



Contents lists available at ScienceDirect

## Artificial Intelligence in Medicine

journal homepage: [www.elsevier.com/locate/aiim](http://www.elsevier.com/locate/aiim)



# A computer vision framework for finger-tapping evaluation in Parkinson's disease

Taha Khan<sup>a,b,\*</sup>, Dag Nyholm<sup>c</sup>, Jerker Westin<sup>a</sup>, Mark Dougherty<sup>a</sup>

<sup>a</sup> Computer Engineering, School of Technology and Business Studies, Dalarna University, Falun 79188, Sweden

<sup>b</sup> School of Innovation, Design and Technology, Mälardalen University, Västerås 72123, Sweden

<sup>c</sup> Department of Neuroscience, Neurology, Uppsala University, Uppsala 75185, Sweden

### ARTICLE INFO

#### Article history:

Received 10 January 2013

Received in revised form 15 October 2013

Accepted 15 November 2013

#### Keywords:

Computer vision  
Motion analysis  
Face detection  
Parkinson's disease  
Finger-tapping

### ABSTRACT

**Objectives:** The rapid finger-tapping test (RFT) is an important method for clinical evaluation of movement disorders, including Parkinson's disease (PD). In clinical practice, the naked-eye evaluation of RFT results in a coarse judgment of symptom scores. We introduce a novel computer-vision (CV) method for quantification of tapping symptoms through motion analysis of index-fingers. The method is unique as it utilizes facial features to calibrate tapping amplitude for normalization of distance variation between the camera and subject.

**Methods:** The study involved 387 video footages of RFT recorded from 13 patients diagnosed with advanced PD. Tapping performance in these videos was rated by two clinicians between the symptom severity levels ('0: normal' to '3: severe') using the unified Parkinson's disease rating scale motor examination of finger-tapping (UPDRS-FT). Another set of recordings in this study consisted of 84 videos of RFT recorded from 6 healthy controls. These videos were processed by a CV algorithm that tracks the index-finger motion between the video-frames to produce a tapping time-series. Different features were computed from this time series to estimate speed, amplitude, rhythm and fatigue in tapping. The features were trained in a support vector machine (1) to categorize the patient group between UPDRS-FT symptom severity levels, and (2) to discriminate between PD patients and healthy controls.

**Results:** A new representative feature of tapping rhythm, 'cross-correlation between the normalized peaks' showed strong Guttman correlation ( $\mu_2 = -0.80$ ) with the clinical ratings. The classification of tapping features using the support vector machine classifier and 10-fold cross validation categorized the patient samples between UPDRS-FT levels with an accuracy of 88%. The same classification scheme discriminated between RFT samples of healthy controls and PD patients with an accuracy of 95%.

**Conclusion:** The work supports the feasibility of the approach, which is presumed suitable for PD monitoring in the home environment. The system offers advantages over other technologies (e.g. magnetic sensors, accelerometers, etc.) previously developed for objective assessment of tapping symptoms.

© 2013 Elsevier B.V. All rights reserved.

## 1. Introduction

Measurement of an individual's ability to tap fingers is an important method of assessing neuromuscular integrity [1]. In finger-tapping tests, subjects are asked to tap their fingers consistently in a rapid succession. The varying amplitude in tapping and the series of inter-tap intervals are variables of interest for symptom assessment. One way to produce these variables is through the 'rapid finger-tapping test' (RFT) in which subjects attempt to reproduce a sequence of stimuli by pinching their index-finger and

thumb together. Stable RFT is unaffected by cultural experience or cognitive and perceptual demands [1] and clinical impairments are readily apparent in tapping signal patterns. For assessment of Parkinson's disease (PD) in particular, RFT is widely applied in clinical settings as the rhythm of the dominant hand finger movements acts as an efficient index for evaluation of brain motor function [2]. Volkow et al. found a strong correlation between the dopamine receptors in PD and the motor task characterized by RFT [3]. Typically, the RFT is visually evaluated in clinical practice which results in coarse resolution to determine the PD status.

PD is characterized by the loss of dopaminergic neurons which causes slowed muscular movement and motor rigidity. PD is a progressive and incurable disease and results in extensive utilization of health and community services. The cost of PD to society is high; the annual cost in the UK is estimated between 449 million to 3.3 billion

\* Corresponding author at: Computer Engineering, School of Technology and Business Studies, Dalarna University, Falun 79188, Sweden. Tel.: +46 760831014.  
E-mail addresses: [tkh@du.se](mailto:tkh@du.se), [tahak@rocketmail.com](mailto:tahak@rocketmail.com) (T. Khan).

pounds [4]. The cost per patient per year in USA is around 10,000 USD where the largest proportion comes from inpatient care while a lower share comes from medication. The number of PD patients suffering from motor dysfunction is increasing in an aging society. Reliable and objective clinical assessments are therefore important as patients need to be followed up frequently with frequent treatment adjustments for many years. An accurate and cost-effective control of PD symptoms may be possible if the desired treatment effects are evaluated quantitatively through computerized methods.

