medRxiv preprint doi: https://doi.org/10.1101/2024.03.13.24304101; this version posted March 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in

perpetuity. It is made available under a CC-BY 4.0 International license.

Article

Development of a Tremor Detection Algorithm for use in an Academic Movement Disorders Center

Mark Saad,¹ Sofia Hefner, Suzann Donovan, Doug Bernhard,¹ Richa Tripathi,¹ Stewart A. Factor,¹ Jeanne Powell,² Hyeokhyen Kwon,³ Reza Sameni,³ Christine D. Esper,¹ and J. Lucas McKay^{1,3*}

- Jean and Paul Amos Parkinson's disease and Movement Disorders Program, Department of Neurology, School of Medicine, Emory University, Atlanta, GA, USA
- Department of Psychology, Laney Graduate School, Emory University, Atlanta, GA, USA
- Department of Biomedical Informatics, School of Medicine, Emory University, Atlanta, GA, USA

Correspondence: lucas@dbmi.emory.edu

Abstract: Tremor, defined as an "involuntary, rhythmic, oscillatory movement of a body part," is a key feature of many neurological conditions, but is still clinically assessed by visual observation. Methodologies for objectively quantifying tremor are promising but remain non-standardized across centers. Our center performs full-body behavioral testing with 3D motion capture for clinical and 4 research purposes for patients with Parkinson's disease, essential tremor, and other conditions. The objective of this study was to assess the ability of several candidate processing pipelines to identify 6 the presence or absence of tremor in kinematic data from movement disorders patients compared to expert ratings from movement disorders specialists. We curated a database of 2,272 separate kinematic data recordings from our center, each of which was contemporaneously annotated as q tremor present or absent by a clinical provider. We compared the ability of six separate processing 10 pipelines to recreate clinician ratings based on F1 score, in addition to accuracy, precision, and recall. 11 We found generally comparable performance across algorithms. The average F1 score was 0.84 ± 0.02 12 (Mean \pm SD; range 0.81 – 0.87), with all F1 confidence intervals overlapping. The highest performing 13 algorithm (cross-validated F1 = 0.87) was a hybrid that used engineered features adapted from 14 an algorithm in longstanding clinical use with a modern Support Vector Machine classifier. Taken 15 together, our results suggest the potential to update legacy clinical decision support systems to 16 incorporate modern machine learning classifiers in order to create better performing tools. 17

Keywords: Motion Capture; Parkinson's Disease; Essential Tremor; Machine Learning, Support Vector Machines, XGBoost

1. Introduction

Tremor, defined as an "involuntary, rhythmic, oscillatory movement of a body part," is a key feature of many neurological conditions, and has been called the most frequent human movement disorder [1]. In Parkinson's disease (PD), the second most common neurodegenerative disorder worldwide [2], tremor that appears while at rest (often a "pillrolling" tremor of the thumb and forefinger) is considered a characteristic sign [3]. However, tremor is a feature of many neurological conditions [3], and can also result from various causes such as trauma or side effects of medications [4]. Furthermore, some oscillatory movements occur that are not tremor; myoclonus and dystonia can produce involuntary, jerking movements that may be rhythmic, but that are not considered to be tremor [5].

Tremor disorders currently require expert diagnosis based on skilled observation using 30 standardized clinical scales, with quantitative measurements approximated by eye and 31 with no automated clinical decision support. Clinicians characterize the features of the 32 tremor, including body distribution, provocative factors, frequency, and gross amplitude 33 Attri- bution (CC BY) license (https:// NOTE: This preprint reports and regaregente that is is in the most in the time time to be a starting t 34 causes and to evaluate potential treatment plans[2]. In PD and Essential Tremor (ET), 35

Citation: Lastname, F.: Lastname, F.: Lastname, F. Tremor Algorithms. Sensors 2023, 1, 0. https://doi.org/

Received Revised: Accepted Published:

Copyright: © 2024 by the authors. Submitted to Sensors for possible open access publication under the terms and conditions of the Creative Commons 4.0/).

18

19

20

21

22

23

24

25

26

27

28

medRxiv preprint doi: https://doi.org/10.1101/2024.03.13.24304101; this version posted March 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

perpetuity. It is made available under a CC-BY 4.0 International license .

Version March 13, 2024 submitted to Sensors

2 of 22

70

71

72

73

74

75

76

77

78

79

80

81

82

83

84

85

86

87

overall tremor severity is measured using standardized clinical scales like the Movement36Disorder Society-Unified Parkinson's Disease Rating Scale Part III (MDS-UPDRS-III) [6],37the Fahn-Tolosa-Marin Tremor Rating Scale (FTM) [5] and The Essential Tremor Rating38Scale (TETRAS) [7]. These give general guidelines for tremor amplitude assessment by39eye, but are not intended to be used with actual measurements (e.g., with calipers or an40anthropometer).41

Recent progress in human activity recognition and edge computing suggests signifi-42 cant potential for automated clinical decision support tools in tremor measurement. Despite 43 this, technologies for identifying tremor have progressed towards standardization and 44 clinical uptake very slowly [1]. In the research domain, various technologies measure 45 human motion, including body-worn sensors [3,7,8], 3D motion capture [9,10], and – most 46 recently - pose recognition from monocular video [11–13]. Digitizing tablets are often 47 used for assessing tremor during tasks like spiral drawing [14,15] and for discriminating 48 tremor from bradykinesia during finger tapping [16]. In fact, recognition of the potential 49 for spectral analysis in assessing tremor dates back to the mid-1960s [17], and differences 50 in tremor frequencies across disorders have been acknowledged for over two decades [3]. 51 Substantial domain knowledge (and in some cases, cultural) gaps between clinicians and 52 engineers further hampers widespread adoption. This is in contrast to fields like cardiology, 53 where automated clinical decision support systems thrive due to large public datasets 54 enabling annual improvements in anomaly detection [18,19]. 55

In our center, we perform comprehensive behavioral testing using 3D kinematic mo-56 tion capture to objectively evaluate abnormal movements in patients with PD, ET, and 57 other conditions [20]. Indications for this procedure include diagnosis adjudication and 58 evaluation for functional neurosurgery, among others. Our behavioral testing paradigm 59 involves multiple standardized upper limb tasks designed to elicit tremor under provoking 60 conditions of rest, posture, and action. Since 2014, we have performed >1500 behavioral 61 tests, using analysis pipelines that were developed organically based on clinician domain 62 knowledge without formal evaluation. A challenge encountered in evaluating tremor anal-63 ysis algorithms is imprecision in the "ground truth" criteria for tremor presence outlined in 64 clinical scales [5–7]. For instance, the MDS-UPDRS-III criterion that "tremor is present but 65 less than 1 cm in amplitude" (corresponding to a score of 1) is clear for a human rater but 66 poses ambiguity for a machine. Questions arise in implementation, such as along which 67 biomechanical axis or axes the amplitude should be measured and what size tremor meets 68 the threshold for being considered "present." 69

The objective of this study is to compare tremor identification algorithms from our clinic for identifying tremors in 3D kinematic data of movement disorder patients. Ground truth labels, recorded in contemporaneous notes by clinicians, are straightforward: tremor present or absent. The main goal is to identify the most accurate algorithm for detecting tremor presence or absence in individual body parts during testing sessions.

