IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received November 18, 2020, accepted January 7, 2021, date of publication January 18, 2021, date of current version January 27, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3052186

Computerized Screening of Essential Tremor and Level of Severity Using Consumer Tablet

MOHAMMOD ABDUL MOTIN[1], (Member, IEEE), JAMES PETERS[2], LAURA PERJU-DUMBRAVA[2], CATHERINE DING[2], MICHAEL ELLER[2], SANJAY RAGHAV[2], SHEIK MOHAMMED ALI[1], (Graduate Student Member, IEEE), PETER KEMPSTER[3], PJ RADCLIFFE[1], AND DINESH KANT KUMAR[1], (Senior Member, IEEE)

[1]Department of Electrical and Biomedical Engineering, School of Engineering, RMIT University, Melbourne, VIC 3001, Australia
[2]Monash Medical Centre, Department of Neurosciences, Melbourne, VIC 3168, Australia
[3]Department of Medicine, School of Clinical Sciences, Monash University, Clayton, VIC 3800, Australia

Corresponding author: Dinesh Kant Kumar (dinesh.kumar@rmit.edu.au)

ABSTRACT Essential tremor (ET) is diagnosed and monitored by movement disorder specialists based on clinical observations. While many ET cases are benign, some require pharmacological and surgical management, and there is a need for tools to assist clinicians in making informed decisions. This work aimed to develop a computerized technique to detect the presence and severity of ET. A set of 6 writing and sketching tasks were performed by 39 subjects on a digital tablet. The position and pressure of contact during the sketching were recorded and analyzed to obtain the dynamics of drawing. ET patients were scored on the Fahn-Tolosa-Marin Tremor Rating Scale by blinded movement disorder neurologists, and then separated into two groups: moderate and severe ET. Drawing tasks were more effective than writing tasks in distinguishing the groups, with drawing horizontal and vertical lines being the most sensitive. A new set of composite index feature was found to be most suitable in separating the three groups, with a Spearman correlation coefficient of 0.72. The technique shows significant differences between controls, patients with moderate tremor and those with severe tremor, with an accuracy of 87.2%. Our computerized analysis significantly outperformed non-specialist clinicians in differentiating ET from control. We conclude that computerized analysis of the dynamics of sketching horizontal and vertical lines is a suitable method to assess the presence and severity of ET.

INDEX TERMS Computerized diagnosis, essential tremor, feature selection, task selection, writing and sketching task.

I. INTRODUCTION

Tremor is an involuntary, rhythmic, oscillatory movement of a body part. Most people have a slight tremor of their unsupported limbs, only noticeable in circumstances of anxiety or fatigue. Pathological tremor, on the other hand, is visible and persistent. The most common form of pathological tremor is essential tremor (ET) [1], [2], which affects roughly 2% of the population [3]. ET is obvious when the limbs are held in posture, affecting tasks such as carrying a cup, feeding, and writing [4]–[6]. A family history of tremor is common in ET. Some patients report improvement with small amounts of alcohol. While the tremor may be intrusive, the disorder is not otherwise disabling in many cases and runs a benign course. A minority, though, has a severe tremor that impairs functional activities of independent living [7]–[9]. Tremor amplitude increases slowly over the decades, so late in life, ET can become more disabling.

At a clinical level, ET has two major problems. Firstly, the disorder lacks a gold standard diagnostic test, and there is uncertainty about the limits of its classification and even whether it represents one or several underlying disease entities [3]. There are grey zones with other tremor disorders such as dystonic tremor and tremor-dominant Parkinson's disease. Imaging-based biomarkers for ET have been investigated, but there are limitations from cost and inconclusive results [10], [11]. Secondly, there is an unmet need for better treatments. Currently prescribed drugs are modestly effective at best. At the severe end of the ET spectrum, functional neurosurgical or interventional neuroradiological procedures for ET exist, yet there is uncertainty around the optimum patient selection. In both of these areas, reproducible, quantitative measurements can assist research into ET.

Simple pen and paper tests are often performed in the clinic, both as a diagnostic aid for ET and as a rough

The associate editor coordinating the review of this manuscript and approving it for publication was Nuno Garcia.

measure of severity for monitoring purposes [4]. Computerized methods using digital tablets [5]–[12] or wearable sensors [13]–[17] have potential as sensitive and quantitative assessments of writing and drawing in ET. This could augment, or even outperform, the standard clinical tools for tremor grading, and would be of particular use in clinical trials of pharmacological treatments of ET.

Spiral drawing is impaired in ET, and a number of studies have identified features corresponding to the shape of the spiral that makes the test more objective. One observation has been the altered axis alignment during spiral sketching in ET cases [12], [18]. The Spiral width variability index is the measure of loop-to-loop spiral width variation, which was reported as a metric for screening ET [8]. Inter-spiral tightness variability based on the 25%–75% range in tightness across hand-sketched spirals has been proposed as a marker of functional (psychogenic) tremor [19]. An alternate method to detect the presence of tremors is the use discrete cosine transform features from hand sketched spiral [5].

Digital tablets provide spatio-temporal location, pen-pressure, azimuth, and altitudes information while writing or sketching. This provides information on the dynamics of strokes and hand-movements such as tremors and has been reported to be more sensitive than shape-based analysis for detecting tremors [5]–[12], [20]. The dynamics of sketching or writing provide movement and pressure information such as speed, acceleration, and pressure variability which can be used to identify the presence of tremor, its amplitude, and frequency [21]. Fourier transform of velocity has been applied to capture the effect of tremor in [22]. However, these dynamic-based techniques used only hand-sketched spirals and most of them did not include shape-based or static information. A few studies [7], [23] have combined shape and temporal information to classify ET and controls. Entropy and fractal dimension based features were applied to the time series of shape information to determine the differences between ET and controls [6], [9]. However, no study has effectively used the dynamics of multiple sketching tasks to identify the presence and severity of the disease.

The aim of this project was to develop a computerized technique for the detection and assessment of the severity of ET. We have investigated the dynamics of 6 handwriting and sketching tasks to identify the tasks and features that best differentiate between controls and ET, and between ET of moderate and high severity. Novel composite index features were developed to enhance the separation between the groups. The recordings were also visually assessed by non-specialist clinicians for comparison with the computerized technique.

