithmically related to 4- and 5-point tremor rating scales. *Brain* **129**, 2660–2666 (2006).

18. Kremer, N. I. *et al.* Supine MDS-UPDRS-III Assessment: An Explorative Study. *J. Clin. Med.* **12**, 3108 (2023).

19. Stacy, M. A. *et al.* Assessment of interrater and intrarater reliability of the Fahn-Tolosa-Marin Tremor Rating Scale in essential tremor. *Mov. Disord. Off. J. Mov. Disord. Soc.* **22**, 833–838 (2007).

20. Becktepe, J. *et al.* Exploring Interrater Disagreement on Essential Tremor Using a Standardized Tremor Elements Assessment. *Mov. Disord. Clin. Pract.* **8**, 371–376 (2021).

21. Alusi, S. H., Worthington, J., Glickman, S., Findley, L. J. & Bain, P. G. Evaluation of three different ways of assessing tremor in multiple sclerosis. *J. Neurol. Neurosurg. Psychiatry* **68**, 756–760 (2000).

22. Tien, R. N. *et al.* Deep learning based markerless motion tracking as a clinical tool for movement disorders: Utility, feasibility and early experience. *Front. Signal Process.* **2**, (2022).

23. Williams, S. *et al.* Accuracy of Smartphone Video for Contactless Measurement of Hand Tremor Frequency. *Mov. Disord. Clin. Pract.* **8**, 69–75 (2021).

24. van Brummelen, E. M. J. *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.* **7**, 4 (2020).

25. Elble, R. J. & McNames, J. Using Portable Transducers to Measure Tremor Severity. *Tremor Hyperkinetic Mov.* **6**, 375 (2016).

26. Hess, C. W. & Pullman, S. L. Tremor: Clinical Phenomenology and Assessment Techniques. *Tremor Hyperkinetic Mov.* **2**, tre-02-65-365-1 (2012).

27. Friedrich, M. U. *et al.* Smartphone video nystagmography using convolutional neural networks: ConVNG. *J. Neurol.* (2022) doi:10.1007/s00415-022-11493-1.

28. Stenum, J., Rossi, C. & Roemmich, R. T. Two-dimensional video-based analysis of human gait using pose estimation. *PLOS Comput. Biol.* **17**, e1008935 (2021).

29. Knorr, S. *et al.* The evolution of dystonia-like movements in TOR1A rats after transient nerve injury is accompanied by dopaminergic dysregulation and abnormal oscillatory activity of a central motor network. *Neurobiol. Dis.* **154**, 105337 (2021).

30. Güney, G. *et al.* Video-Based Hand Movement Analysis of Parkinson Patients before and after Medication Using High-Frame-Rate Videos and MediaPipe. *Sensors* **22**, 7992 (2022).

31. Stenum, J. *et al.* Applications of Pose Estimation in Human Health and Performance across the Lifespan. *Sensors* **21**, 7315 (2021).

32. Esteva, A. *et al.* Deep learning-enabled medical computer vision. *Npj Digit. Med.* **4**, 1–9 (2021).

33. Seethapathi, N., Wang, S., Saluja, R., Blohm, G. & Kording, K. P. Movement science needs different pose tracking algorithms. Preprint at https://doi.org/10.48550/arXiv.1907.10226 (2019).

34. Cimorelli, A., Patel, A., Karakostas, T. & Cotton, R. J. Portable in-clinic video-based gait analysis: validation study on prosthetic users. 2022.11.10.22282089 Preprint at https://doi.org/10.1101/2022.11.10.22282089 (2022).

35. Williams, S., Fang, H., Relton, S. D., Graham, C. D. & Alty, J. E. Seeing the unseen: Could Eulerian video magnification aid clinician detection of subclinical Parkinson's tremor? *J. Clin. Neurosci. Off. J. Neurosurg. Soc. Australas.* **81**, 101–104 (2020).

36. Williams, S. *et al.* The discerning eye of computer vision: Can it measure Parkinson's finger tap bradykinesia? *J. Neurol. Sci.* **416**, 117003 (2020).

37. Peach, R. *et al.* Quantitative assessment of head movement dynamics in dystonia using visual perceptive deep learning: a multi-centre retrospective longitudinal cohort study. 2023.09.11.23295260 Preprint at https://doi.org/10.1101/2023.09.11.23295260 (2023).

38. Morinan, G. *et al.* Computer vision quantification of whole-body Parkinsonian bradykinesia using a large multi-site population. *Npj Park. Dis.* **9**, 1–12 (2023).

39. Perera, T. *et al.* Deep brain stimulation wash-in and wash-out times for tremor and speech. *Brain Stimul. Basic Transl. Clin. Res. Neuromodulation* **8**, 359 (2015).

40. Cooper, S. E., McIntyre, C. C., Fernandez, H. H. & Vitek, J. L. Association of Deep Brain Stimulation Washout Effects With Parkinson Disease Duration. *JAMA Neurol.* **70**, 95–99 (2013).

41. GitHub - google/mediapipe: Cross-platform, customizable ML solutions for live and streaming media. https://github.com/google/mediapipe.

42. opencv/opencv. (2023).

43. Mathis, A. *et al.* DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nat. Neurosci.* **21**, 1281–1289 (2018).

44. Nath, T. *et al.* Using DeepLabCut for 3D markerless pose estimation across species and behaviors. *Nat. Protoc.* **14**, 2152–2176 (2019).

45. Motulsky, H. J. & Brown, R. E. Detecting outliers when fitting data with nonlinear regression – a new method based on robust nonlinear regression and the false discovery rate. *BMC Bioinformatics* **7**, 123 (2006).