An electrocardiographic apparatus has been used to examine tapping velocity [5,6] and tapping amplitude has been estimated from 3D images captured from infrared cameras [7]. The system could record 3D motion of fingers but required a large and costly laboratory setup. In recent years, lightweight magnetic sensors and accelerometers [8–11] were utilized for movement analysis, however with a drawback that only a basic analysis could be performed which was irrelevant to the routine assessment of PD in clinical environment. These systems computed tapping velocity and acceleration which were then correlated with the disease status to verify a relationship between the quantities of these features and PD. Besides, some other important features for identification of tapping symptoms in PD, such as tapping rhythm and fatigue, were not investigated. Moreover, the equipments used in these systems were invasive as well as expensive and required supervision for monitoring the tapping symptoms. It has also been discovered that accelerometer readings are affected by gravitational artifacts [12]. All these equipments required appropriate environmental settings to conduct diagnosis.

At-home treatment systems provide an economical solution to track disease progression as well as support patients having physical limitations to access medical facilities. An example is 'Objective PD Measurement System' [13] comprising of dexterity and mobility measurement devices which communicate to a central server via internet. The PD related motor impairment symptoms can be recorded in the server and can be accessed by clinicians and researchers through a web interface. In this way the patient's symptoms can be objectively rated. The investigator can quickly assess treatment effect by grouping together different tests. A web-based video analysis of RFT can be particularly useful when incorporated in systems where tapping videos can be stored in a centralized server for computerized processing.

A recent computer-vision (CV) based approach is 'Virtual Touchpad Interface' [14]. It includes a mobile interface that instructs patients to perform the test. A webcam is attached to the system to record the test which can be evaluated in real time. The assessments are made using contour algorithms to determine if the index-finger and thumb are touching or are apart. However this system had similar limitations to other sensor-based systems [8–11], i.e. it could only compute tapping speed and velocity, and it did not take into account other important clinical features (such as amplitude and rhythm) for symptom level discrimination. Further the algorithm was unable to solve problems in dyskinesias where the patient may not control the body movements and may not keep his hands at one location consistently, i.e. over the testing pad in this case.

We propose a new and unique CV framework for objective assessment of RFT based on quantitative motion analysis of index-fingers combined with the face detection method. We claim that tapping characterization is possible through quantification of rapid opposition of index-finger against the thumb through video analysis. It is a low-cost approach as it does not require any specialized hardware and can be performed in a more natural setup. An added advantage is that the subject's face in tapping video may be localized and blurred to avoid ethical issues in publishing and data sharing.

## 2. Methods and materials

### 2.1. Patients and data

The data have been acquired from a multi-center clinical study called 'DIREQT' (Duodopa Infusion: Randomized Efficacy and Quality of life Trial) [15] conducted during 2002–2003 at 5 different medical facilities in Sweden. The study involved video estimates of film footage in which the patients performed standardized movements such as RFT, alternating finger movement test, oscillatory hand movement test, sitting and arising from a chair and frontal gait. The raters scored the motor functions utilizing the unified Parkinson's disease rating scale (UPDRS) [16]. The motor examination of finger-tapping is part III in UPDRS abbreviated as UPDRS-FT. The scores in UPDRS-FT are ranged between 0 and 4 points where '0' represents normal tapping, '1' represents mildly impaired, '2' represents moderately impaired, '3' represents severely impaired, and '4' represents inability to perform tapping.

The patient data in the present study consisted of RFT recordings from 13 patients (5 females and 8 males) suffering from advanced PD. The mean total UPDRS score of these patients was 50.45 (range 14–92; SD  $\pm$  18.12) on the UPDRS scale 0 (normal state) to 108 (total motor impairment). The patients were aged between 50 and 75. The recordings took place at different timings of the day, i.e. from 9 am till 5 pm with a rest of 30 min in between. A total of 17 recordings took place each day for each patient. Six patients were recorded for 2 days while two others were videotaped for 3 days with a gap of a week in between the recording days. Five patients could show up for a single day for recordings. The tapping videos were rated by two clinicians based upon the tapping performance of each patient using the UPDRS-FT.

During the video recordings, the patients were seated on a chair against a plain wall behind. The videos were recorded for each patient facing toward a pivoted camera. During RFT, the patients positioned their hands above the shoulders besides the face. They were asked to stretch the index-finger vertically against the thumb as much as possible. The tapping had to be done as fast as possible. The visual features of interest for the two clinicians were the tapping rate, amplitude between index-finger and thumb, hesitations, halts, and decrement in the amplitude. The raters classified tapping symptoms in 4 symptom levels in UPDRS-FT (i.e. '0', '1', '2' and '3'). Each patient was videotaped for 10 s at a frame rate of 25 fps and a frame resolution of  $352 \times 288$  pixels.

In addition to the patient group, RFT from 6 healthy controls (HC), aged between 40 and 60 years, were recorded with the same video configuration. A total of 84 recordings were made over a span of two weeks such that each individual was recorded once every day in a week. The recorded videos from the two groups (patient and HC) were fed to the CV-based system to produce and analyze the tapping signals. A total of 471 videos, 387 patients (i.e. 23 days  $\times$  17 videos/day = 391 videos; 4 bad quality videos were removed) and 84 HC (i.e. 7 days  $\times$  2 video/day  $\times$  6 subjects = 84 videos) were utilized for tapping classification. The investigation was approved by the local ethics committee and written informed consent was obtained from the HC and the patient group.