2. Materials and Methods

2.1. Data sources

We compared algorithm performance using a database of 2,272 recordings made during standard clinical exams of N = 50 arbitrarily selected clinic patients. Aspects of the testing paradigm have been described previously [9,20,21]; more detail is provided below. In 42 patients (84%) the primary diagnosis was either Parkinson's disease or essential tremor. Detailed demographic and clinical characteristics are shown in Table 1.

2.2. Behavioral testing paradigm

Behavioral testing was captured through 3D optical motion capture with 60 reflective markers on standardized bony landmarks during a 1-hour clinical assessment in our facility (Figure 1). Assessments were billed under Current Procedural Terminology (CPT [22]) codes 96000 and 96004. All patients with Parkinson's disease were asked to hold their antiparkinsonian medications for at least 12 hours prior to the study visit (the practically-

Version March 13, 2024 submitted to Sensors

Value
50
32 (64%)
18 (36%)
65 (12)
36 - 83
29 (58%)
13 (26%)
6 (12%)
1 (2%)
1 (2%)





Figure 1. Clinical motion capture facility. Our center uses a custom set of 60 retroreflective kinematic markers for most cases. Markers on the hands (arrows) enable tremor measurement. From top to bottom, the markers highlighted are R.Wrist, R.Thumb.M3, and R.Finger3.M3 (A). After data collection, analysis is performed using a deidentified wire frame or representation of the individual, preserving privacy (B). Our 650 square feet center is used for both clinical and research applications (C). The origin of the kinematic coordinate system is superimposed.

defined OFF state [23]). At the time of testing, the average time since the last medication 88 dose was 13±5 hours. Tasks included goal-directed upper limb movements, static postures, 89 and walking, and were designed to provoke various tremors [2]. For instance, seated 90 finger-to-nose pointing with the right arm (coded *sit-point-right* or *sit-point-1* in data files) 91 aimed to elicit action-provoked tremor in the right upper extremity and rest or postural 92 tremor in the legs, left upper extremity, torso, head, and neck (Table B.2). On average, 93 kinematic data recordings were 27 ± 9 seconds long and ranged from 3 to 92 seconds. 94

2.3. Kinematic data recording, processing, and export

Data were captured using a 3D motion capture system (Motion Analysis Corporation, 96 Rohnert Park, CA, USA) with 10 cameras recording at 120 Hz. Following testing completion, 97 clinic staff manually postprocessed kinematic data using standard interpolation features 98 in Motion Analysis Cortex software for quality control. Occasional low-pass or similar 99 filters were applied on an as-needed basis to address noise in individual markers, but no 100 consistent additional filtering occurred. Each recording's kinematic data was exported 101 into standard *.trc tabular format. A typical .trc file for a 30-second recording at 120 Hz 102 comprises 3600 rows (30 seconds \times 120 Hz) and 180 columns (60 markers \times 3 axes) of 103 kinematic data. Due to changes in marker labels and occasional missing data, each .trc file 104 was divided into separate .csv files for each body extremity in the accompanying dataset. 105 These files are compatible with standard Python, R, Matlab, or similar software libraries. 106 Summaries of the contents of example files are provided in Table B.1. 107

Version March 13, 2024 submitted to Sensors

Algorithm **Kinematic Data** Classifier Features Aggregation A1r **Rule-Based** Velocity Engineered Winner-take-all A1s SVM A2r Position Engineered Average Rule-Based A2s SVM B1 SVM Position Data-driven Average B2 Position Data-driven XGBoost Average

Table 2. Comparison of tremor identification algorithms. All algorithms operate on spectral features of kinematic data.

2.4. Annotations

Annotations were taken contemporaneously during the exam for the clinicians' own 109 use while preparing exam notes. Because tremor is intermittent in nature and typically does 110 not appear across more than a few isolated body regions, annotations typically included 111 separate entries for particular body parts during each particular recording. For example, 112 the annotation "Left hand: present, F3 and thumb" was used to indicate that tremor was 113 present on the third finger (F3) and thumb of a particular trial. Therefore, the annotations 114 were converted by the study team into separate annotations for each body extremity during 115 each recording. For example, "mild bilateral rest hand tremor" was converted into the 116 annotation "tremor present" for each of the left and right hands. As the presence or absence 117 of tremor in other body extremities was ambiguous in this case, no annotations were 118 provided for other body extremities. In cases where the absence of tremor was described in 119 the original notes ("this gentleman does not have tremor") tremor was labeled as "tremor 120 absent" for all extremities. Cases in which dyskinesia or dystonic posturing was present 121 were labeled as "tremor absent." 122

2.5. Spectral composition of kinematic data

All algorithms used initial preprocessing to isolate spectral (or "frequency-domain") 124 features of recorded data based on the substantial amount of established research in 125 this area. Tremor frequencies vary from 0.5 Hz for palatal tremor to 18 Hz for primary 126 orthostatic tremor, [2] with the majority of parkinsonian and essential tremor typically 127 occurring between 4-12 Hz. [3] Volitional movements during behavioral testing typically 128 occur primarily at lower frequencies, while higher frequency ranges are prone to artifacts 129 related to aliasing or electrical or other noise. For this reason, tremor data are typically 130 processed by band-pass filtering. Typical ranges include 1 to 16, 0.5 to 15, or 2 to 30 Hz [1]. 131 All of the tremor detection algorithms examined employed some initial band-pass or other 132 filtering, described below. 133

2.6. Algorithms

Identifying tremor is a process that uses the rich information embedded within motion 135 data from kinematic markers on each extremity during to determine whether tremor is 136 present or absent in a particular testing session. Although this use case is unique, like many 137 general machine learning problems, this process can be broken down into two basic steps. 138 The first step is *feature engineering*: extracting information ("features") from raw kinematic 139 marker data. The second step is *classifier development*: creating a classifier based on the 140 extracted features that determines whether tremor is present or absent. During classifier 141 development, in particular, it is important to perform some hyperparameter optimization 142 to identify the optimal operating point for a given algorithm. 143

In this study, we compare six algorithms for identifying tremor (Table 2). The first two (A1r and A2r) were developed organically over several years based on clinical expertise and signal processing heuristics. As implemented in our clinic, both algorithms derive engineered spectral features from the kinematic data which are then input into simple 147

108

123

medRxiv preprint doi: https://doi.org/10.1101/2024.03.13.24304101; this version posted March 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in

perpetuity. It is made available under a CC-BY 4.0 International license

Version March 13, 2024 submitted to Sensors

5 of 22

rule-based classifiers to determine whether tremor is present or absent. While these algo-148 rithms were developed iteratively over several years with access to the clinical dataset, no 149 comprehensive hyperparameter tuning was performed, potentially leading to suboptimal 150 parameter settings. 151

To create a more fair comparison with the modern machine learning algorithms (B1 and 152 B2), we also examined the performance of these algorithms when the features identified by 153 each (summarized in Table 3) were used as inputs to a well-established machine learning 154 model (Support Vector Machines, SVMs [24]) trained and evaluated with 5-fold cross-155 validation. To distinguish these algorithms from the related algorithms with rule-based 156 classifiers, these implementations are referred to as A1s and A2s. 157

The final two algorithms (B1 and B2) were developed specifically for this study 158 based on standard modern machine learning best practices. Both B1 and B2 use basic 159 preprocessing and spectral features together with well-established machine learning models 160 to identify optimal operating points. The details of each algorithm are described below. 161

2.6.1. Velocity Spectral Peak Detection (Algorithm A1r)

The oldest algorithm in use in our center was developed iteratively between 2014 and 163 2020. The key feature of this algorithm is that it performs numerical differentiation on 164 kinematic data prior to feature identification in the frequency domain. It uses a winner-take-165 all approach to aggregate tremor features across kinematic markers on a given extremity 166 (described below). An example of tremor identification using Algorithm A1r is presented 167 in Figure 2. This algorithm was implemented in Matlab (Version R2022b; The Mathworks, 168 Natick, MA, USA).