II. MATERIALS AND METHODS

A. PARTICIPANTS

Ten men and nine women diagnosed with ET were recruited from the Movement Disorders Service at Monash Health. Their mean age was 67.2 ± 13.0 and the mean duration of tremor symptoms was 21.7 ± 19.0 years. All complied with the axis 1 definition of ET in the 2018 Consensus Statement on the classification of tremors [3]. No subject met any of the Axis 1 exclusion criteria for ET. Twenty healthy participants (9 men and 11 women) acted as controls for automated writing and drawing tasks. Their mean age was 64.3 ± 12.0.

The study was conducted in accordance with the human experiments Helsinki Declaration (revised 2004) and approved by the Monash Health and RMIT University Human Research Ethics Committees (HREC Project Number: 184981). All participants in this study gave their written informed consent before data recording. The demographic and clinical information of healthy control and ET participants are shown in Table 1.

TABLE 1. Demographic and clinical information of the age-matched healthy controls and ET patients.

Demographic	Control	Essential Tremor
Number of Participants	20	19
Age (years)	64.3±12.0	67.2 ±13.0
Gender (Male, Female)	9, 11	10, 9
Dominant Hand (right, left)	20, 0	14, 5
Disease Duration (years)	N/A	21.7±19
Mean Tremor Scale (FTM)	N/A	26.52±14.33

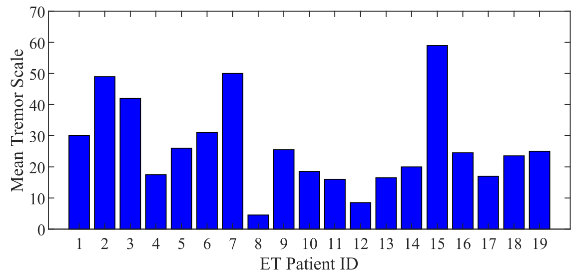


FIGURE 1. Average Fahn-Tolosa-Marin (FTM) tremor scale of each ET patients. The average FTM tremor scale varies from 3.2 to 59.

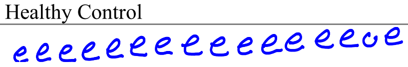
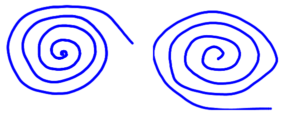
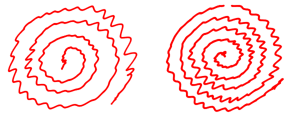
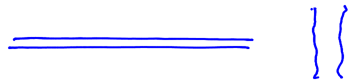
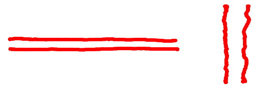
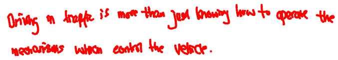
TABLE 2. Demographic and clinical information of moderate ET and Severe ET patients.

Demographic	Moderate ET	Severe ET
Number of Participants	13	6
Age (years)	65.5±14.8	70.6 ±8.0
Gender (Male, Female)	5, 8	5, 1
Dominant Hand (right, left)	8, 5	6, 0
Mean Tremor Scale (FTM)	18.69±6.58	43.50±11.43

B. ET SEVERITY SCORING

Scoring on the Fahn-Tolosa-Marin (FTM) Tremor Rating Scale [24] was performed by two blinded assessors from videotaped recordings. The FTM score of each ET patients are shown in Fig. 1. The ET patients were grouped into moderate ET (FTM $<$ 30) and severe ET (FTM greater than or equal to 30). The demographic and clinical information of moderate and severe ET patients are shown in Table 2.

TABLE 3. The details of the six different tasks that were performed by the age-matched control subjects and ET patients. Among six different tasks, four were naturally handwritten tasks and two were guided hand-sketched tasks.

Input Signal	Description	Sample Image Recorded from Age-Matched Healthy Control	Sample Image Recorded from ET patients
Task 1	Writing letter "e" several times		
Task 2	Writing letters "bd" several times		
Task 3	a. Drawing a spiral in a clockwise direction b. Drawing a spiral in an anti-clockwise direction		
Task 4	Drawing two horizontal and vertical line		
Task 5	Writing a sentence		
Task 6	Writing name		

C. DATA ACQUISITION

Dynamics of handwriting and sketching were recorded using a digital tablet (Wacom Intuos Pro Large, A3 sized) with a pressure-sensor mounted ink-pen. This was chosen because it provides the user with the feel of a regular pen and with a large sheet of paper and is perceived as comfortable and convenient to elderly subjects. Customized software was developed in *c#* which was then integrated into the tablet to record the pen trajectories *(x, y)*, pen tip pressure between the pen and the tablet surface, azimuth–the angle between pen and vertical plane of pad surface, altitude–the angle between pen and pad, and time stamp. The data was recorded at 133 Hz of sampling rate, analyzed in real-time to obtain the dynamic measures using the customized software and stored as.csv files. The tablet was placed as was comfortable to the participants.

D. WRITING AND DRAWING TASKS

The data was collected when the participants performed six different writing and drawing tasks (Table 3). The first task required repeated writing of the letter "e" which is considered as a basic assessment of fine motor skills of writing [25], [26]. The second task was writing "bd" repeatedly as the writing strokes are distinctively affected by tremor [27]. Both of these tasks have letters with vertical strokes. Task 3 consisted of drawing a spiral between dotted lines in a clockwise (3a) and then anti-clockwise (3b) direction. In Task 4, the participant again drew between dotted lines; this time two horizontal and vertical lines. The drawing tasks with displayed dots are language independent and less dependent on visual feedback. Task 5 was the handwriting task that required attention and visuospatial memory compared with task 1 and 2. Writing one's signature was Task 6, and this represents the natural writing style of the person.