46. Lakens, D. Equivalence Tests: A Practical Primer for t Tests, Correlations, and Meta-Analyses. *Soc. Psychol. Personal. Sci.* **8**, 355–362 (2017).

47. Lakens, D., Scheel, A. M. & Isager, P. M. Equivalence Testing for Psychological Research: A Tutorial. *Adv. Methods Pract. Psychol. Sci.* **1**, 259–269 (2018).

48. The jamovi project (2021). jamovi (Version 1.6) [Computer Software]. Retrieved from https://www.jamovi.org.

49. RStudio Team. R Studio. (2020).

50. GraphPad Software. GraphPad Prism.

51. 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. Off. J. Mov. Disord. Soc.* **26**, 1548–1552 (2011).

52. Lauk, M. *et al.* Side-to-side correlation of muscle activity in physiological and pathological human tremors. *Clin. Neurophysiol. Off. J. Int. Fed. Clin. Neurophysiol.* **110**, 1774–1783 (1999).

53. Sandoe, C. *et al.* Predictors of deep brain stimulation outcome in tremor patients. *Brain Stimulat.* **11**, 592–599 (2018).

54. Favilla, C. G. *et al.* Worsening essential tremor following deep brain stimulation: disease progression versus tolerance. *Brain* **135**, 1455–1462 (2012).

55. Agarwal, S. & Biagioni, M. C. Essential Tremor. in *StatPearls* (StatPearls Publishing, 2023).

56. Pan, Z. *et al.* Clinical application of an automatic facial recognition system based on deep learning for diagnosis of Turner syndrome. *Endocrine* **72**, 865–873 (2021).

57. Biderman, D. *et al.* Lightning Pose: improved animal pose estimation via semi-supervised learning, Bayesian ensembling, and cloud-native open-source tools. *bioRxiv* 2023.04.28.538703 (2023) doi:10.1101/2023.04.28.538703.

58. Wang, X., Garg, S., Tran, S. N., Bai, Q. & Alty, J. Hand tremor detection in videos with cluttered background using neural network based approaches. *Health Inf. Sci. Syst.* **9**, 30 (2021).

59. Williams, S. *et al.* Computer vision of smartphone video has potential to detect functional tremor. *J. Neurol. Sci.* **401**, 27–28 (2019).

60. Nunes, A. S. *et al.* Automatic Classification and Severity Estimation of Ataxia From Finger Tapping Videos. *Front. Neurol.* **12**, (2022).

61. Vidailhet, M., Roze, E. & Jinnah, H. A. A simple way to distinguish essential tremor from tremulous Parkinson's disease. *Brain* **140**, 1820–1822 (2017).

62. Chiu, S. Y. *et al.* Ataxia and tolerance after thalamic deep brain stimulation for essential tremor. *Parkinsonism Relat. Disord.* **80**, 47–53 (2020).

63. Kadirvelu, B. *et al.* A wearable motion capture suit and machine learning predict disease progression in Friedreich's ataxia. *Nat. Med.* **29**, 86–94 (2023).

64. Ilg, W. *et al.* Digital Gait Biomarkers Allow to Capture 1-Year Longitudinal Change in Spinocerebellar Ataxia Type 3. *Mov. Disord.* **37**, 2295–2301 (2022).

65. Needham, L. *et al.* The accuracy of several pose estimation methods for 3D joint centre localisation. *Sci. Rep.* **11**, 20673 (2021).

66. Chatzis, T., Stergioulas, A., Konstantinidis, D., Dimitropoulos, K. & Daras, P. A Comprehensive Study on Deep Learning-Based 3D Hand Pose Estimation Methods. *Appl. Sci.* **10**, 6850 (2020).

67. Cleeves, L. & Findley, L. J. Variability in amplitude of untreated essential tremor. *J. Neurol. Neurosurg. Psychiatry* **50**, 704–708 (1987).

68. Neumann, W.-J., Gilron, R., Little, S. & Tinkhauser, G. Adaptive Deep Brain Stimulation: From Experimental Evidence Toward Practical Implementation. *Mov. Disord.* **n/a**, (2023).

69. Fox, M. D. & Deuschl, G. Converging on a Neuromodulation Target for Tremor. *Ann. Neurol.* **91**, 581–584 (2022).

70. Malekmohammadi, M. *et al.* Kinematic Adaptive Deep Brain Stimulation for Resting Tremor in Parkinson's Disease. *Mov. Disord.* **31**, 426–428 (2016).

71. Fasano, A. *et al.* Gait ataxia in essential tremor is differentially modulated by thalamic stimulation. *Brain* **133**, 3635–3648 (2010).

72. Groppa, S. *et al.* Physiological and anatomical decomposition of subthalamic neurostimulation effects in essential tremor. *Brain* **137**, 109–121 (2014).

73. Deuschl, G., Bain, P., Brin, M. & Committee, A. H. S. Consensus Statement of the Movement Disorder Society on Tremor. *Mov. Disord.* **13**, 2–23 (1998).

74. Fahn, S., Tolosa, E. & Marin, C. Clinical Rating Scale for Tremor. **2**, 271–280 (1988).

75. Haglin, J. M., Jimenez, G. & Eltorai, A. E. M. Artificial neural networks in medicine. *Health Technol.* **9**, 1–6 (2019).

76. Anvari, F. & Lakens, D. Using anchor-based methods to determine the smallest effect size of interest. *J. Exp. Soc. Psychol.* **96**, 104159 (2021).

77. Quantifying Tremor in Essential Tremor Using Inertial Sensors—Validation of an Algorithm. *IEEE J. Transl. Eng. Health Med.* **9**, 2700110 (2020).