Feature extraction.

Raw kinematic displacement data for each marker of a given extremity are zero-phase 171 lowpass filtered (20 Hz), centered, and passed through a Savitzky-Golay derivative filter to 172 obtain smooth velocity estimates in each of the x, y, and z dimensions. The power spectral 173 densities (PSDs) of the velocity components for each marker are obtained using Welch's 174 method and combined using the Euclidean norm. The combined PSD of each marker is then converted to log scale, smoothed using a Savitzky-Golay filter, and converted back 176 to a linear scale for spectral analysis. Spectral features are summarized in Table 3. More 177 details on feature calculation are available in the documentation for powerbw.m. 178

Rule-based classification.

To detect a peak that would indicate the presence of tremor, the peak power and 180 the corresponding center frequency were first detected for each kinematic marker using 181 functionality integrated in the Matlab function powerbw.m. A significant peak should be 182 narrow and symmetric about the center frequency, so any peak with a bandwidth greater 183 than 2 Hz or nonsymmetric power to the left and right of the peak would cast doubt on 184 the presence of a physiological tremor. Indicators of bandwidth and symmetry are derived 185 using powerbw.m and subjected to threshold rules to determine the presence or absence of 186 tremor. Peaks with center frequencies above 10 Hz would also be deemed unreasonable; 187 therefore, central frequencies above 10 Hz are also interpreted as tremor absence. To 188 aggregate features across markers of a given extremity, the algorithm proceeds to detect a 189 tremor on each marker independently. The tremor features for the marker with the largest 190 tremor amplitude on a given recording are used as representative of the entire extremity. 191

2.6.2. Amplitude Spectral Peak Detection (Algorithm A2r)

The amplitude spectral peak detection was established in our center primarily to 193 provide tremor identification in the amplitude, rather than velocity domain, in order to 194 enable direct comparison with clinical magnitude cutoffs. The key feature of this algorithm 195 is that it converts all kinematic data from kinematic markers on a given extremity to the 196 frequency domain prior to aggregation with a max procedure. Therefore, the spectral 197

169

170

162

175

179

Version March 13, 2024 submitted to Sensors

Algorithm Feature Description A1r F CENTER Tremor frequency (Hz) AMPLITUDE_MM_P_S Tremor amplitude (mm/s) BW 3 dB bandwidth (Hz) HI_F Left frequency border (Hz) LO_F Right frequency border (Hz) MAX_POWER Maximum power level of the power spectrum (dB/Hz) HI POWER Power level at right frequency border (dB/Hz)LO POWER Power level at left frequency border (dB/Hz) RELATIVE POWER Proportion of total power A2r F CENTER Tremor frequency (Hz) AMPLITUDE_MM Tremor amplitude (mm) PROMINENCE Peak prominence (mm) WIDTH Peak width (Hz)

Table 3. Spectral features calculated by clinical algorithms A12r/s and A2r/s.

features identified for a given extremity reflect a combination of kinematic markers, rather 198 those of a single dominant markers. An example of tremor identification using Algorithm 199 A2r is presented in Figure 3. This algorithm was implemented in Matlab (Version R2022b; 200 The Mathworks, Natick, MA, USA). 201

Feature extraction.

Raw kinematic data of all markers on a given extremity are high-pass filtered with a 203 4th order Butterworth filter with corner frequency 2 Hz using filtfilt.m in Matlab. The two-204 sided frequency spectrum is calculated using the fast Fourier transform and converted into 205 the single-sided frequency spectrum of each axis of each kinematic marker. The single-sided 206 frequency spectra of each x, y, z component of all markers on each extremity are combined 207 using a max procedure to create an aggregate spectrum for the extremity that represents the 208 most severe tremor at each frequency. The aggregate spectrum is subsequently smoothed 209 with a Savitsky-Golay 3rd-order polynomial smoothing filter.'Frequency peaks in the 210 smoothed spectrum are then identified with the heuristic-based findpeaks.m method in Matlab software using default arguments. 212

Rule-based classification.

Classification proceeds in two steps. First, the central frequency of the dominant fre-214 quency peak identified by findpeaks.m is compared to maximum and minimum threshold 215 values. Peaks with central frequency < 3.5 Hz or > 10 Hz are considered non-physiological and are discarded. If these conditions are met, the amplitude of the peak is compared to 217 the a simple threshold value (0.1 mm) to determine tremor presence. This threshold value 218 was determined over trial and error. 219

2.6.3. Support Vector Machines with Engineered Spectral Features (Algorithms A1s and A2s)

We also examined the performance of Algorithms A1r and A2r when the final classifi-222 cation steps were altered from the heuristic rule-based implementations to Support Vector 223 Machines (SVMs). Support vector machines (SVM) are a widely recognized approach for 224 classification tasks [25]. An SVM is a supervised machine learning algorithm that works by 225 identifying an optimal hyperplane in an augmented feature domain that separates obser-226 vations into distinct classes. In this case, observations that fall on one or other side of the 227 hyperplane are classified as tremor present or absent. Importantly, the feature domain can 228

6 of 22

211

213

202

216

220

Version March 13, 2024 submitted to Sensors

7 of 22

be augmented with features derived via nonlinear functions (here, radial basis functions) 229 in order to accommodate linearly-non separable classes in the original data. Here, we 230 extracted the spectral features identified by each algorithm (summarized in Table 3) and 231 used them as inputs to two separate SVMs with 5-fold cross-validation and radial basis 232 function kernels. 233

2.6.4. Modern Classifiers (Algorithms B1 and B2)

The final two algorithms (B1 and B2) were developed specifically for this study. They 235 use basic preprocessing and spectral features together with well-established machine 236 learning models to identify tremor in kinematic data. 237

Feature extraction.