E. VISUAL ASSESSMENT

Two independent medical professionals who were not neurologists volunteered to examine the Task 3a and 3b drawings. This assessment was included to give a broad comparison with the ability of the non-specialized eye to detect features of ET on a standard clinical drawing task. Each assessor was given a description [4] with illustrative samples of the use of pen-and-paper drawing assessments in the diagnosis of tremor disorders. They were then provided the de-identified static spiral images with the corresponding duration of spiral drawing and asked to label these based on whether it was drawn by a control or ET participant.

F. PRE-PROCESSING

The data were segmented using the pen-tip pressure to separate on tablet strokes from the on-air movement; pressure = 0 labeled as 'in-air', while pressure > 0 as 'on tablet'. To remove outliers and artefacts caused by accidental touch of the pen or wrist on the tablet, two steps of pre-processing was conducted:

Step 1: The total length d of each segment was calculated as follows:

$$d = \sum_{i}^{N} \sqrt{(x_i - x_0)^2 + (y_i - y_0)^2}$$

where (x_0, y_0) is the starting point and (x_i, y_i) is the i^{th} sample in the cartesian coordinate trajectory while sketching, $i = 1, 2 \ldots\ldots N$. N is the total number of samples in each segment. The segments of length, $d <$ 2mm were considered as noise for task 3 and 4 and were discarded–2mm chosen empirically.

Step 2: To remove points due to accidental contact on the tablet, median filtering was applied. If the distance between two adjacent points is five times higher than the median value of the distances between the previous five adjacent points, was considered as accidental touch and removed from the time series.

G. FEATURE EXTRACTION FROM WRITING AND DRAWING TASKS

Seventy-nine features were computed from handwriting and drawing tasks- details of these are in Table 4. The 79 features

TABLE 4. A set of 79 computed features from six different writing and drawing tasks.

Features Type	Feature Acronym	Feature Description	Features No
Static Features	*avgStrLenAir, medStrLenAir, maxStrLenAir, minStrLenAir, stdStrLenAir, iqrStrLenAir, avgStrLen, medStrLen, maxStrLen, minStrLen, stdStrLen, iqrStrLen*	Average, median, maximum, minimum, standard deviation, and interquartile range of pen stroke length in air and on tablet surface	6×2=12
Static Features	*inAirT, onSurT, totalT, ratioAirtoSurT, satPreDur, noPause,*	The time in air, on tablet surface, total time, ration of air to surface, time duration with saturated pressure, and number of pauses.	6
Velocity	*avgVx, medVx, maxVx, stdVx, avgVy, medVy, maxVy, stdVy,* ratioV_xV_y	Average, median, maximum, and standard deviation of pen tip velocity in *x* and *y* direction. Ratio of pen tip median velocity in *x* and *y*.	4×2+1=9
Acceleration	*avgAx, medAx, maxAx, stdAx, avgAy, medAy, maxAy, stdAy, ratioAxAy*	Average, median, maximum, and standard deviation of pen tip acceleration in *x* and *y* direction. Ratio of median acceleration in *x* and *y*.	4×2+1=9
Jerk	*avgJx, medJx, maxJx, stdJx, avgJy, medJy, maxJy, stdJy, ratioJxJy*	Average, median, maximum, and standard deviation of pen tip jerk in *x* and *y* direction. Ratio of median jerk in *x* and *y*.	4×2+1=9
Speed	*avgS, medS, maxS, stdS*	Average, median, maximum, and standard deviation of pen tip speed.	4
Angular Speed	*avgAngS, medAngS, maxAngS, stdAngS,*	Average, median, maximum, and standard deviation of pen tip angular speed.	4
Directional Changes	$D_x, D_y, D_{xy}, Diff_{xy}$	Total number of time direction changes in *x* direction, *y* direction, both *x* or *y* direction, and absolute differences of direction changes in either *x* or *y* direction	4
Pen Pressure	*avgP, medP, maxP, stdP, skewP, kurtP,*	Average, median, maximum, and standard deviation of pen tip pressure applied on the tablet surface.	6
Pen Pressure Differences	*avgDiffP, medDiffP, maxDiffP, stdDiffP, skewDiffP, kurtDiffP*	Average, median, maximum, and standard deviation of pen tip pressure difference applied on the tablet surface.	6
Composite index	*CIsPDT, CIaPDT, CImPDT, CIPS, CIPST, CIPD, CIPA, CIDA*	Composite index of standard deviation of pressure difference and total time, average of pressure difference and total time, median of pressure difference and total time, pressure and speed, speed, average pressure, speed and task execution time, average pressure and bi-directional changes, average pressure and median acceleration in *x*, and bi-directional change and median acceleration in *x*	8

can be broadly classified into four categories: 18 static, 39 dynamic, 12 pressure, and 8 composite indices.

H. FEATURE SELECTION

An appropriate number of features improve the classification of data, but excessive numbers can result in overtraining, increased error and computational complexity [28]. Feature selection was performed using statistical tests to identify the most suitable set of features. The distribution of the data using Shapiro-Wilk test [29] was found not to be normal and non-parametric distribution-free Kruskal-Wallis test was applied to identify statistically significant features. The features with $p < 0.05$ for between-group differences were selected [30]. These features were ranked based on Spearman correlation coefficient and the top ten were selected for classification analysis. All computation, including statistical analysis, was performed using Matlab2018b (MathWorks).

I. SUPPORT VECTOR MACHINE

The selected ten features were classified by support vector machine (SVM) for both two-class and multi-class classification. In the multi-class problem, SVM was used to classify control vs moderate ET vs severity of ET and architecture is shown in Fig. 2a.

J. HIERARCHICAL SUPPORT VECTOR MACHINE

The twenty selected features were incorporated in H-SVM as a two-layer problem where the first layer detects the ET, and the second layer classifies ET to moderate and severe. SVM based multi-layer hierarchical multi-class classification as shown in Fig. 2b.