Figures

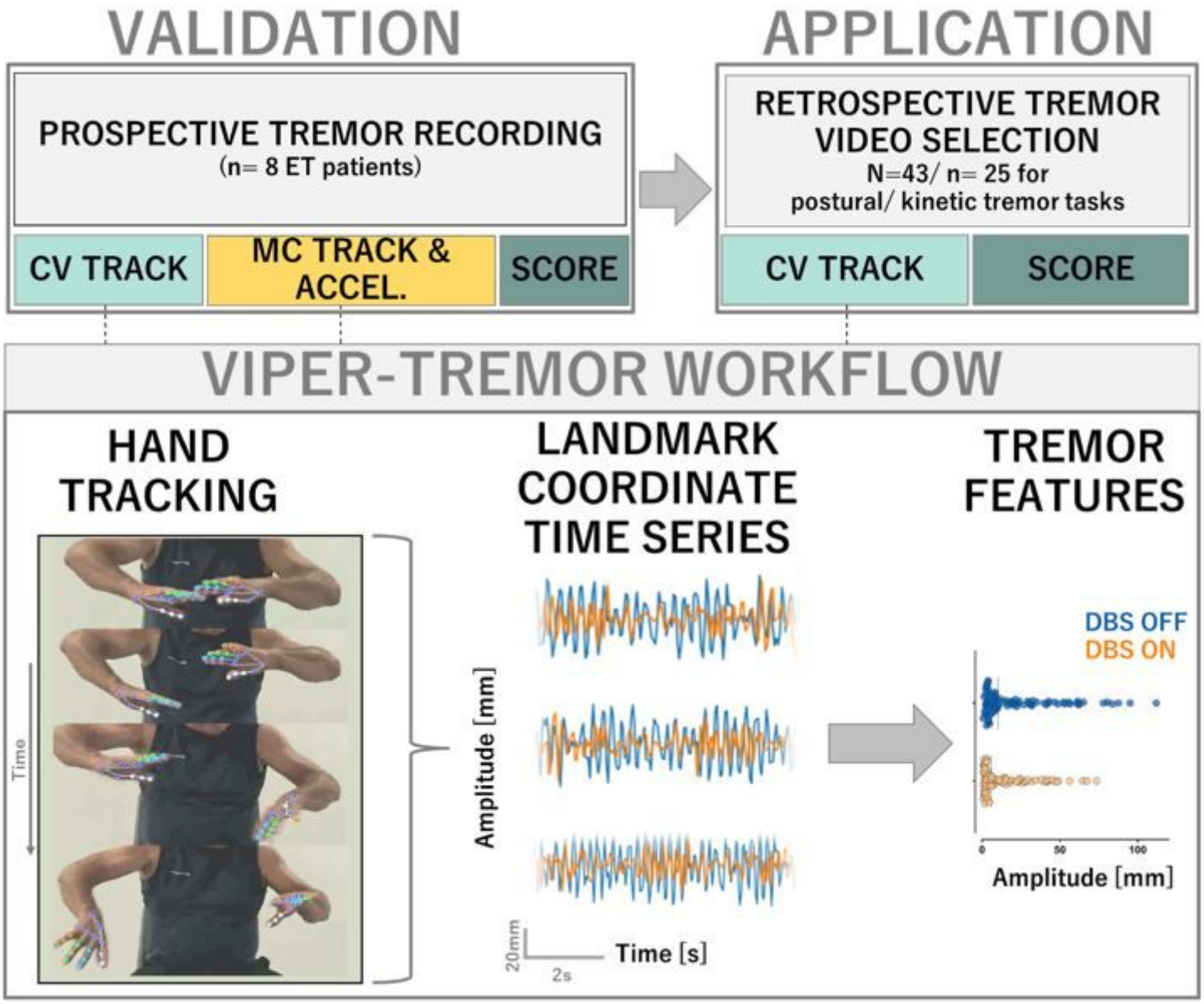

Figure 1

Workflow. The study was conducted into two independent cohorts/phases. In the prospective validation phase, upper limb tremor was recorded with a state-of-the-art motion capture (MC) and wrist-worn accelerometer (ACCEL) setup, as well as with a smartphone videocamera for subsequent computer vision-based analysis (CV). Additionally, tremor was scored using the Fahn-Tolosa-Marin clinical tremor rating scale. Time series data extracted from motion capture, accelerometer, and computer vision analysis was passed into a common analytical pipeline to compute canonical tremor features38 for clinical and inter-methodological correlation. Finally, the computer vision-based tremor analysis was applied to a retrospective video dataset of ET patients before and under chronic DBS treatment for additional clinical validation in an independent dataset. Given the largely unstandardized video recordings, video selection criteria were applied in order to ensure sufficient data quality, in accordance with best practice considerations in computer vision movement analysis26,32. Subjects were included for

tremor analysis if: (i) at least one preoperative as well as one postoperative follow-up video under active DBS was available, (ii) video material was free of excessive camera movements, (iii) video material did not contain visual distortion such as different zoom depths, (iv) tracked limbs were fully visible in the video frame throughout the recording, (v) video material contained at least 5 seconds of continuous arm holding for postural and 3 repetitions of finger-to-nose pointing per side for kinetic tremor assessment.