In order to decouple tremulous movements from slow body movements due to sub-239 jects' displacements and non-tremulous activities, the vector position of each kinematic 240 marker on a given extremity is initially calculated as a measure of its instantaneous distance 241 from the origin of the kinematic reference frame. This is done by calculating the Euclidean 242 norm of the x, y, and z coordinates at all time instants, resulting in a single signal per sensor, 243 as a function of time. The resulting signals are bandpass filtered between 1 Hz and 20 Hz 244 with a linear-phase finite impulse response (FIR) filter design using a hamming window 245 of order 80. The signals are next decimated from 120 Hz to 40 Hz to further focus on the 246 spectral range of interest. Next, the spectra of each sensor's signal is estimated by using 247 sliding windows of 3s and 2.75s overlap with a 120-point discrete Fourier transform (DFT). 248 The Welch power spectral density (PSD) estimation method with a Hamming window of 249 120 samples is used for PSD estimation, followed by a Gaussian-shaped moving average 250 with a standard deviation of 1 Hz, to further smooth the spectra, sharpening the dominant 251 frequencies and making them more distinguishable for the classifier. This results in 120 252 points of two-sided PSD with a spectral resolution of 0.33 Hz (40 Hz/120). The first 61 253 PSD values (corresponding to the DC component and one-sided spectrum) are used as 254 the spectral feature vector of each sensor. The average feature vector calculated across all 255 kinematic markers on a given extremity are then used as inputs to each of the classifiers 256 described below. 257

B1: SVM classification.

In algorithm B1, the 61-point one-sided average spectral features were directly pro-259 vided to an SVM as feature vectors. We considered SVM models with both linear and 260 radial basis function (RBF) kernels. A standard stratified 5-fold cross validation scheme 261 was performed, by splitting the data into 5 nonoverlapping splits, using 4 splits for training 262 and the left-out split for validation. The stratification ensured that each fold retained 263 approximately equal proportions of the two class labels. 264

B2: XGBoost classification.

In algorithm B2, the 61-point one-sided average spectral features were directly pro-266 vided to XGBoost as feature vectors. XGBoost is also a widely recognized approach for 267 classification tasks [26]. The XGBoost classifier operates by iteratively constructing an 268 ensemble of decision trees and refining them based on a specified loss function. The pro-269 cedure for loading the features was analogous to the SVM process, again using stratified 270 5-Fold cross-validation to ensure balanced representation across data splits. The classifier 271 was configured to bypass label encoding, opting instead for the 'logloss' evaluation metric. 272 This probability-centric metric enables the future extension of the proposed scheme for 273 estimating probabilities of tremulous events, instead of a binary decision. 274

238

234

258

medRxiv preprint doi: https://doi.org/10.1101/2024.03.13.24304101; this version posted March 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in

perpetuity. It is made available under a CC-BY 4.0 International license

8 of 22

2.7. Performance Metrics

We compared the ability of six separate processing to recreate clinician ratings of 276 tremor presence/absence based on F1 score, which is a harmonic mean of precision and 277 recall [9]: 278

$$Precision = \frac{TP}{TP + FP} \tag{1}$$

$$Recall = \frac{TP}{TP + FN}$$
(2)

$$F1 \ score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$
(3)

TP is a true positive that represents the total of successfully classified tremor-positive 279 records, FP is a false positive that represents the total number of misclassified tremor-280 negative records, total misclassified class windows, and FN is a false negative that repre-281 sents the total number of misclassified tremor-positive records. 282

We compared performance across classifiers using a confidence interval (CI) approach. 283 For each of classifiers A1s, A2s, B1, and B2, we calculated the average (μ) and standard 284 deviation (σ) of F1 scores observed across each of n = 5 folds during cross-validation. We then used the sample standard deviations to estimate 95% CIs as $\mu \pm 1.96 \times \frac{\sigma}{\sqrt{n}}$. We applied similar analyses to secondary outcomes including Accuracy, Precision, Recall, and AUC. 287 For classifiers A1r and A2r, we were only able to calculate point estimates of F1 and other 288 outcomes as these classifiers were developed manually with access to the entire dataset. 289

Finally, we characterized the contributions of different frequency bands to classification 290 with SHAP (SHapley Additive exPlanations) [27] plots. SHAP plots visually represent how 291 much each feature contributes to the classification of each observation as one class or the 292 other. This is analogous to the visual representation of factor loadings in familiar techniques 203 like principal components analysis (PCA) but is adapted for nonlinear techniques like 294 XGBoost. 295

3. Results

3.1. Characteristics of annotations

The most frequent clinical annotation was "present." Clinicians used a wide range 298 of qualitative labels for tremor size. A description of the mapping between raw clinician 299 provided labels and dichotomized dataset labels is provided in Table 4. Annotations 300 most frequently referred to the hands (37%), although annotations for all extremities were 301 present. The frequencies of appearance of various body extremities are described in Table 5. 302 Chi-squared tests identified significant differences in reporting frequencies between the 303 arms and legs (P \ll 0.001) and between the arms and head/torso (P \ll 0.001). 304

3.2. Model performance

Classification performance metrics for all models are reported in Table 6. The overall 306 highest performance, as assessed with average F1 score across five cross-validation folds, 307 was observed in algorithm A1s. However, all classifiers performed well. Across models, the 308 average F1 score was 0.84 ± 0.02 (range [0.81 - 0.87]), and all F1 confidence intervals were 309 overlapping. Figure 4 compares the performance of Algorithms B1 and B2 with operating 310 point information superimposed for the other algorithms. 311

We further performed feature importance using SHAP values, for all spectral features 312 used as inputs to algorithms B1 and B2. The SHAP plots are shown in Figure 5. Unlike 313 typical SHAP plots that are oriented vertically, in Figure 5 the feature importance is shown 314 on the y axis and the features - characteristic frequencies within the kinematic data - are 315 shown on the x axis, which allows visualization of the SHAP plots as a kind of spectrum. 316 We noted that high activation of features in the frequency range between 4.3 Hz and 7.0 Hz 317 were identified as significant for identifying tremor presence vs. absence (red), while high 318

275

285 286



297

Version March 13, 2024 submitted to Sensors

9 of 22

321

Table 4. Mapping between clinician-provided labels and training labels in training data. The "Other" label aggregates annotations with fewer than 10 observations and annotations for which no indicator of size was provided (e.g., "RH tremor").

	Label in	Dataset
Raw Label	Absent	Present
Absent	1476	0
Dystonia, dyskinesia, or other abnormal posture or movement	68	0
Present	0	288
Not much	0	24
Very slight or very trace	0	10
Slight or trace	0	78
Intermittent	0	10
Mild	0	121
Mild to moderate	0	34
Moderate	0	55
Moderate to severe	0	17
Significant	0	32
Severe	0	17
Other, or no indicator of size	0	42

activation of features in the frequency range between 0.7Hz and 1.3Hz were identifying 319 tremor absence vs. presence (blue). 320

4. Discussion and Conclusion

The objective of this study was to assess the ability of several candidate processing 322 pipelines to identify the presence or absence of tremor in kinematic data from movement 323 disorders patients compared to expert ratings from movement disorders specialists. We 324 found generally comparable performance across algorithms; the average F1 score was 325 0.84 ± 0.02 (Mean ±SD; range 0.81 - 0.87), with all F1 confidence intervals overlapping. 326 Notably, the highest performing algorithm (cross-validated F1 = 0.87) was Algorithms 327 A1s, which was a version of the oldest algorithm in clinical use in our center that had been 328 modified such that the manually-engineered features were used as inputs to a modern 329 SVM with radial basis function kernels to accommodate linearly non-separable data. 330