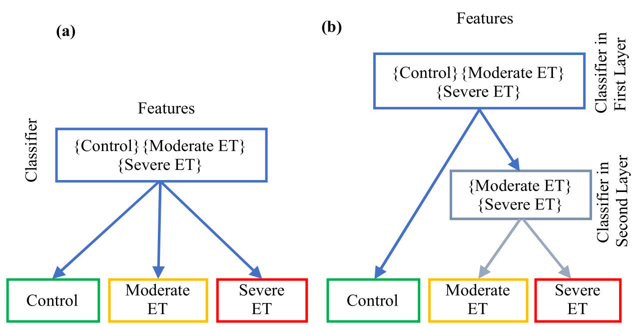


FIGURE 2. Multi-class classification model using (a) Single layer and (b) Multi-layer hierarchical classifier. In hierarchical classifier, the first layer is used to distinguish control vs moderate ET and severe ET and the second layer for moderate ET vs severe ET.

The first layer classifies the control vs ET, and the second layer classifies moderate ET vs severe ET. Instead of using the same feature set, the top ten distinctive features of each layer were used for classification. Compared to a single-layer multi-class classifier, the multi-layer hierarchical classifier has been reported to have better classification accuracy and robustness [31]. It has also been found to be less sensitive to the datasets being imbalanced or skewed [31].

K. MODEL TRAINING AND TESTING PROCEDURE

In this study, we have employed a leave one subject out cross validation where each subject was picked from the pool of the whole dataset and the rest are used for training the model.

TABLE 5. The performance evolutions of six different tasks for two class classification with leave one subject out cross-validation.

Task Name	Accuracy (%)	Specificity (%)	Sensitivity (%)	F score
Task 1	69.23	70.00	68.42	0.69
Task 2	66.66	70.00	63.15	0.66
Task 3a	76.92	85.00	68.42	0.76
Task 3b	79.48	90.00	68.42	0.78
Task 4	**87.18**	**90.00**	**84.21**	**0.87**
Task 5	71.80	90.00	52.63	0.62
Task 6	64.10	75.00	52.63	0.62

The average performance of each subject is reported as overall performance.

L. PERFORMANCE METRICS

In this study, four statistical measures called accuracy, sensitivity, specificity and F score were applied to evaluate the performance of the proposed model, which were calculated as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 10,$$

$$Sensitivity = \frac{TP}{TP + FN} \times 100,$$

$$Specificity = \frac{TN}{TN + FP} \times 100,$$

$$\text{F Score} = 2 \times \frac{Sensitivity \times Sensitivity}{Sensitivity + Sensitivity} \times 100,$$

where TP (true positive) is the number of ET subject detected as ET, TN (true negative) is the number of controls detected as control, FP (False positive) is the number of controls detected as ET, and FN (False Negative) is the number of ET detected as control.

III. RESULTS

The classification result of leave-one-out validation for two-class (control vs ET) and multi-class (control vs moderate ET and severe ET) using support vector machine (SVM) classifier with linear kernel have been presented in Table 5 and Table 6 respectively. For two-class, the classification accuracy using writing tasks (task 1, 2, 5 and 6) varies from 64.10% to 71.80% while for drawing tasks (task 3a, 3b and 4) it is in the range of 76.92% to 84.62%. The details of the writing and drawing tasks are presented in Method Section. The sensitivity, specificity, and F score of writing and drawing tasks for two class classification are listed in Table 5. Multi-class classification using multi-class SVM (M-SVM), the overall classification accuracy using writing tasks varies from 43.58% to 53.84% while for drawing tasks from 66.66% to 74.35%. Multi-class classification using proposed Hierarchical SVM (H-SVM) has an accuracy of 53.84% to 64.69% for the writing tasks and 74.36% to 87.18% for the drawing tasks (Table 6).

The screening of drawn spirals was done by visual inspection performed by two independent evaluators and is shown in Table 7. In both cases, the inter-rater variability and misclassification-rate were high, with sensitivity, specificity and accuracy for the two evaluators being 55.5 and 66.6%, 44.4 and 66.6% and 50 and 66.6% respectively. In comparison, the overall performance of the computerized analysis had specificity of 90.00%, sensitivity of 84.21%, and accuracy of 87.18% (Table 7).

TABLE 6. The accuracy of six different tasks for multi-class classification with leave one subject out cross-validation using M-SVM and H-SVM classifier.

Task Name	M-SVM Classifier			
	Overall	Control	Moderate ET	Severe ET
Task 1	53.84	65.00	38.46	50.00
Task 2	50.28	90.00	20.30	50.00
Task 3a	66.66	90.00	23.07	83.33
Task 3b	69.23	90.00	30.76	83.33
Task 4	74.35	85.00	46.15	100.0
Task 5	53.84	65.00	38.46	50.00
Task 6	43.58	65.00	13.60	50.00
Task Name	**H-SVM Classifier**			
	Overall	Control	Moderate ET	Severe ET
Task 1	64.10	85.00	30.76	66.66
Task 2	53.84	55.00	46.15	66.66
Task 3a	74.36	85.00	53.84	83.33
Task 3b	79.48	90.00	53.84	100.0
Task 4	**87.18**	**90.00**	**84.62**	**83.33**
Task 5	66.66	80.00	53.84	50.00
Task 6	53.84	80.00	15.38	50.00

TABLE 7. The Performance evaluation of screening ET from control by visual examination of spiral tasks by two different evaluators. The proposed model predicted responses are compared with the evaluators.

Evaluator	Accuracy (%)	Misclassification Rate (%)	Sensitivity (%)	Specificity (%)
Evaluator-1	50	50	55.5	44.4
Evaluator-2	66.6	33.3	66.6	66.6
Proposed	**87.18**	**12.8**	**84.21**	**90.00**

A. RANK OF THE FEATURES

The features were individually tested for statistical significance using non-parametric distribution-free Kruskal-Wallis test and Spearman's correlation was performed on those with $p < 0.05$. Spearman's rank correlation coefficients of the tasks and their features are presented in Table 8 and 9. The tables show that drawing tasks (3a, 3b, and 4) outperform the writing tasks and thus are more suitable for distinguishing between controls and ET, and ET moderate and severe. The rank of features shows that composite index features (combining both dynamic and shape features) of the drawing tasks have the highest correlation for distinguishing between ET and control. For writing tasks, both shape and dynamic features have comparable correlation coefficients. Comparing moderate ET vs severe ET, the dynamic features obtained from drawing tasks had a higher correlation than the shape-based ones. Of the top ten features of drawing tasks based on Spearman's rank correlation coefficients, most of the features were from dynamic and composite index feature sets. Writing tasks had a lower correlation, and both shape and dynamic features had comparable correlation coefficients.