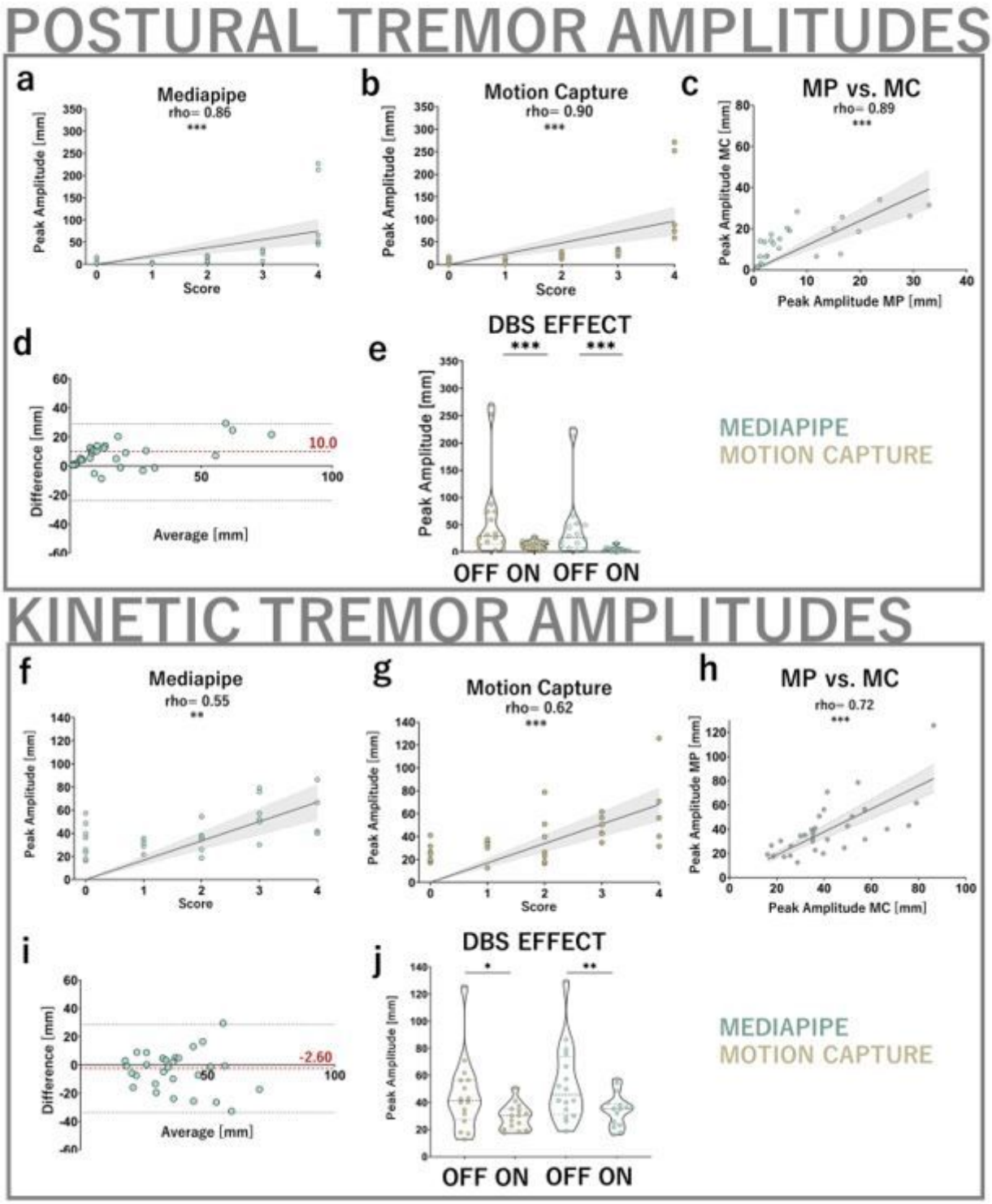

Figure 2

Tremor amplitude analysis. Postural: a-c Both the computer vision and motion capture outcomes show strong and significant agreement with clinical postural tremor scores (MP: rho= 0.86, MC: rho= 0.90, both p< .001) as well as excellent inter-methodological agreement (rho= 0.89, p<.001). **d** Mean deviation of computer vision-derived amplitude measurements from motion capture is 10mm and Bland-Altman plotting shows no systematic relationship of measurement and error magnitudes. Of note, computer vision and motion capture recordings were not identical (see methods) but reflect sequential recordings. **e** Computer vision and motion capture-derived amplitudes are sensitive to DBS effect, similar to clinical scores (effect size 0.94 for motion capture, 1.00 for Mediapipe and clinical scores, all p<.001, Durbin-Conover corrected). **Kinetic: f-h** Mediapipe and motion capture-derived amplitudes show strong and significant correlations with clinical scores (MP: rho= 0.55, p= .01, MC: rho= 0.62, p< .001,) while being strongly intercorrelated (MP vs. MC: rho= 0.72, p< .001). **i**Mediapipe exhibits an accuracy for kinetic tremor tracking of -2.6 mm, as measured by the mean absolute deviation from motion capture. Bland-Altman plotting reveals no systematic relationships of error and measurement magnitudes. **j** Mediapipe and motion capture-derived amplitudes are responsive to DBS effect, with similar effect sizes (MP: 0.69, p= .003, MC: 0.60, p= .022).