### 2.2. System description

The block diagram of the proposed CV based algorithm for RFT quantification is shown in Fig. 1. In the first step the face of the subject is detected. The video-frame is split into two images from the center coordinates of a face-rectangle. Two regions of interest (ROI) are located in each of the images representing areas of index-finger movements of left and right hands, respectively. In the next step, motion of the index-finger is estimated in each ROI. The movement of index-finger tip coordinates produces a tapping

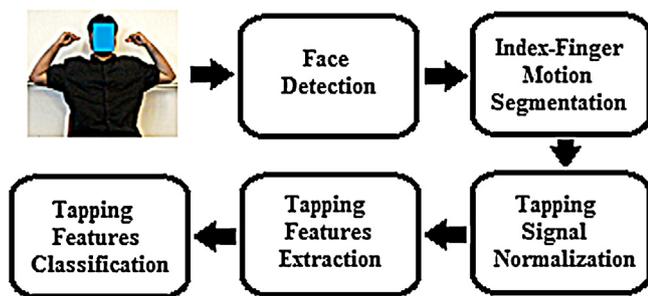


Fig. 1. System diagram of video-based finger-tapping quantification.

time-series which is normalized by face-rectangle height. The tapping features are extracted and classified using a support vector machine (SVM) classifier to characterize between the symptom severity levels based on the clinical scores.

### 2.3. Face detection

Face detection for tapping analysis is a novel approach motivated from the observation that the length of an adult person's hand is approximately equal to the height of that person's face (Fig. 2a.i and a.ii)[17]. In previous video based methods for tapping analysis, the tapping signal amplitude was affected by distance of the camera from the subject's hand which could not be corrected [18]. This amplitude calibration problem led to problematic estimation of tapping features for symptom characterization. There are several advantages of human face incorporation in tapping analysis where the frontal face position is readily apparent in the videos and both hands are positioned besides the face. First, the tapping signal can be normalized by face height to cope with varying camera positions during recordings. The nearer the camera is to the subject, the larger the finger-tapping amplitude and face height. The opposite is true if the camera is placed at a farther distance. The tapping signal normalization would calibrate the tapping amplitude resulting in correct estimation of tapping features. Secondly, since tapping is performed by placing the hands besides the face, the ROI besides the face may be located to track the index-finger movement of both hands in reference to the face-rectangle coordinates (Fig. 2c). Interfering head movements during tapping may also be located and eliminated. Third, the algorithm would be able to blur the face to comply with ethical restrictions when publishing the videos (Fig. 2b). Further, the subjects may be assessed for dystonia symptoms by analyzing their facial expressions (a possible extension to this work).

Face detection in this framework has been performed using a face tracker available in the OpenCV library developed by Turk [19]. The face tracker is based upon the eigenfaces recognition algorithm where for a given face image and a set of face images (called eigenfaces); a Euclidean distance is computed between eigenfeatures in the new image and each of the example faces. The eigenfaces are considered as a set of standard face eigenfeatures derived from principal component analysis of many face pictures. Any human face structure may be considered as a combination of these standard features. A face detector called the 'Haar Cascade Classifier' [20] is used to examine each image location and classify it as 'face' or 'not face'. The classification results in identifying a face-rectangle representing facial features in video frames (Fig. 2b). The OpenCV provides an option to choose between the classifiers for frontal or sideways face detection. In our case the subjects were facing toward the camera so the classifier used was 'Haar Cascade Frontal Face' classifier. Face classification is performed with a reliability of 98% in nearly 96,000 video frames. The 2% failure in face detection is because of the subjects moving their head sideways during

finger-tapping. This minor loss may be ignored by skipping the frames later in the tapping recording.

Once the face-rectangle with top-left coordinates  $x_{\min}$ ,  $y_{\min}$  and bottom right coordinates  $x_{\max}$ ,  $y_{\max}$  is computed, the video frame is divided into two images from the face-rectangle center. Since the hands are positioned over the shoulders besides the face, two rectangles are drawn (twice the size of face-rectangle) besides the face-rectangle to cover hand movements in each image (Fig. 2c). These rectangles represent the two regions of interest (i.e. ROI<sub>left</sub> and ROI<sub>right</sub>) for motion detection of index-fingers in left and right hands, respectively. The ROIs can be adjusted to cover bigger areas depending upon the background steadiness during RFT.

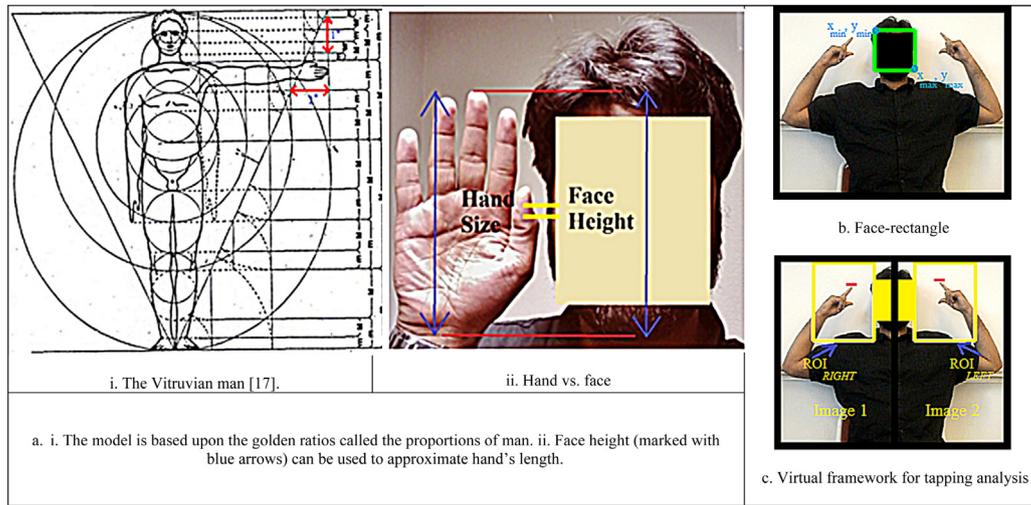
### 2.4. Index-finger motion segmentation