TABLE 8. The list of the top ten features of six different tasks based on Spearman rank order correlation coefficient to differentiate ET from age-matched healthy controls. The features are represented by the acronym and Spearman correlation coefficient is in the brackets.

Task Name	Feature (correlation coefficient)
Task 1	*stdStrLen (0.48), iqrStrLen (0.47), D_{xy} (0.39), D_y (0.38), maxStrLen (0.37), minStrLen (0.37), onSurT (0.36), D_x (0.35), medStrLen (0.34), ratioA_xA_y (-0.32)*
Task 2	*ratioJ_xJ_y (0.38), maxStrLen (0.37), minStrLen (0.37), ratioJ_xJ_y (0.37), CIDA (0.36), D_{xy} (0.34), medP (0.33), satPreDur (0.33), stdJ_x (-0.33), maxV_x (-0.32)*
Task 3a	*CIDA (0.68), CIPD (0.64), D_{xy} (0.60), D_x (0.60), medA_x (0.58), D_y (0.57), Diff$_{xy}$ (0.55), CIPA (0.53), CIaPDT (0.53), avgAccX (0.52)*
Task 3b	*CIDA (0.66), D_x (0.63), medA_x (0.61), CIPD (0.61), D_{xy} (0.60), D_y (0.60), CIPA (0.56), avgA_x (0.51), medA_y (0.50), skewP (-0.49)*
Task 4	*CIDA (0.72), D_{xy} (0.64), medA_x (0.63), D_y (0.62), CIPD (0.61), medA_y, (0.60), D_x (0.57), CIPA (0.52), medV_y (0.50), CIaPDT (0.45)*
Task 5	*kurtP (0.47), skewP (-0.42), medAngS (-0.37, medStrLen (0.37), medV_x (-0.36), ratioA_xA_y (0.32), D_{xy} (0.32), D_x (0.32, medA_x (-0.32) medDiffP (-0.32)*
Task 6	*maxStrLenAir (0.48), D_{xy} (0.45), D_x (0.44), inAirT (0.43) CIsPDT (0.43), stdStrLenAir (0.42), ratioAST (0.42) CIDA (0.42), D_y (0.41) CIaPDT (0.41)*

TABLE 9. The list of top ten features of six different tasks based on Spearman rank order correlation coefficient to differentiate moderate ET from severe ET. The features are represented by their acronym and the Spearman correlation coefficient in brackets.

Task Name	Feature (correlation coefficient)
Task 1	*satPreDur (0.77), avgP (0.72), medP (0.72), CIPD (0.72), noPause (-0.67), maxP (0.66), medStrLen (0.62), avgStrLen (0.62), medStrLenAir (0.57), D_x (0.54)*
Task 2	*satPreDur (0.76), CIPD (0.74), maxStrLen (0.68), minStrLen (0.68), stdStrLen (0. 68), iqrStrLen (0.66), maxDiffP (0.64), CIPA (0.64), medJ_x (0.64) ratioA_xA_y (0.63)*
Task 3a	*medA_y (0. 87), medJ_y (0. 87), medA_x (0. 83), CIPS (0.80), CIPA (0.80), CIDA (0.80), CIPST (0.79), CIPD (0.79), D_x (0.78), D_{xy} (0.77)*
Task 3b	*medA_y (0. 85), medA_x (0. 84), CIDA (0.81), CIPA (0.81), medJ_y (0.76), D_{xy} (0.72), avgA_x (0.72), avgA_y (0.70), D_y (0.70), avgV_x (0.66)*
Task 4	*medJ_x (0.87), medA_y (0.84), medA_x (0.83), CIPA (0.80), CIDA (0.80), medA_y (0.77), D_x (0.76), CIPA (0.76), medV_y (0.74), CIaPDT (0.74)*
Task 5	*ratioV_xV_y (-0.53), onSurT (-0.43), maxV_x (-0.40), maxA_x (-0.40), maxJ_x (-0.40), D_x (-0.39), D_{xy} (-0.36), maxAngS (-0.35), CIDA (-0.32), D_y (-0.32)*
Task 6	*CIDA (0.72), maxV_y (0.68), maxA_x (0.68), maxJ_x (0.68), CIPST (0.68), CIPD (0.68), D_y (0.66), CIsPDT (0.66), CIaPDT (0.66), D_x (0.64)*

B. ROBUSTNESS OF THE MODEL

The higher sample size is considered necessary for the training to represent the phenomena being modeled. However, with limited labeled data samples, which is often the case with medical data, the appropriateness of the feature selection and thus the resultant model needs to be tested for robustness when trained with the small training set. For this purpose, the minimum number of data points (participants) that were necessary to train the model for accurate classification was determined for each classification problem, control vs ET, and for ET severity. This was only performed for task 4 which was found (Table 5 and 6) to be the most discriminative task.

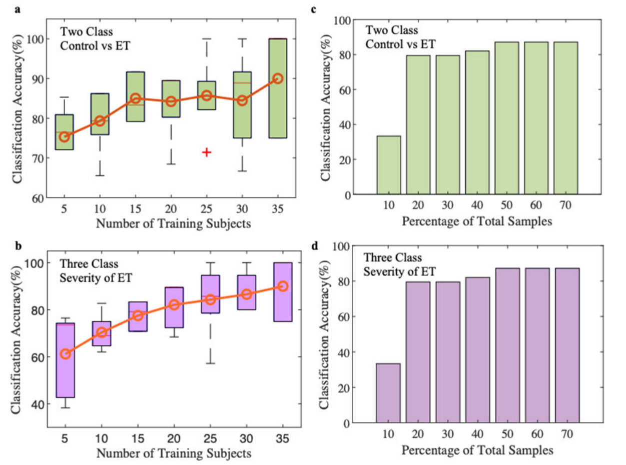


FIGURE 3. Evaluation of model performance with a different number of training subjects and a certain portion of total samples. The boxplot represents the distribution of overall accuracy of the model for a different number of training subjects from 5 to 35 for a) control vs ET and b) control vs moderate ET vs Severe ET. The box represents the 1st, median, and 3rd quartile of the overall accuracy using a varying number of subjects from the training pool randomly for five iterations. (c) control vs ET and d) Severity of ET- show the bar chart the overall classification accuracy based on the percentage of total samples (10% to 70%) of each subject used for training the model and leave-one-out cross-validation.