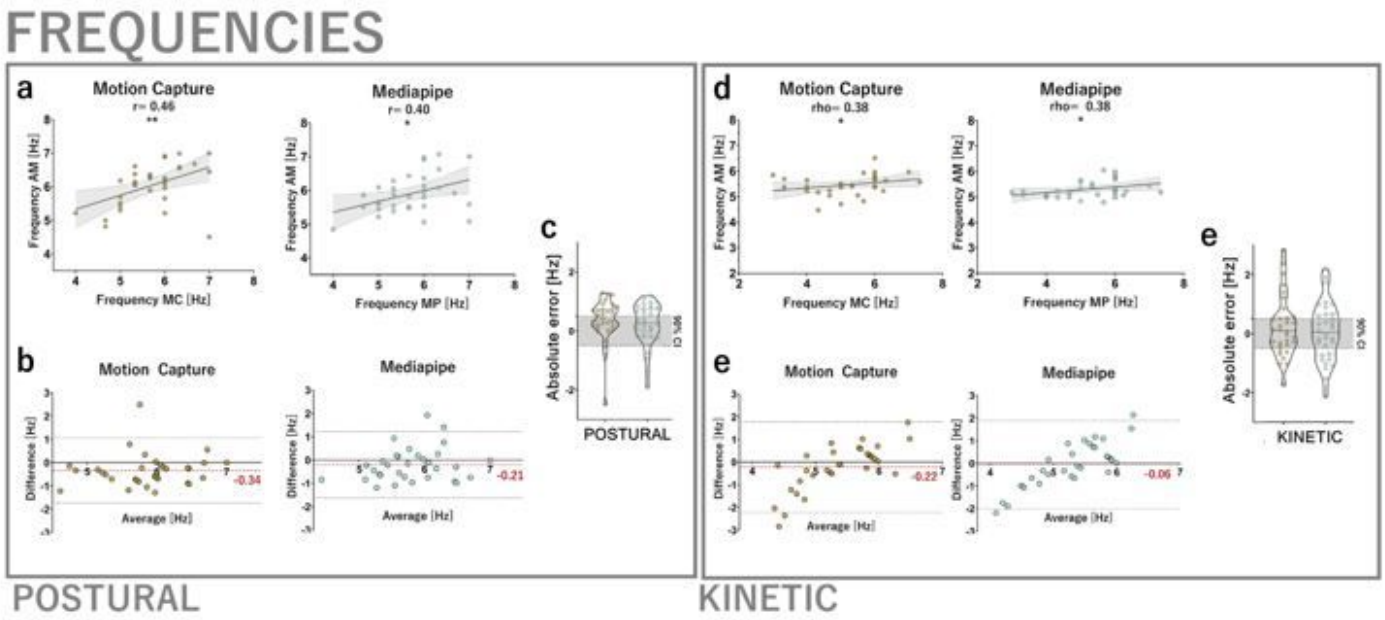

Figure 3

Tremor frequency analysis. a In the postural tremor condition, Mediapipe (MP) demonstrates strong correlations with accelerometer measurements (AM), similar to motion capture (MC), when estimating tremor frequencies (MP: r= 0.40, p< .05, MC: r= 0.46, p< .01). **b** Accuracies of postural tremor frequency measurements were calculated to be -0.34 Hz for motion capture and -0.21 Hz for Mediapipe. Bland-Altman plots revealed no systematic relationship of error and measurement magnitudes. **c** Absolute errors of postural tremor frequency measurements are equivalent to motion capture. **d** In the kinetic tremor condition, Mediapipe and motion capture-derived frequencies showed similar correlation strengths

to accelerometric frequency measurements (rho= 0.38, p< .05). **e** Accuracies of kinetic tremor frequency measurements were calculated to be -0.22 Hz for motion capture and -0.06 Hz for Mediapipe, however both with systematic relationships of error and measurement magnitudes. **f** Absolute errors of kinetic tremor frequency measurements are equivalent to motion capture.