These results suggest some points that may be generally useful in settings with site-331 specific, legacy clinical decision support systems. In particular, in our clinic's implemen-332 tation, the existing algorithms A1r and A2r lacked a clear separation between feature 333 identification and classification steps. We anticipate that this may be the case in other 334 centers with site-specific, legacy systems. Refactoring legacy code to separate these two 335 separate steps may provide an important opportunity to introduce updated classifier archi-336 tectures into these systems without discarding the rich domain knowledge that is embedded 337 in the derivation of engineered features. In our case, although some of the engineered 338 features (e.g., dominant frequency) could be trivially discovered by classifiers with generic 339 spectral features (like B1 and B2), other features (e.g., symmetry of the dominant spectral 340 peak) reflect clinical domain expertise that automated searches could miss given limited 341 training data. We anticipate this as a common issue and recommend that centers utilizing 342 legacy data processing routines refactor their algorithms to distinguish between feature 343 extraction and classification to address this potential limitation and enhance algorithm 344 performance. 345

Further, when we visualized the Receiver-Operator Characteristic curves (Figure 4), 346 we found that the clinical algorithms A1r and A2r were tuned to penalize false positive rate 347 at the expense of some sensitivity in clinical use. Because these algorithms were originally 348 without hyperparameter tuning, this was not done intentionally on the part of the clinicians 349

Version March 13, 2024 submitted to Sensors

10 of 22

Extremity	Absent	Present	Total
Head/Torso	296	56	352
Head	76	52	
Shoulders	71	2	
Thorax	78	1	
Pelvis	71	1	
Arms	767	569	1336
L_Dist_Arm	73	5	
R_Dist_Arm	72	1	
L_Hand	237	280	
R_Hand	242	279	
L_Prox_Arm	71	3	
R_Prox_Arm	72	1	
Legs	481	103	584
L_Dist_Leg	81	21	
R_Dist_Leg	81	8	
L_Foot	76	30	
R_Foot	81	15	
L_Prox_Leg	81	21	
R_Prox_Leg	81	8	

Table 6. Algorithm performance comparison. Performance for hand-tuned algorithms A1r and A2r is reported across the entire dataset. Performance for other numerically optimized algorithms is reported across 5 separate data folds (Mean [95% CI]).

Metric	A1r	A2r	A1s	A2s	B1	B2
F1-score	0.81	0.85	0.87	0.85	0.82	0.82
			[0.85 - 0.89]	[0.82 - 0.88]	[0.78 - 0.86]	[0.76 - 0.88]
Accuracy	0.89	0.91	0.92	0.91	0.89	0.92
2			[0.91 - 0.93]	[0.89 - 0.93]	[0.87 - 0.91]	[0.89 - 0.95]
Precision	0.87	0.93	0.93	0.89	0.91	0.85
			[0.90 - 0.96]	[0.86 - 0.92]	[0.89 - 0.93]	[0.80 - 0.90]
Recall	0.75	0.78	0.81	0.80	0.78	0.80
			[0.79 - 0.83]	[0.77 - 0.83]	[0.73 - 0.83]	[0.70 - 0.90]

using these tools. Refactoring code could give clinicians the opportunity to tailor the 350 balance of precision and recall to the clinical task at hand. 351

Analysis of SHAP plots revealed interesting information about the spectral composi-352 tion of tremor. We noted that high activation of features in the frequency range between 353 4.3 Hz and 7.0 Hz were identified as significant for identifying tremor presence vs. absence 354 (red), while high activation of features in the frequency range between 0.7Hz and 1.3Hz 355 were identifying tremor absence vs. presence (blue) (Figure 5). This is consistent with 356 literature using various sensing modalities that have described tremor [2] as producing 357 frequency band activity around 5Hz, with voluntary movement producing lower frequency 358 band activity below 3Hz. In particular, these results show that low frequency movement is 359 not informative for detecting tremor; in fact, these frequencies have negative predictive 360 value, suggesting that voluntary movements have the potential to be interpreted as false 361 positives. 362

This was an informative result, as both clinical algorithms A1r and A2r were designed 363 with features engineered to capture spectral information that was informative about tremor 364 presence (presumably around 5Hz), but imposed no penalties on low frequency information 365 that indicated that it was absent. This could be interpreted to mean that the original 366 developers of these algorithms exhibited some cognitive confirmation or similar bias when 367

Version March 13, 2024 submitted to Sensors

11 of 22



Figure 2. Example of tremor identification with algorithm A1r. Algorithm A1r operates on each kinematic marker on a given extremity, and estimates the central frequency (Hz) and spectral power density (db/Hz) of the highest-amplitude tremor observed across markers.

designing the features to represent the aspects of the behavior they "knew about," while 368 neglecting equivalent kinematic information that was informative about the absence of 369 tremor. The ability of modern ML to discover features may provide a unique opportunity 370 to complement engineered features created with domain expertise.

4.1. Limitations

Our primary aim was to develop a generic tremor identification algorithm that could be 373 used across extremities, behavioral tasks, and diagnoses. Although the resulting algorithm 374 is almost certainly not optimal in all settings, this approach generally aligns with clinical 375 best practices and represents an important first step in the development of a comprehensive 376 clinical decision support tool for tremor. However, this necessarily comes with the limitation 377 that this tool may not be appropriate for all tremor identification tasks. 378

Further, our tremor assessment approach does not use scripted voluntary movements 379 and weight application in order to isolate mechanical, volitional, and pathophysiological 380 causes of tremor [28]. 381

4.2. Unique contributions

Our method assesses tremor across the body, a unique capability. In a recent machine 383 learning review on tremor applications, only 14% (5/37) explored body parts beyond 384 hands or distal arms [1]. This instrumentation is certainly convenient and almost certainly 385 sufficient for tremor characterization with frequency [2] or amplitude [29,30]. However, 386 we know that signal processing approaches like correlation across body regions provide 387 additional diagnostic insight for discriminating, for example, parkinsonian from orthostatic 388 tremors [28]. With full body data, end-to-end machine learning approaches (e.g., [9]) have 389 significant potential to discover these and other features automatically. Other more subtle 390 tremor features like distractibility [28] seem more likely to be characterized in full body 391 data. Further, our testing approach imposes few, if any constraints on the participants' 392

371 372

Version March 13, 2024 submitted to Sensors

12 of 22



Figure 3. Example of tremor identification with algorithm A2r. Algorithm A2r operates simultaneously on all kinematic markers on a given extremity, and estimates the central frequency (Hz) and amplitude (mm) of the highest-amplitude tremor present.

natural movements. This complicates data analysis compared to methods that confine movements to a single plane, (e.g., [13]) but may improve external validity.

4.3. Conclusion

Here, we sought to assess the ability of several candidate processing pipelines to 396 identify the presence or absence of tremor in kinematic data from movement disorders 397 patients compared to expert ratings from movement disorders specialists. We found 398 that many solutions offered acceptable performance. The best individual-performing 399 algorithm was a modernization of one of the oldest algorithms in constant clinical use in 400 our center. In general, updating legacy clinical decision support systems to incorporate 401 modern machine learning classifiers may result in better performing tools and associated 402 decreases in provider time and improved outcomes. 403

Funding: This study was supported by philanthropic funding to SAF.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration 405 of Helsinki, and approved by the Institutional Review Board of Emory University under IRB protocol 406 00002688 approved June 2, 2021.