Two measures were used to test the robustness of the model- (i) based on the number of participants used for training, and (ii) the proportion of the data of each participant. The first step was to obtain the performance by increasing the number of participants from 5 to 35. The training data was a random subset of the participants from each class while the class balance was maintained for training the model. Each step was iterated five times and the results were averaged and the results are shown in Fig. 3a and 3b. The figure shows that accuracy improved with increasing the number of training subjects and plateaued with 20 subjects, with classification accuracy reaching 84.20% for two-class and 82.10% for three-class.

The model robustness was also assessed by determining the minimum amount of data from each participant that is needed to train the model to differentiate between ET and controls and between ET severity levels, and the graphs are demonstrated in Fig. 3c and 3d. The data used for training is shown as a percentage of the full recording. The figure shows that accuracy was about 75% for 20% data and plateaued at 85% when 50% of the data was used to train the model. These two tests demonstrate that the features obtained from ranking (Table 5, 6 and 7) were suitable for identifying ET from controls and moderate ET from severe ET.

IV. DISCUSSION

There is a need for computerized techniques to detect and monitor the severity of ET. This study investigated features of different writing and sketching tasks for computerized detection and assessment of ET. The results show that for standardized sketching tasks, computerized analysis shows a good correlation with the presence of ET and the severity of the disease. This technique, therefore, shows promise for applications that require objective intra-individual assessments of severity, blinded clinical trials of new therapies or measuring the outcome of pharmacological or surgical interventions. The degree by which the computerized analysis outperformed visual assessment by non-specialized physicians emphasises the potential value of our approach. The selection of the tasks, the associated features, and the method of classification are discussed below:

A. TASK SELECTION

An important observation of this study is that the drawing tasks outperformed the writing task in differentiating between control and ET, and between moderate and severe ET, with the sketching of horizontal and vertical lines (Task 4) being the most effective. The drawing of the horizontal and vertical lines requires long strokes which was found to be the most effective task for discriminating between the groups. Among the writing tasks, the group difference was the highest for task 5 (writing a long sentence). This task has a cognitive loading component which could cause psychological stress and enhance the tremor in ET patients.

B. SKETCHING AND WRITING FEATURES

The standard deviation of stroke-length on the tablet surface (*stdStrLen)* was higher in ET compared with controls indicating that there was large variability in their stroke-length. It was also observed that the maximum stroke length *(maxStrLenAir)* in the air was significantly higher in ET. This indicates that when the ET patients write, they spend significantly more time hovering above the tablet, which could be because of their poor fine-motor control, indecisiveness or both. Severe ET patients have constant high pressure*(satPreDur)* while writing. This could relate to impairments of repetitive movement in ET that are independent of the tremor [32].

Among the drawing tasks, it was found that the difference between ET and controls was most significant for *CIDA*, the composite index of bi-directional changes and median acceleration in x. Higher *CIDA* indicates either a greater number of total directional changes or larger acceleration- both of which would indicate the presence of tremor. There was a significant difference in the acceleration ($medA_x$, $medA_y$) and jerk ($medJ_x$, $medJ_y$) between moderate and severe ET. This confirms that the severity of ET is based on the amplitude and frequency of tremor.

C. CLASSIFICATION

The statistical analysis shows that the feature set with significant differences between ET and control is different from the feature set required to differentiate between moderate and severe ET. This motivated the use of the hierarchical classification method, where the first step was to detect ET from control, and the next step was to detect the severity of the disease. Such an approach also has robustness on problems such as class imbalance and skewed datasets [31]. The incorporation of the hierarchical multi-layer classifier improves the accuracy ∼13% over conventional single layer multi-class classifier (see Table 6). However, the performance metrics for classifying moderate ET for both the classifiers and for each of the six tasks is much less than for controls and severe ET. The relatively lower classification accuracy of moderate ET may be because of partial overlap in the feature spaces with the control groups.

D. LIMITATIONS OF THE STUDY AND FUTURE DIRECTIONS

The proposed approach has shown that computerized analysis can be used to distinguish ET from controls and moderate from severe ET. One limitation is that this study did not investigate the difference between ET and other disorders with tremor symptoms. We have not compared PD and ET tremor. While this is important and has been attempted by other authors, but it was not the focus of this study. The FTM scoring was only performed by blinded movement disorders specialists, and the small *post hoc* examination of static spirals by two non-specialized physicians (Table 7) has been included to provide a general benchmark for classification by visual inspection. This minor part of this study should not be used to infer a level of performance for specialized neurologists. Another shortcoming of this study is the size of the dataset of 39 (19 + 20) which, while comparable with other studies, is not sufficient to investigate within the subgroups. Thus, there is a need for a larger dataset. It is also important to test patients over multiple visits to study the repeatability of these experiments.

V. CONCLUSION

In this work, we have developed a computerized system for detecting ET and for monitoring its level of severity. The novelties of this work are threefold. Firstly, several writing and sketching tasks have been compared and those that best differentiate ET from control, and classify the severity of the disease, have been identified. Secondly, a wide range of features reported in the literature have been investigated and a novel set of composite features to best discriminative ET from control as well level of severity of ET have been developed. Finally, a hierarchical multi-layer classifier has been incorporated, which outperformed the single layer support vector machine. This work has the potential of being used to detect and monitor ET symptoms, which can support clinicians and be used for measuring improvement from medication or other treatments.