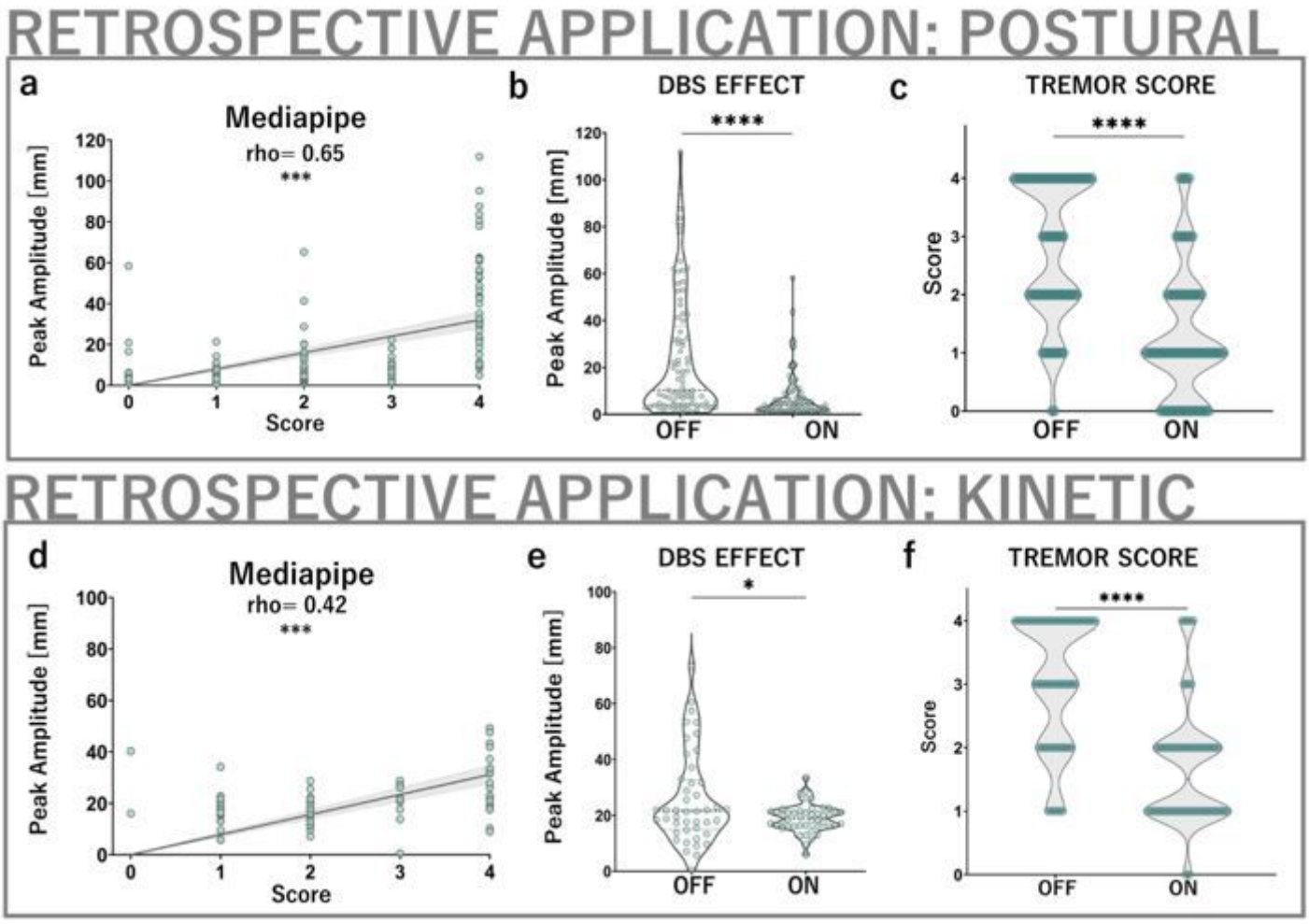

Figure 4

Application of computer vision tremor analysis in an independent, retrospective cohort. a-b Computer vision-derived postural tremor amplitude measurements are strongly correlated to clinical scores (MP: rho= 0.65, p< .001) as well as responsive to DBS effects (MP: r= 0.49 (95% CI [0.34, 0.61]), p< .001; score: r= 0.61 (95% CI [0.49, 0.71], p< .001, **b-c**). **d** For kinetic tremor, Mediapipe-derived amplitude measurements are substantially correlated to respective clinical scores (rho = 0.42, p< .001) and responsive to the effect of DBS (MP: r= 0.37, 95% CI [0.03, 0.61], p= 0.025;score: r = 0.64, 95% CI [0.48, 0.75], p< .001, **e-f**).