Informed Consent Statement: Patient consent was waived due to the retrospective nature of this study.

Data Availability Statement: Data are available upon reasonable request from the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

413 414 415

303

394

395

404

407

408

409

410

411

Version March 13, 2024 submitted to Sensors





(b) PRC

Figure 4. The average receiver operating characteristic (ROC) and precision-recall (PRC) curves for the SVM and XGBoost classifiers using spectral features of the spatial positions of the sensors. The shades correspond to ± 1 standard deviations of each curve across the five fold cross-validation. Refer to text for details.



Figure 5. SHAP (SHapley Additive exPlanations) plot illustrating the contribution of each spectral feature across the Nyquist band, to the tremor prediction results. Each column on the plot represents a specific feature's contribution to the prediction. Positive SHAP values drive the model's output towards the tremor class, while negative values drive towards the non-tremor class. The color intensity indicates the magnitude of the feature value, with red denoting high values and blue indicating low values. Notice the significance of the frequency range between 4.3 Hz–7 Hz in identifying tremor. Frequencies below 3 Hz (corresponding to slow motions of the subject) are not informative for detecting tremor.

medRxiv preprint doi: https://doi.org/10.1101/2024.03.13.24304101; this version posted March 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in pernetuity.

perpetuity. It is made available under a CC-BY 4.0 International license .

Version March 13, 2024 submitted to Sensors

14 of 22

CPT	Current Procedural Terminology
ET	Essential tremor
FTM	Fahn-Tolosa-Marin Tremor Rating Scale
MDS-UPDRS-III	Movement Disorder Society-Unified Parkinson's Disease Rating Scale Part III
PD	Parkinson's disease
SHAP	SHapley Additive exPlanations
SVM	Support vector machine
TETRAS	The Essential Tremor Rating Scale

References

- De, A.; Bhatia, K.P.; Volkmann, J.; Peach, R.; Schreglmann, S.R. Machine Learning in Tremor Analysis: Critique and Directions. *Movement Disorders* 2023, *38*, 717–731. https://doi.org/10.1002/mds.29376.
- Bhatia, K.P.; Bain, P.; Bajaj, N.; Elble, R.J.; Hallett, M.; Louis, E.D.; Raethjen, J.; Stamelou, M.; Testa, C.M.; Deuschl, G.; et al. Consensus Statement on the classification of tremors. from the task force on tremor of the International Parkinson and Movement Disorder Society. Movement Disorders: Official Journal of the Movement Disorder Society 2018, 33, 75–87. https://doi.org/10.1002/ mds.27121.
- 3. Deuschl, G.; Bain, P.; Brin, M. Consensus statement of the Movement Disorder Society on Tremor. *Movement Disorders* **1998**, 13, 2–23. https://doi.org/10.1002/mds.870131303.
- Testa, C.M.; Haubenberger, D.; Patel, M.; Caughman, C.Y.; Factor, S.A. Tremor in Medicine and Other Secondary Tremors. In *Tremors*; Oxford University Press, 2022; [https://academic.oup.com/book/0/chapter/369585583/chapterag-pdf/49059420/book_43955_section_369585583.ag.pdf]. https://doi.org/10.1093/med/9780197529652.003.0007.
- 5. Fahn, S.; Tolosa, E.; Marin, C. Clinical Rating Scale for Tremor. In *Parkinson's Disease and Movement Disorders*; 1988; pp. 225–234.
- Goetz, C.; et al. Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): scale presentation and clinimetric testing results. *Movement disorders: official journal of the Movement Disorder Society* 2008, 23, 2129–2170.
- Elble, R.; Comella, C.; Fahn, S.; Hallett, M.; Jankovic, J.; Juncos, J.L.; Lewitt, P.; Lyons, K.; Ondo, W.; Pahwa, R.; et al. Reliability of a New Scale for Essential Tremor. *Mov Disord* 2012, 27, 1567–9, [23032792]. https://doi.org/10.1002/mds.25162.
- 8. Mancini, M.; Shah, V.V.; Stuart, S.; Curtze, C.; Horak, F.B.; Safarpour, D.; Nutt, J.G. Measuring freezing of gait during daily-life: an open-source, wearable sensors approach. *Journal of NeuroEngineering and Rehabilitation* **2021**, *18*, 1–13.
- Kwon, H.; Clifford, G.; Genias, I.; Bernhard, D.; Esper, C.; Factor, S.; McKay, J. An Explainable Spatial-Temporal Graphical Convolutional Network to Score Freezing of Gait in Parkinsonian Patients. *Sensors (Basel)* 2023, 23. Place: Switzerland, https://doi.org/10.3390/s23041766.
- 10. Filtjens, B.; Nieuwboer, A.; D'cruz, N.; Spildooren, J.; Slaets, P.; Vanrumste, B. A data-driven approach for detecting gait events during turning in people with Parkinson's disease and freezing of gait. *Gait Posture* **2020**, *80*, 130–136.
- Li, W.; Chen, X.; Zhang, J.; Lu, J.; Zhang, C.; Bai, H.; Liang, J.; Wang, J.; Du, H.; Xue, G.; et al. Recognition of Freezing of Gait in Parkinson's Disease Based on Machine Vision. *Frontiers in Aging Neuroscience* 2022, 14, 921081. https://doi.org/10.3389/fnagi.20 22.921081.
- Güney, G.; Jansen, T.S.; Dill, S.; Schulz, J.B.; Dafotakis, M.; Hoog Antink, C.; Braczynski, A.K. Video-Based Hand Movement Analysis of Parkinson Patients before and after Medication Using High-Frame-Rate Videos and MediaPipe. *Sensors (Basel, Switzerland)* 2022, 22, 7992. https://doi.org/10.3390/s22207992.
- Friedrich, M.; Roenn, A.J.; Palmisano, C.; Alty, J.; Paschen, S.; Deuschl, G.; Ip, C.W.; Volkmann, J.; Muthuraman, M.; Peach, R.; et al. Visual Perceptive Deep Learning for Smartphone Video-Based Tremor Analysis: VIPER-Tremor. Preprint, In Review, 2023. https://doi.org/10.21203/rs.3.rs-3692906/v1.
- 14. Elble, R.J.; Brilliant, M.; Leffler, K.; Higgins, C. Quantification of essential tremor in writing and drawing. *Movement Disorders: Official Journal of the Movement Disorder Society* **1996**, *11*, 70–78. https://doi.org/10.1002/mds.870110113.
- Elble, R.J.; Pullman, S.L.; Matsumoto, J.Y.; Raethjen, J.; Deuschl, G.; Tintner, R.; Tremor Research Group. Tremor amplitude is logarithmically related to 4- and 5-point tremor rating scales. *Brain: A Journal of Neurology* 2006, 129, 2660–2666. https: //doi.org/10.1093/brain/awl190.
- 16. Bronte-Stewart, H.; Gala, A.; Wilkins, K.; Pettruci, M.; Kehnemouyi, Y.; Velisar, A.; Trager, M. The digital signature of emergent tremor in Parkinson's disease. *Res. Sq.* **2023**.
- 17. Randall, J.E.; Stiles, R.N. POWER SPECTRAL ANALYSIS OF FINGER ACCELERATION TREMOR. *Journal of Applied Physiology* **1964**, *19*, 357–360. https://doi.org/10.1152/jappl.1964.19.2.357.
- Zhang, C.; Wang, G.; Zhao, J.; Gao, P.; Lin, J.; Yang, H. Patient-specific ECG classification based on recurrent neural networks and clustering technique. In Proceedings of the 2017 13th IASTED International Conference on Biomedical Engineering (BioMed).
 IEEE, 2017, pp. 63–67.
- Reyna, M.A.; Kiarashi, Y.; Elola, A.; Oliveira, J.; Renna, F.; Gu, A.; Perez Alday, E.A.; Sadr, N.; Sharma, A.; Kpodonu, J.; et al. Heart murmur detection from phonocardiogram recordings: The George B. Moody PhysioNet Challenge 2022. preprint, Health Informatics, 2022. https://doi.org/10.1101/2022.08.11.22278688.