Motion segmentation algorithms generally work by comparing the incoming video frame to a previously referenced frame. Deviations from the referenced frame are analyzed to attribute the difference with the incoming frame. The camera should be stationary so that a static scene model (background) may be built to detect movement regions. The motion analysis of index-fingers in ROI<sub>left</sub> and ROI<sub>right</sub> has been performed using the motion-template gradient algorithm [21]. The motion templates record location of moving pixels in video frames at each time-stamp. A motion history image (MHI) is kept to track the latest movements together with the latest timestamps. An important feature of this method is that motion can be detected in small regions of a frame even in low resolution imagery; this is particularly suited to detect fine finger movements in tapping.

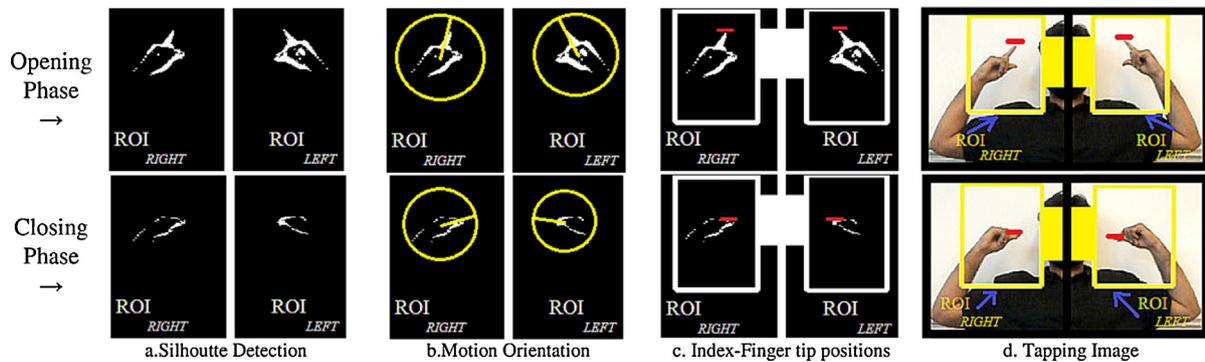
Motion-template gradient is a five step algorithm: (1) silhouette detection, (2) MHI updates, (3) motion gradient calculation, (4) motion orientation calculation and (5) motion segmentation. Silhouettes are the basic components of motion detection. Silhouettes are detected in ROI<sub>left</sub> and ROI<sub>right</sub> separately (Fig. 3a) by computing the absolute pixel difference between an incoming frame and previous 4 frames. An absolute difference greater than 40 pixels is kept as a threshold to identify the motion silhouettes representing index-finger movements. Experiments showed that an absolute difference less than 40 pixels may represent insignificant motion details which may arise due to fine movements of fingers posterior to the index-finger. The motion silhouettes are then binarized and stored in two memory buffers (one for left and the other for the right hand) to track movement of index-fingers in successive video frames. A time stamp (a current system time) is associated with the motion silhouettes to identify a recent motion. The previous frames consisting of a sequence of motion silhouettes are referred to as MHIs based on which the motion templates are constructed by comparing the silhouettes in a recent frame with the MHIs. The buffer size is specified as 5 MHIs. A time limit of 1 s is kept to store previous MHIs after which the oldest MHI is removed with an addition of a recent MHI.

An overall motion is detected by computing the gradients of MHIs in the memory buffer by convolving MHIs with a Sobel filter. The gradients of MHIs point orthogonal to silhouettes' boundaries. The output of the gradient function is a mask which gives the angle of gradients' direction (motion orientation). Overall motion orientation has been computed in ROI<sub>left</sub> and ROI<sub>right</sub>, respectively, by taking the vector-sum of all the gradient orientations inside the mask as shown in Fig. 3b.