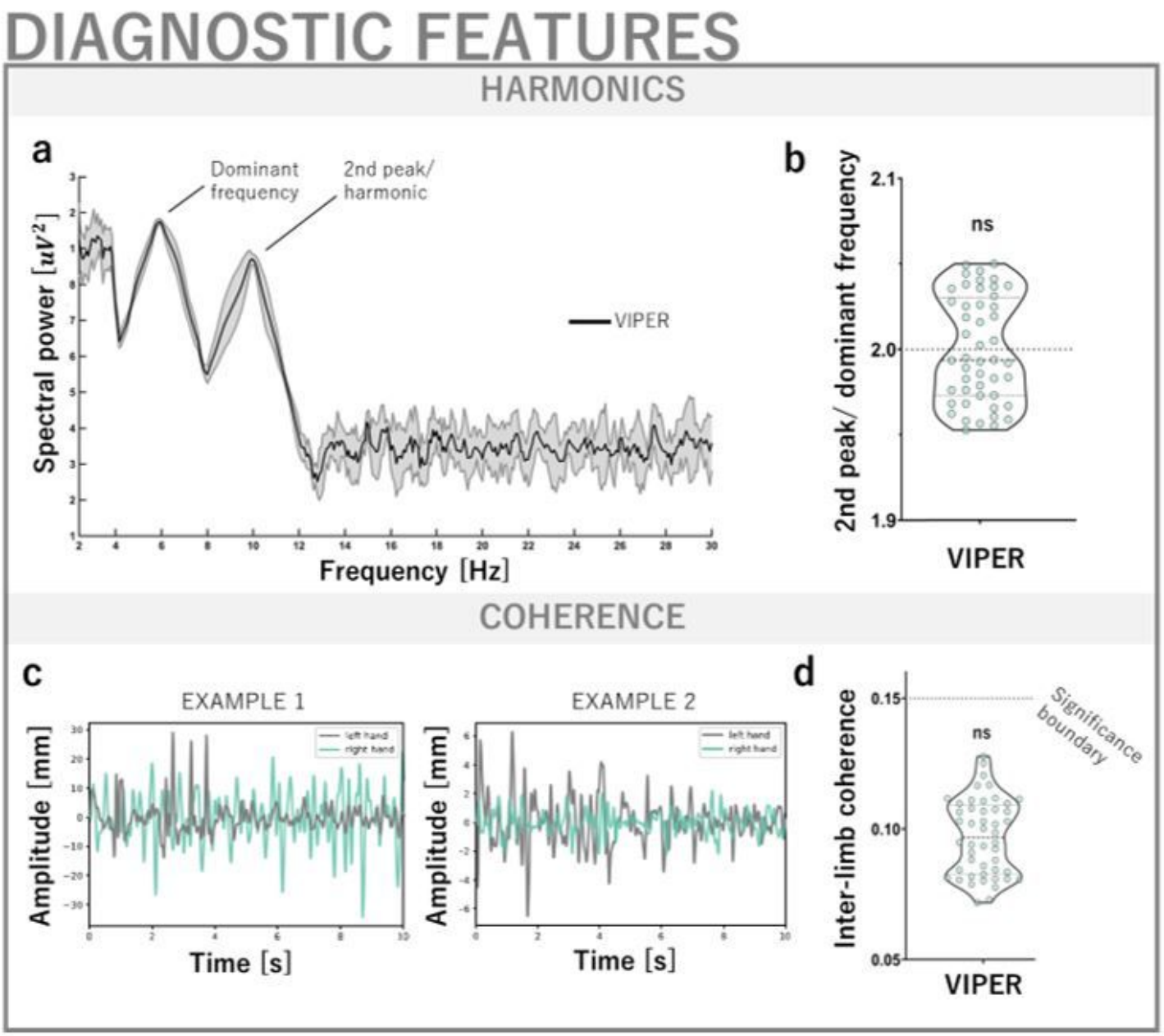

Figure 5

Using VIPER-Tremor to augment diagnostic insight. a Example power spectrum derived with VIPER-Tremor. In addition to the primary, dominant frequency peak, there is a clear cut second peak, i .e. a harmonic. **b** In line with previous electromyographic work[51] demonstrating the diagnostic value of the presence of a harmonic at twice the dominant tremor frequency, we divided the frequencies of the second and first frequency peak and performed a one-sample t-test against 2.0, which was not significant (p= .99, SD of discrepancy 0.03). **c** Mediapipe-derived tremor signals from both hands are not coherent. **d** On a group level, interlimb coherence values fall below the time series length-dependent significance value of 0.15 (median MP: median 0.097, maximum 0.13), in line with electromyographic literature[51].

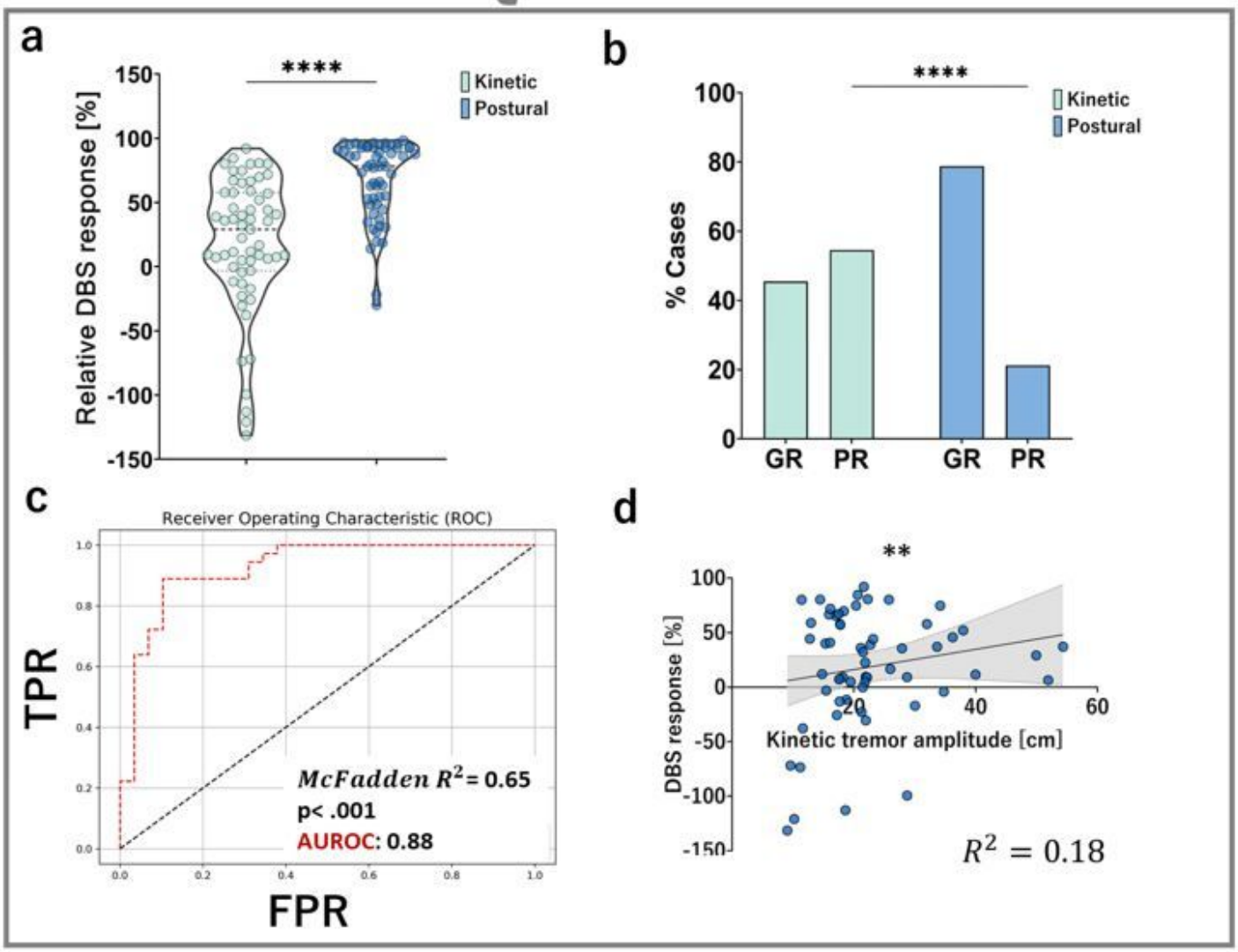

Figure 6

Using VIPER-Tremor for DBS effect quantification and prognostication. **a** DBS exerts stronger effects on postural than kinetic tremor as measured by relative amplitude reduction (p< .001, effect size= 0.72). **b** Binarizing the sample into good (GR) and poor responders (PR) using a combined criterion of ≥2 cm tremor amplitude and ≤30% amplitude decrease in DBS ON yields a significant greater fraction of poor responders in the context of kinetic tremor, which is an important driver of functional disability (55% vs. 21% of cases, p< .001). **c** Preoperative hand kinematic features are strong and significant predictors of DBS outcome ($\chi 2$= 58.4, p< .001, McFadden R^2= 0.65). Baseline kinetic tremor amplitude is found to be a significant and independent predictor of DBS outcome for kinetic tremor. Leave-one-out-cross-validation evaluation yields an AUROC of 0.88, a balanced accuracy of 0.88, F1-score 0.89. **d** Moreover, baseline kinetic tremor amplitude is a significant predictor of the relative DBS response in a linear model (p= .012), as measured by the percentual reduction of kinetic tremor amplitude (R^2= 0.18, Pearson's r= 0.35, p= .005).