REFERENCES

[1] E. D. Louis, "Essential tremor," *New England J. Med.*, vol. 345, no. 12, pp. 887–891, 2001.
[2] P. Crawford, III, and E. E. Zimmerman, "Differentiation and diagnosis of tremor," *Amer. Family Physician*, vol. 83, no. 6, pp. 697–702, 2011.
[3] K. P. Bhatia, P. Bain, N. Bajaj, R. J. Elble, M. Hallett, E. D. Louis, J. Raethjen, M. Stamelou, C. M. Testa, and G. Deuschl, "Consensus statement on the classification of tremors. From the task force on tremor of the international parkinson and movement disorder society," *Movement Disorders*, vol. 33, no. 1, pp. 75–87, Jan. 2018.
[4] J. Alty, J. Cosgrove, D. Thorpe, and P. Kempster, "How to use pen and paper tasks to aid tremor diagnosis in the clinic," *Practical Neurol.*, vol. 17, no. 6, pp. 456–463, Dec. 2017.

[5] J. Solé-Casals, I. Anchustegui-Echearte, P. Marti-Puig, P. M. Calvo, A. Bergareche, J. I. Sánchez-Méndez, and K. Lopez-de-Ipina, "Discrete cosine transform for the analysis of essential tremor," *Frontiers Physiol.*, vol. 9, p. 1947, Jan. 2019.

[6] K. Lopez-de-Ipiña, A. Bergareche, P. de la Riva, M. Faundez-Zanuy, P. M. Calvo, J. Roure, and E. Sesa-Nogueras, "Automatic non-linear analysis of non-invasive writing signals, applied to essential tremor," *J. Appl. Log.*, vol. 16, pp. 50–59, Jul. 2016.

[7] K. Lopez-de-Ipina, J. Solé-Casals, M. Faúndez-Zanuy, P. Calvo, E. Sesa, J. Roure, U. Martinez-de-Lizarduy, B. Beitia, E. Fernández, J. Iradi, J. Garcia-Melero, and A. Bergareche, "Automatic analysis of Archimedes' spiral for characterization of genetic essential tremor based on Shannon's entropy and fractal dimension," *Entropy*, vol. 20, no. 7, p. 531, Jul. 2018.

[8] E. D. Louis, A. Gillman, S. Boschung, C. W. Hess, Q. Yu, and S. L. Pullman, "High width variability during spiral drawing: Further evidence of cerebellar dysfunction in essential tremor," *Cerebellum*, vol. 11, no. 4, pp. 872–879, Dec. 2012.

[9] K. López-de-Ipiña, J. Solé-Casals, M. Faundez-Zanuy, P. Calvo, E. Sesa, U. Martinez de Lizarduy, P. De La Riva, J. Marti-Masso, B. Beitia, and A. Bergareche, "Selection of entropy based features for automatic analysis of essential tremor," *Entropy*, vol. 18, no. 5, p. 184, May 2016.

[10] S. L. Pullman, "Spiral analysis: A new technique for measuring tremor with a digitizing tablet," *Movement Disorders*, vol. 13, no. S3, pp. 85–89, Oct. 2008.

[11] K. Lopez-de-Ipiña, A. Bergareche, P. de la Riva, M. Faundez-Zanuy, P. Calvo, and F. Zelarain, "Analysis of non-invasive writing signals, applied to essential tremor: A nonlinear approach," in *Proc. 3rd IEEE Int. Work-Conf. Bioinspired Intell.*, Jul. 2014, pp. 56–59.

[12] E. D. Louis, Q. Yu, A. G. Floyd, C. Moskowitz, and S. L. Pullman, "Axis is a feature of handwritten spirals in essential tremor," *Movement Disorders*, vol. 21, no. 8, pp. 1294–1295, 2006.

[13] S. Barrantes, A. J. Sánchez Egea, H. A. González Rojas, M. J. Martí, Y. Compta, F. Valldeoriola, E. Simo Mezquita, E. Tolosa, and J. Valls-Solè, "Differential diagnosis between Parkinson's disease and essential tremor using the smartphone's accelerometer," *PLoS ONE*, vol. 12, no. 8, Aug. 2017, Art. no. e0183843.

[14] A. Hossen, M. Muthuraman, Z. Al-Hakim, J. Raethjen, G. Deuschl, and U. Heute, "Discrimination of parkinsonian tremor from essential tremor using statistical signal characterization of the spectrum of accelerometer signal," *Bio-Med. Mater. Eng.*, vol. 23, no. 6, pp. 513–531, 2013.

[15] A. Hossen, M. Muthuraman, J. Raethjen, G. Deuschl, and U. Heute, "Discrimination of parkinsonian tremor from essential tremor by implementation of a wavelet-based soft-decision technique on EMG and accelerometer signals," *Biomed. Signal Process. Control*, vol. 5, no. 3, pp. 181–188, Jul. 2010.

[16] X. Zheng, A. Vieira Campos, J. Ordieres-Meré, J. Balseiro, S. Labrador Marcos, and Y. Aladro, "Continuous monitoring of essential tremor using a portable system based on smartwatch," *Frontiers Neurol.*, vol. 8, p. 96, Mar. 2017.

[17] N. Mahadevan, C. Demanuele, H. Zhang, D. Volfson, B. Ho, M. K. Erb, and S. Patel, "Development of digital biomarkers for resting tremor and bradykinesia using a wrist-worn wearable device," *NPJ Digit. Med.*, vol. 3, no. 1, pp. 1–12, Dec. 2020.

[18] M. Michalec, N. Hernandez, L. N. Clark, and E. D. Louis, "The spiral axis as a clinical tool to distinguish essential tremor from dystonia cases," *Parkinsonism Rel. Disorders*, vol. 20, no. 5, pp. 541–544, May 2014.

[19] C. W. Hess, A. W. Hsu, Q. Yu, R. Ortega, and S. L. Pullman, "Increased variability in spiral drawing in patients with functional (psychogenic) tremor," *Human Movement Sci.*, vol. 38, pp. 15–22, Dec. 2014.

[20] R. J. Elble, R. Sinha, and C. Higgins, "Quantification of tremor with a digitizing tablet," *J. Neurosci. Methods*, vol. 32, no. 3, pp. 193–198, Jun. 1990.

[21] J. A. Sisti, B. Christophe, A. R. Seville, A. L. A. Garton, V. P. Gupta, A. J. Bandin, Q. Yu, and S. L. Pullman, "Computerized spiral analysis using the iPad," *J. Neurosci. Methods*, vol. 275, pp. 50–54, Jan. 2017.