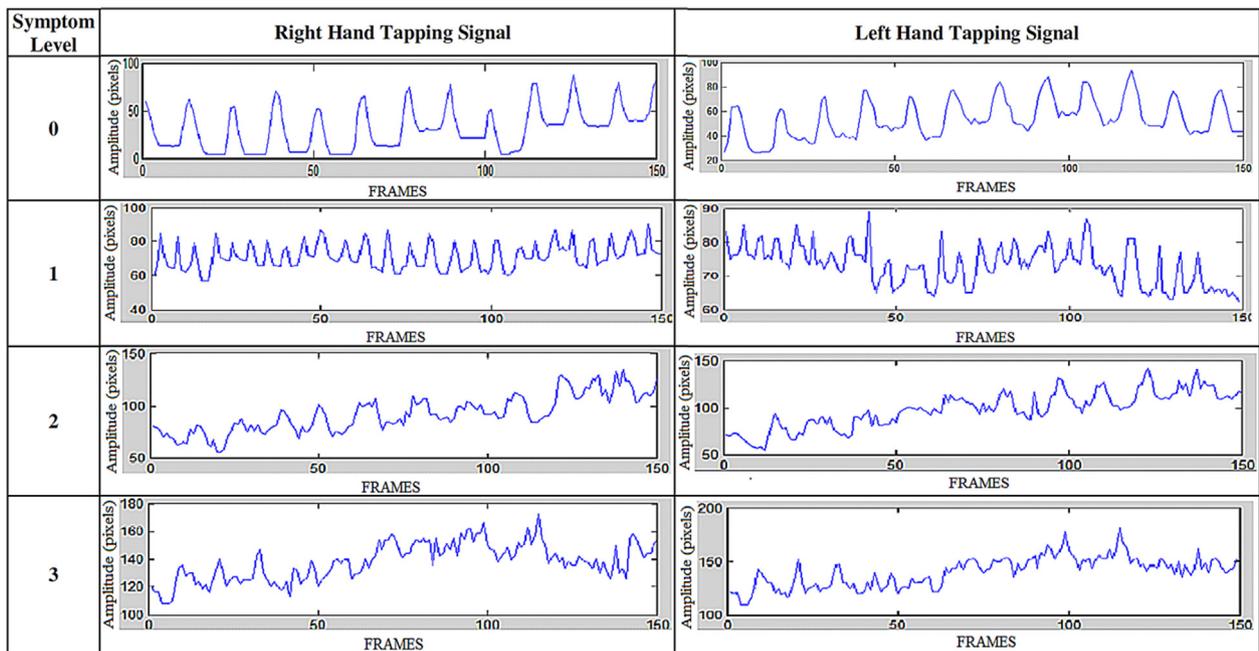
Once motion orientation is computed, the local motion is segmented within ROI<sub>left</sub> and ROI<sub>right</sub>, respectively. MHIs in the buffers are scanned for silhouette regions with a recent timestamp to search for the most recent motion toward the motion orientation. Floodfilling is done to isolate the locally found motion silhouettes in ROI<sub>left</sub> and ROI<sub>right</sub>, respectively. The motion silhouettes are merged as connected components representing the moving pixels of hand.



**Fig. 2.** The golden ratio and the role of face detection in finger-tapping quantification. (a) (i) The model is based upon the golden ratios called the proportions of man. (ii) Face height (marked with arrows) can be used to approximate hand's length. (b) Face-rectangle. (c) Virtual framework for tapping analysis.



**Fig. 3.** Index-fingers motion detection. Motion orientation of index-fingers in  $ROI_{left}$  and  $ROI_{right}$ , respectively, is determined to identify the opening and closing phases of tapping. Tip of index-fingers are marked to locate finger coordinates over time to produce a tapping signal.



**Fig. 4.** The tapping signals produced by the video processing of representative clinically rated samples. The tapping rhythm deteriorates as well as the amplitude decreases with the increase in symptom levels. Non-dominant (left) hand tapping signals deteriorates more than the dominant (right) hand.

The top position of the connected component (Fig. 3c and d) is determined to locate the tip of an index-finger and a position-marker is placed over the tip coordinates. The moving  $y$ -coordinate (amplitude) of this position-marker over time forms the tapping time-series. The signal values are divided by the face-rectangle height to normalize the effect of distance variation between camera and subject for a correct estimation of tapping amplitude. Fig. 4 illustrates tapping signals acquired by the video processing of representative clinically rated samples. It can be observed that the rhythm in these signals deteriorates with the increasing symptom levels. The tapping signals produced by the left hand depict more deterioration. This is because the subjects have less control over their non-dominant (left) hand which resulted in varying amplitude and hesitations in the tapping signal as compared to the dominant (right) hand.

### 2.5. Baseline calibration and the measurement accuracy

The varying hand position in the video frames shifts the baseline of tapping signal from the horizontal axis (Fig. 4). One way to calibrate the baseline is to apply a moving average on the signal values so that the signal lies over the horizontal zero-axis. This can be done by negating each signal value with a moving average of 4 neighboring values (2 neighbors on left and 2 on right) as given in Eq. (1).

$$S_{i-2} = T_i - \frac{T_{i-2} + T_{i-1} + T_{i+1} + T_{i+2}}{4} \quad \text{for } i = 3, \dots, n-2 \quad (1)$$

where  $T_{1,\dots,n}$  and  $S_{1,\dots,n-4}$  are the original and the calibrated tapping signals, respectively, and  $n$  is the signal length. The baseline calibration was particularly useful in a situation where the index-finger movements were affected due to dyskinesias and the patients could not keep control of their hands constantly at one position during finger tapping (Fig. 6II.A). The calibrated signal is then filtered using the 3-standard deviation rule to remove possible outliers (noise) in the signal.

In order to assess the accuracy of this algorithm in measuring the index finger motion, we compared the tapping signals produced using our non-invasive motion-based method, and the tapping signals produced using an invasive marker-based method. To perform this comparison, two red-color passive markers were attached on the tip of the index-finger and thumb on both hands of the HC (Fig. 5a). The RFT was carefully recorded after ensuring that there was no other red-color object placed in the background.