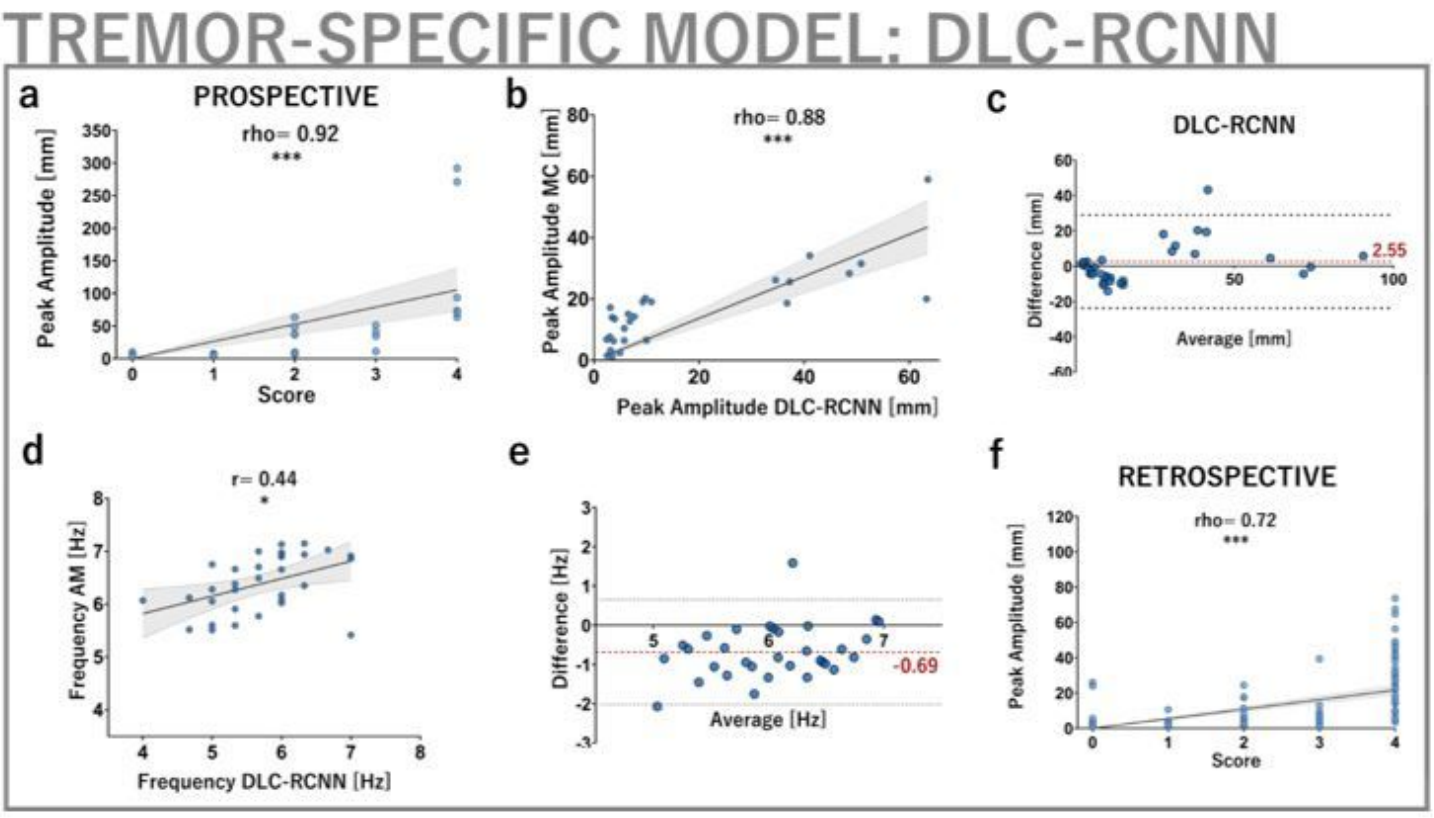

Figure 7

Application of a disease-specific convolutional neural network for postural tremor analysis across cohorts. a-c In the prospective cohort, DLC-RCNN-derived amplitude measurements are strongly correlated to clinical scores (rho= 0.72, p< .001) and motion capture (rho= 0.88, p< .001). Mean absolute error is 2.55 mm (95% CI [-2.11, 7.29]) with no systematic relationship to measurement magnitudes. **d-e** DLC-RCNN frequency measurements are moderately correlated to accelerometer (r= 0.44, p< .05) with a mean absolute error of -0.69 Hz [95% CI -0.93, 0.44]. **f**In the retrospective cohort, DLC-RCNN-derived postural tremor amplitudes show a moderate correlation to clinical scores (rho= 0.72, p< .001). DLC-RCNN however failed to capture and measure kinetic tremor in both cohorts.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryFigure1.svg
- SupplementaryFigure2.svg
- SupplementaryFigure3.svg
- SupplementaryFigure4.svg
- SupplementaryFigure5.svg
- SupplementaryMaterial.docx