417

424

425

426

427

428

429

430

431

432

433

434

435

436

437

438

439

440

441

442

443

444

445

446

447

448

449

450

451

452

453

454

455

456

457

458

459

medRxiv preprint doi: https://doi.org/10.1101/2024.03.13.24304101; this version posted March 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

perpetuity. It is made available under a CC-BY 4.0 International license .

Version March 13, 2024 submitted to Sensors

15 of 22

471

472

473

474

475

476

480

- Tripathi, R.; McKay, J.; Esper, C.E. Use of 3D motion capture for kinematic analysis in movement disorders. *Practical Neurology* 466
 2023, *in press*.
- Naiji Gong.; Hyeokhyen Kwon.; Gari Clifford.; Christine Esper.; Stewart Factor.; Johnathan McKay. Phenotyping Motorsubtypes of Parkinsonism from Full-body Kinematics using Machine Learning (P6-11.007). Neurology 2023, 100, 4306. https://doi.org/10.1212/WNL.000000000203887.
- 22. American Medical Association. CPT Current Procedural Terminology. https://www.ama-assn.org/amaone/cpt-current-procedural-terminology. Accessed: 2023-12-28.
- 23. McKay, J.; Goldstein, F.; Sommerfeld, B.; Bernhard, D.; Perez Parra, S.; Factor, S. Freezing of Gait can persist after an acute levodopa challenge in Parkinson's disease. *NPJ Parkinsons Dis* **2019**, *5*, 25. Place: United States, https://doi.org/10.1038/s41531-019-0099-z.
- 24. Cortes, C.; Vapnik, V. Support-vector networks. *Machine learning* **1995**, *20*, 273–297.
- 25. Cortes, C.; Vapnik, V. Support-Vector Networks. Mach. Learn. 1995, 20, 273–297. https://doi.org/10.1023/A:1022627411411.
- 26. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, pp. 785–794, [arxiv:cs/1603.02754]. https://doi.org/10.1145/2939672.2939785.
- 27. Lundberg, S.; Lee, S.I. A Unified Approach to Interpreting Model Predictions, 2017, [arxiv:cs, stat/1705.07874].
- Vial, F.; Kassavetis, P.; Merchant, S.; Haubenberger, D.; Hallett, M. How to Do an Electrophysiological Study of Tremor. *Clinical Neurophysiology Practice* 2019, 4, 134–142. https://doi.org/10.1016/j.cnp.2019.06.002.
- Shaikh, A.G. Tremor analysis separates Parkinson's disease and dopamine receptor blockers induced parkinsonism. *Neurological Sciences* 2017, *38*, 855–863. https://doi.org/10.1007/s10072-017-2852-6.
- Sally R. Williams.; J. Lucas McKay.; Douglas Bernhard.; Mark Saad.; Anh-Thu N. Vu.; Richa Tripathi.; Stewart A. Factor.; Christine D. Esper. Quantitative motion analysis and clinical characteristics of Holmes tremor as compared to other tremor types (S32.008).
 Neurology 2022, *98*, 1842.
- Altman, D.G.; Bland, J.M. Measurement in Medicine: The Analysis of Method Comparison Studies. *Journal of the Royal Statistical Society Series D: The Statistician* 2018, 32, 307–317, [https://academic.oup.com/jrsssd/article-pdf/32/3/307/49920574/jrsssd_32_3_307.pdf]. https://doi.org/10.2307/2987937.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual 491 author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to 492 people or property resulting from any ideas, methods, instructions or products referred to in the content. 493 medRxiv preprint doi: https://doi.org/10.1101/2024.03.13.24304101; this version posted March 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in percetuity.

perpetuity. It is made available under a CC-BY 4.0 International license .

494

Appendix A Comparison of tremor features identified by algorithms A1r and A2r

We performed some additional analyses to compare the tremor features identified by clinical algorithms A1r and A2r. We compared tremor frequencies identified by clinical algorithms A1r and A2r using a Bland-Altman approach [31]. Because each of the clinical algorithms produced an estimate of tremor amplitude whether or not a tremor was detected, we compared the ranges of amplitudes obtained when tremor was present or absent according to ground-truth labels with two-sample Kolmogorov-Smirnov tests.

Overall, algorithms A1r and A2r identified very similar central tremor frequency estimates, with average values 4.8 ± 1.0 Hz and 4.7 ± 0.6 Hz, respectively. Bland-Altman analysis between the results of the two algorithms identified a bias of -0.1 Hz between the two algorithms, with 95% limits of agreement (-9.0, 0.7) Hz (Figure A.1). 502

With both algorithms, the range of identified amplitudes for which tremor was rated 505 present according to expert labels had some overlap with the range of amplitudes for which 506 tremor was rated absent. This suggests that a simple amplitude-based threshold would be 507 insufficient to discriminate tremor presence using either approach. With algorithm A1r, the 508 average amplitude when tremor was present was $|102.2 \pm 13.40, 1.9 - 944.7|$ mm/s ([Mean 509 \pm SD, range]) compared to [26.4 \pm 35.1, 0.4 – 199.0] mm/s when tremor was absent. With 510 algorithm A2r, the average amplitude when tremor was present was $[3.07 \pm 3.00, 0.3 - 24.3]$ 511 mm compared to $[0.13 \pm 0.12, 0.01 - 0.59]$ mm when tremor was absent. The cumulative 512 densities identified by both algorithms showed separation between cases labeled as tremor 513 absent and present and highly significant (P << 0.001) two-sample Kolmogorov-Smirnov 514 tests. Visual inspection of cumulative amplitude distributions (Figure A.2) for the two 515 algorithms suggested that A2r provided better separation, although this could not be 516 compared directly due to the different units used by the algorithms. 517

Version March 13, 2024 submitted to Sensors



Figure A.1. Comparison of tremor frequencies identified by clinical algorithms A1r and A2r.



Figure A.2. Comparison of tremor amplitudes identified by clinical algorithms A1r and A2r, stratified by ground-truth labels of tremor presence or absence.

medRxiv preprint doi: https://doi.org/10.1101/2024.03.13.24304101; this version posted March 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

perpetuity. It is made available under a CC-BY 4.0 International license .