Each frame of the recorded video was transformed from RGB into HSV (Hue-Saturation-Value) color space [22]. The values of hue, saturation and value were normalized in the range 0–255. Red color was segmented in  $ROI_{\text{left}}$  and  $ROI_{\text{right}}$ , respectively using the hue range of '210–255' and '0–10', saturation range of '127–255', and the value range of '127–255'. The segmented pixels were merged as connected components representing the red-color passive markers. The top position of connected component was determined and a position-marker was placed over the pixel coordinates. The moving  $y$ -coordinate (amplitude) of position-marker over time was used to produce the tapping time-series. This time-series was normalized using the face height. The baseline was adjusted using Eq. (1). A right hand tapping signal is shown in Fig. 5b.

The same video was then used to produce another tapping signal using the motion-based method. The signal was normalized and calibrated, and then was compared with the calibrated tapping signal produced using the marker-based method. It can be seen in Fig. 5a that the position-markers for both motion-based and the marker-based methods are overlapping between each other in the closing and opening phases of tapping. Also the tapping signals produced by these methods are similar and are nearly overlapping (Fig. 5b). A minor translation in the motion-based signal can be

noticed which is due to the fact that the motion algorithm has to buffer previous 4 frames to compute motion in the next frame. A peak-finder was utilized to locate peaks and valleys in these signals. The magnitude difference of peaks and valleys between the two signals for both left and right hand was computed.

In a total of 168 tapping signals (right and left hand tapping signals in 84 videos of HC), the magnitude difference was  $3.314 \pm 0.09$  pixels between the peaks and  $2.44 \pm 0.05$  pixels between the valleys within the 95% confidence interval. This proves that a motion-based marker-less system can produce a tapping signal similar to a marker-based system with a very minor difference of accuracy in measuring the index-finger motion. The motion-based algorithm offers several other advantages compared to a marker-based algorithm because the accuracy of color segmentation is highly dependent upon the colors in the background and is more sensitive to illumination changes as compared to the motion based algorithm. Secondly, a motion-based system is non-invasive does not require markers, and does not cause any discomfort to the patients in tapping.

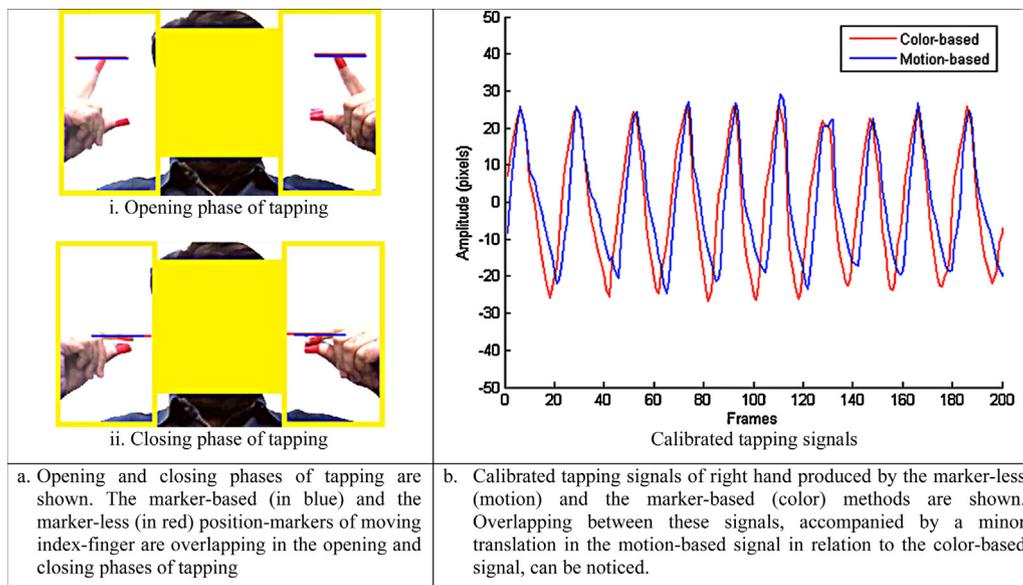
### 2.6. Feature extraction

The biological rhythm in tapping is a representative feature of brain function [23]. The disturbed rapid alternating movements are the sign of brain dysfunction known in medical term as 'arrhythmokinesia'. Shimoyama et al. [2] found that in the normal subjects (1) the tapping frequency is lowered with advancing age, (2) women tap slower than men, and (3) tapping with the dominant hand is faster than the non-dominant hand and is the correct representative of brain function. According to them, the tapping rhythm of the dominant hand is a generalized feature which can distinguish between a person with PD and a normal subject. In PD, the tapping amplitude varies depending upon patient's condition, i.e. it may reduce in bradykinesias (reduced muscular movement) or may increase in dyskinesias (involuntarily muscular movements caused by dopaminergic medication). There is a need of a signal feature which can be used to identify both conditions.