[22] D. Haubenberger, D. Kalowitz, F. B. Nahab, C. Toro, D. Ippolito, D. A. Luckenbaugh, L. Wittevrongel, and M. Hallett, "Validation of digital spiral analysis as outcome parameter for clinical trials in essential tremor," *Movement Disorders*, vol. 26, no. 11, pp. 2073–2080, Sep. 2011.

[23] A. P. Legrand, I. Rivals, A. Richard, E. Apartis, E. Roze, M. Vidailhet, S. Meunier, and E. Hainque, "New insight in spiral drawing analysis methods–application to action tremor quantification," *Clin. Neurophysiol.*, vol. 128, no. 10, pp. 1823–1834, Oct. 2017.

[24] S. Fahn, E. Tolosa, and C. Marín, "Clinical rating scale for tremor," *Parkinson's Disease Movement Disorders*, vol. 2, pp. 271–280, 1993.

[25] W. Cobbah and M. C. Fairhurst, "Computer analysis of handwriting dynamics during dopamimetic tests in Parkinson's disease," in *Proc. 26th Euromicro Conf. EUROMICRO Inform., Inventing Future*, Sep. 2000, pp. 414–418.

[26] P. Zham, "Classification of handwriting kinematics in automated diagnosis and monitoring of Parkinson's disease," Ph.D. dissertation, RMIT Univ., Melbourne, VIC, Australia, 2019.

[27] P. Zham, S. P. Arjunan, S. Raghav, and D. K. Kumar, "Efficacy of guided spiral drawing in the classification of Parkinson's disease," *IEEE J. Biomed. Health Informat.*, vol. 22, no. 5, pp. 1648–1652, Sep. 2018.

[28] B. Mwangi, T. S. Tian, and J. C. Soares, "A review of feature reduction techniques in neuroimaging," *Neuroinformatics*, vol. 12, no. 2, pp. 229–244, Apr. 2014.

[29] N. M. Razali and Y. B. Wah, "Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests," *J. Stat. Model. Analyt.*, vol. 2, no. 1, pp. 21–33, 2011.

[30] W. H. Kruskal and W. A. Wallis, "Use of ranks in one-criterion variance analysis," *J. Amer. Stat. Assoc.*, vol. 47, no. 260, pp. 583–621, Dec. 1952.

[31] S. Datta, C. K. Karmakar, A. S. Rao, B. Yan, and M. Palaniswami, "Automated scoring of hemiparesis in acute stroke from measures of upper limb co-ordination using wearable accelerometry," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 4, pp. 805–816, Apr. 2020.

[32] E. Goubault, H. P. Nguyen, F. S. Ayachi, S. Bogard, and C. Duval, "Do bradykinesia and tremor interfere in voluntary movement of essential tremor patients? Preliminary findings," *Tremor Other Hyperkinetic Movements*, vol. 7, p. 459, Jun. 2017.

MOHAMMOD ABDUL MOTIN (Member, IEEE) received the B.Sc.Eng. and M.Sc.Eng. degrees in electrical and electronic engineering from the Rajshahi University of Engineering and Technology, Rajshahi, Bangladesh, in 2011 and 2014, respectively, and the Ph.D. degree from the Department of Electrical and Electronic Engineering, University of Melbourne, Parkville, VIC, Australia, in 2020. He is currently working with RMIT University, Australia. His research interests include biomedical signal processing and modeling, time series analysis, machine learning, embedded systems, monitoring neurological disorders, and sensor devices for wearable health care monitoring.

JAMES PETERS is currently with the Department of Neurosciences, Monash Medical Centre, Melbourne, VIC, Australia.

LAURA PERJU-DUMBRAVA is currently with the Department of Neurosciences, Monash Medical Centre, Melbourne, VIC, Australia.

CATHERINE DING is currently with the Department of Neurosciences, Monash Medical Centre, Melbourne, VIC, Australia.

MICHAEL ELLER is currently with the Department of Neurosciences, Monash Medical Centre, Melbourne, VIC, Australia.

SANJAY RAGHAV is currently an Associate Professor with RMIT University and a Senior Lecturer with Monash University. He is also a Senior Consultant Neurologist with the Monash Medical Centre and Frankston Hospital. He is an assigned Supervisor and a Mentor to RACP trainees. He has a special interest in movement disorders which include tremors, Parkinson's disease, tics, dystonia, and Huntington's disease. In addition, he is involved in several research projects related to movement disorders at RMIT University and Monash University. Several of his research papers and articles have been published in esteemed peer-reviewed medical journals and presented at National and International conferences. He is actively involved in teaching and training of Neurology Registrar and Fellows.

SHEIK MOHAMMED ALI (Graduate Student Member, IEEE) received the M.S. degree in electrical engineering from the Indian Institute of Technology (IIT) Madras, Chennai, India, in 2013. He is currently pursuing the Ph.D. degree with the Department of Electrical and Electronic Engineering, RMIT University, Melbourne, Australia. From 2013 to 2018, he has worked in academia and industry in different roles. His research interests include biomedical signal processing, healthcare and wearable technologies, biomedical devices and systems, sensors in intelligent vehicle, and intelligent transport.

PETER KEMPSTER is currently a Neurologist with the Monash Medical Centre, Clayton, VIC, Australia. His main research interests include movement disorders have been the neuropharmacology and natural history of Parkinson's disease.

PJ RADCLIFFE is an academic with RMIT University, Melbourne, Australia. He has authored more than 100 articles in peer reviewed journals, conferences, and book chapters.

DINESH KANT KUMAR (Senior Member, IEEE) received the B.E. degree from the Indian Institute of Technology (IIT) Madras, India, and the Ph.D. degree from IIT, Delhi, India, in 1990. He has been a Professor with RMIT University, Melbourne, since 1997, where he leads research in bio-signals, and his passion is to develop devices for affordable diagnostics. He has experience in two start-ups. He has authored more than 420 articles and seven books. He has been a Senior Professorial Fellow in Brazil, and a VAJRA Fellow in India. He is on the TGA (Australia) advisory panel for medical devices.