18 of 22

518

Appendix B Additional dataset details

This section provides some additional details on the data format and coding scheme. 519 During each individual behavioral test, the laboratory 3D motion capture system records 520 the instantaneous position of all kinematic markers on the body (typically 60) and exports 521 these data to a standard *.trc tabular data format with some minimal header information. A 522 typical *.trc* file for a 30-second recording at 120 Hz comprises 3600 rows (30 seconds \times 120 523 Hz) and 180 columns (60 markers \times 3 axes) of kinematic data. Because each file includes 524 data from markers on different extremities, for which tremor may be absent or present on a 525 give trial, the columns of the *.trc file corresponding to markers on each extremity must be 526 separated prior to analysis. Our clinical data processing pipeline maintains the mappings 527 between kinematic markers and extremities in an *.xml file (markers.xml). To avoid the 528 burden of parsing these files, the data supplied with the paper are provided in two different 529 formats. Each deidentified .trc file is provided as originally exported, as well as divided 530 into separate .csv files for each body extremity in the accompanying dataset. These files are 531 compatible with standard Python, R, Matlab, or similar software libraries. Summaries of 532 the contents of example files are provided in Table B.1. Descriptions of kinematic marker 533 locations are provided in Tables B.3 through B.5. 534

Information about the testing condition used during each recording is provided as part of the individual file names, using the nomenclature provided in Table B.2. For example, the file *data/HH/std-arms-extended1-TP.trc* designates participant HH standing with arms extended forward along the X axis of the laboratory 1. In pointing and spiral movement tasks, the suffix *right* or *left*, or 1 or 2, are appended to the filename denoting the extremity involved. In some cases, extra motor or cognitive tasks were introduced to intensify tremor provocation, with supplementary information appended to the base task codes.

Example file	std-arms-extended1-	std-arms-extended1-
	TP.trc	TP/R_Hand.csv
Description	Tabular data with header exported by motion capture soft- ware	Portion of .trc file cor- responding to extrem- ity R_Hand
Columns		
Number	182	Variable
Contents	Index	Index
	Time (seconds)	Time (seconds)
	Whole-body kinematic	Extremity kinematic
marker data arranged		marker data from
	as x, y, z	R_Hand arranged as x,
	-	y, z
Load methods	loadTrc.m	readtable.m
		pandas.read_csv
		 readr::read_csv

Table B.1. Example file descriptions and load methods. In some cases .trc files contain additional columns with derived variables that should be ignored.

Table B.2. Nomenclature for behavioral tasks employed in testing.

Code	Task
sit-rest	Seated with arms at sides
sit-arms-extended	Seated, with arms extended anteriorly and parallel to the floor
sit-UEopp	Seated, with arms in a "T" pose parallel to the ground with fingers of each hand opposed
sit-point	Seated, performing a finger-to-nose pointing task with the indicated extremity (<i>right/1</i> or <i>left/2</i>)
sit-spiral	Seated, performing a spiral movement with the indicated extremity (<i>right/1</i> or <i>left/2</i>)
std-rest	Standing with arms at sides
std-arms-extended	Standing, with arms extended out parallel to the ground
std-UEopp	Standing, with arms in a "T" pose parallel to the ground with fingers of each hand opposed
walk-thru	Comfortable walking from one end to the other of the motion capture space
TUG	Sequential "timed up and go" walking tasks

Table B.3. Kinematic marker descriptions for markers on the trunk. Markers that appear on both sides of the body are listed for the right side only and are coded beginning with "R." Replacing this character with "L" will designate the corresponding marker on the left side of the body.

Extremity	Marker Code	Description
Head	Front_Head	Center of forehead, on cap
	JAW	Mental protuberance
	Jaw	Mental protuberance
	RBHD	Right back head, on cap
	RFHD	Right front head, on cap
	Rear_Head	Rear of head, on cap
	TopHead	Top of head
	Top_Head	Top of head
Shoulders	C7	Seventh cervical vertebra
	RBAK	Right scapula (asymmetry marker)
	RSHO	Right acromioclavicular joint
	R_Shoulder	Right acromion process
	STRN	Xiphoid process
Pelvis	LASI	Left anterior superior iliac spine
	RASI	Right anterior superior iliac spine
	RIC	Right iliac crest
	RPSI	Right posterior superior iliac spine
	R_ASIS	Right anterior superior iliac spine
	V_Sacral	Sacrum
Thorax	CLAV	Clavicular notch
	R_Clavicle	Right clavicle
	R_Scap_Inf	Right scapula inferior angle
	R_Scapula	Right supraspinous fossa
	T10	10th thoracic vertebra

Table B.4. Kinematic marker descriptions for markers on the arms. Markers that appear on both sides of the body are listed for the right side only and are coded beginning with "R." Replacing this character with "L" will designate the corresponding marker on the left side of the body.

Extremity	Marker Code	Description
R_Dist_Arm	RFRM	Lateral surface of forearm
	RWRA	Radial side of wrist
	RWRB	Ulnar side of wrist
	R_Forearm	Lateral surface of forearm
	R_Radius	Right styloid process of radius
	R_Ulna	Mid region of ulna
R_Hand	RFIN	Third finger, first metacarpal joint
	RFINGM2	Third finger, second metacarpal joint
	RFINGM3	Third finger, most distal segment
	RTHM1	Thumb, first metacarpal
	RTHM2	Thumb, second metacarpal
	RTHM3	Thumb, most distal segment
	R_Finger3_M1	Third finger, first metacarpal joint
	R_Finger3_M2	Third finger, second metacarpal joint
	R_Finger3_M3	Third finger, most distal segment
	R_Hand	Radial surface of wrist
	R_Thumb_M1	Thumb, first metacarpal
	R_Thumb_M2	Thumb, second metacarpal
	R_Thumb_M3	Thumb, most distal segment
	R_Wrist	Radial surface of wrist
R_Prox_Arm	RELB	Right lateral epicondyle
	R_BicepsLateral	Lateral surface of upper arm
	R_Biceps_Lateral	Lateral surface of upper arm
	R_Elbow	Right lateral epicondyle
	R_Elbow_Medial	Right medial epicondyle

Table B.5. Kinematic marker descriptions for markers on the legs. Markers that appear on both sides of the body are listed for the right side only and are coded beginning with "R." Replacing this character with "L" will designate the corresponding marker on the left side of the body.

Extremity	Marker Code	Description
R_Dist_Leg	RANK	Lateral aspect of ankle
C	RANKM	Medial aspect of ankle
	RTIB	Midpoint of tibia
	R_Ankle	Lateral aspect of ankle
	R_Ankle_Medial	Medial aspect of ankle
	R_Shank	Midpoint of tibia
R_Foot	RFTM	Dorsal/medial surface of foot midway
		between ankle and toe
	RHEE	Distal surface of heel
	RTOE	Third metatarsal
	R_Hallux	Dorsal surface of big toe
	R_Heel	Distal surface of heel
	R_MedFoot	Dorsal/medial surface of foot midway
		between ankle and toe
	R_Toe	Third metatarsal
R_Prox_Leg	RKNE	Lateral aspect of flexion-extension axis of knee
	RKNEM	Medial aspect of flexion-extension axis
		of knee
	RTHI	Upper lateral 1/3 surface of thigh
	R_Knee	Lateral aspect of flexion-extension axis of knee
	R_Knee_Medial	Medial aspect of flexion-extension axis of knee