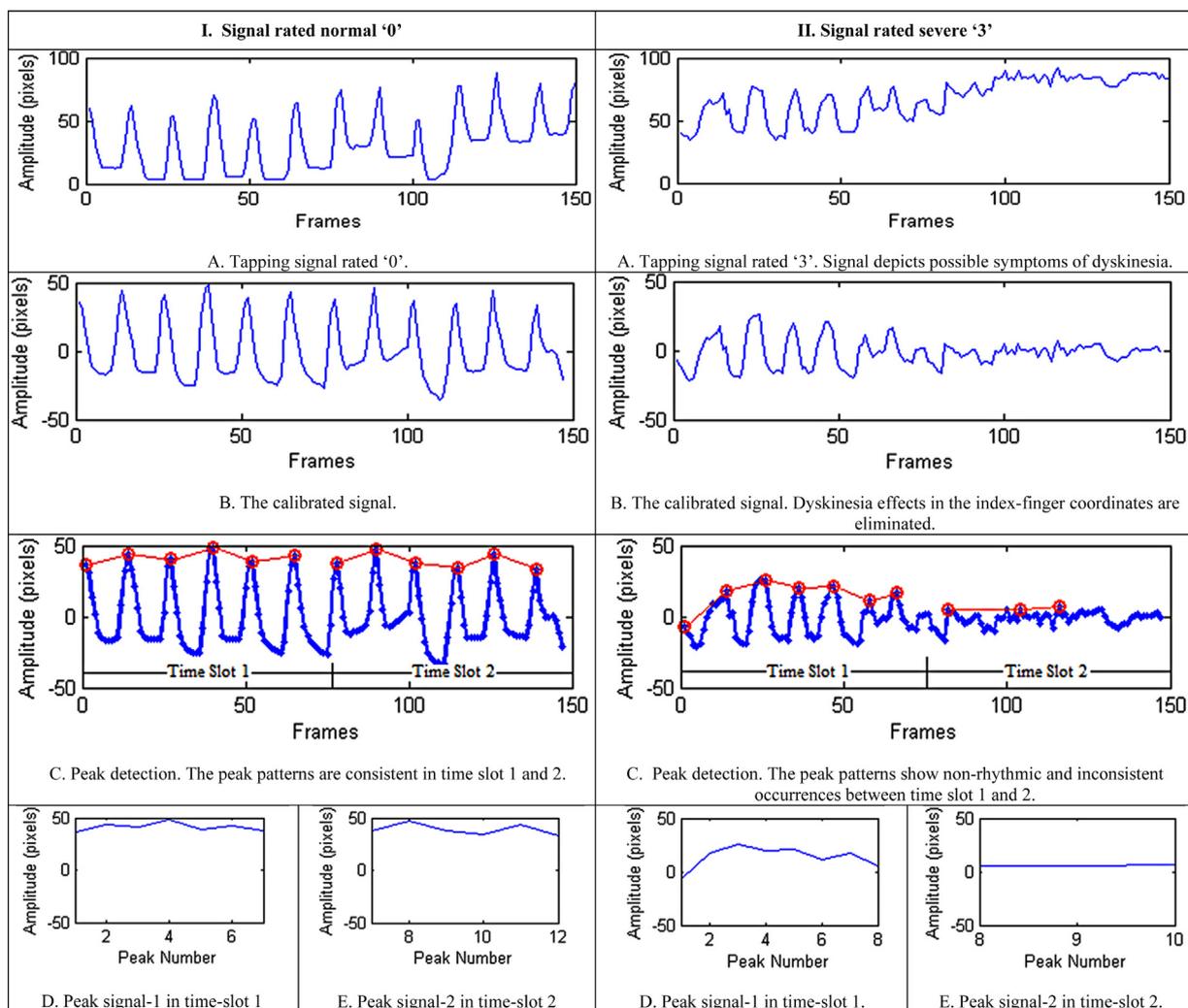
The tapping features have been estimated from the dominant hand tapping signal which was the right hand for all subjects. The signal baseline is calibrated (Section 2.5) before computing these features. A novel feature called the 'cross-correlation between the normalized peaks' (CCNP) is used for estimating the consistency and rhythm in tapping. In this approach, the signal  $S_i$  (Eq. (1)) is split into two time-slots. Time-slot 1 consists of signal values from 2 to 5 s (75 frames) of the tapping video. Time-slot 2 consists of signal values from 6 to 9 s (75 frames). A peak-finder algorithm is applied over the tapping signal to find out peak magnitudes and locations (Fig. 6IC and IIC). A new 'peak-signal' is constructed by joining the peak points within each time slot. In case of 'normal' rated tapping signal, peak-signal-1 and peak-signal-2 exhibited similarity in peak occurrences as well as in the peak magnitudes representing rhythm and tapping consistency over time. This similarity can also be considered a measure of subjects' stamina as the tapping amplitude reduces due to fatigue in later stages of tapping.

By contrast, the peak-signal-2 in an 'impaired' signal (Fig. 6IIE) represents the slot where the subject hesitated for a while. This has resulted in a rather straighter line with lower magnitude as compared to peak-signal-1 (Fig. 6IID). Cross-correlation is applied to measure the similarity between peak-signals 1 and 2 using Eq. (2).

$$CCNP_{P_1 P_2}(m) = \frac{1}{N} \sum_{n=1}^{N-m+1} P_1(n)P_2(n+m-1) \quad \text{for } m = 1, 2, \dots, N \quad (2)$$



**Fig. 5.** A comparison between the marker-based and the marker-less methods of tapping. (a) Opening and closing phases of tapping are shown. The marker-based (in blue) and the marker-less (in red) position-markers of moving index-finger are overlapping in the opening and closing phases of tapping. (b) Calibrated tapping signals of right hand produced by the marker-less (motion) and the marker-based (color) methods are shown. Overlapping between these signals, accompanied by a minor translation in the motion-based signal in relation to the color-based signal, can be noticed. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article).



**Fig. 6.** Cross-correlation between the normalized peaks (CCNP). The signal rated 'normal' shows rhythm and consistency in tapping over time. A comparison between the tapping signals rated 'normal' and 'severely-impaired' shows that the CCNP values are larger in normal tapping.

where  $P_1$  and  $P_2$  are peak-signals 1 and 2 of length  $N$ , respectively, and  $M$  is the number of correlation points. In case if  $P_1$  and  $P_2$  have different lengths, the shorter signal is zero-padded to the length of the longer signal.

The CCNP showed high correlation values between peak-signals 1 and 2 in 'normal' case. By contrast, low correlation between peak-signals 1 and 2 is observed in the 'impaired' case. The average between CCNP values ( $\text{Avg}_{\text{CCNP}}$ ) is computed (Eq. (3)) and used in feature analysis as the representative feature of tapping rhythm in the opening phases of index-finger.

$$\text{Avg}_{\text{CCNP}} = \frac{\sum_{m=1}^M \text{CCNP}_{P_1 P_2}(m)}{M} \quad (3)$$

Similarly, the valley-signals 1 and 2 are derived from the calibrated tapping signal  $S_i$  so that the rhythm in closing phases of index-finger can be analyzed separately. Cross-correlation between the valley-signals 1 and 2 is computed using Eq. (4). The average between CCNV ( $\text{Avg}_{\text{CCNV}}$ ) values is computed using Eq. (5).

$$\text{CCNV}_{V_1 V_2}(m) = \frac{1}{N} \sum_{n=1}^{N-m+1} V_1(n) V_2(n+m-1) \quad \text{for } m = 1, 2, \dots, N \quad (4)$$

$$\text{Avg}_{\text{CCNV}} = \frac{\sum_{m=1}^M \text{CCNV}_{V_1 V_2}(m)}{M} \quad (5)$$

where  $V_1$  and  $V_2$  are valley-signals 1 and 2 of length  $N$ , and  $M$  is the number of correlation points. In case if  $V_1$  and  $V_2$  have different lengths, the shorter signal is zero-padded to the length of the longer signal.

Other tapping features extracted from the calibrated tapping signal  $S_i$  and utilized in symptom classification are listed in Table 1. These features incorporate clinical symptoms visually inspected by clinicians to rate between the UPDRS-FT levels. The 'tapping speed', 'average maximum amplitude of finger taps' and 'variation coefficient (VC) in maximum amplitude' have been investigated previously to detect slowed pace, amplitude reduction and fatigue in tapping, respectively. These features were extracted using magnetic sensors [11] attached with the index-finger and thumb for measurements of tapping movements, whereas here they are computed from the calibrated tapping signal  $S_i$  produced by the CV-based system.

Some features such as total number of taps ( $T_n$ ) are computed by applying a peak-detector over the tapping signal  $S_i$  where the number of peak locations represents the number of taps. The difference between the number of taps in time slots 1 and 2 ( $\Delta T_n$ ) can be used to detect fatigue in the late stage of tapping (Eq. (6)).

$$\Delta T_n = T_{n1} - T_{n2} \quad (6)$$

where  $T_{n1}$  and  $T_{n2}$  are the number of peaks in time slots 1 and 2 of tapping signal, respectively.

A maximum tap amplitude ( $T_A$ ) can be computed by summing up the peak magnitude and the magnitude of valley adjacent left to the peak. The difference between the averages of maximum amplitude of finger taps in time slots 1 and 2 ( $\Delta A$ ) can be used to estimate amplitude reduction caused by fatigue in the late stage of tapping (Eq. (7)).

$$\Delta A = \frac{\sum_{k=1}^{n_1} T_{A1k}}{n_1} - \frac{\sum_{k=1}^{n_2} T_{A2k}}{n_2} \quad (7)$$

where  $T_{A1k}$  and  $T_{A2k}$  are the maximum amplitude of taps in time slots 1 and 2, respectively, and  $n_1$  and  $n_2$  are number of taps in time slots 1 and 2.

The VC in maximum amplitude of finger taps ( $VA$ ) can be used to estimate amplitude reduction (Eq. (8)). Similarly, the difference

between the VCs in maximum amplitude of finger taps in time slots 1 and 2 ( $\Delta VA$ ) can be used to estimate the difference in amplitude variation in the early and the later stages of tapping (Eq. (9)).

$$VA = \frac{\sqrt{(1/n) \sum_{k=1}^n (T_{Ak} - \bar{T}_A)^2}}{\bar{T}_A} \quad (8)$$

$$\Delta VA = VA_1 - VA_2 \quad (9)$$

where  $\bar{T}_A$  is the average between maximum amplitude of taps  $T_{Ak}$  in the tapping signal  $S_i$ , and  $n$  is the total number of taps in that signal.  $VA_1$  and  $VA_2$  are the VCs in maximum amplitude of finger taps in time slots 1 and 2 of tapping signal, respectively.

The interval from one peak location to the adjacent peak location represents a tap interval. Similarly, a tap distance covered by the index-finger can be computed by summing up the absolute of the signal values in a single tap interval. Tapping speed in a tap interval can be computed by dividing the tap distance by the tap interval. The average tapping speed ( $T_s$ ) can be computed using Eq. (10).

$$T_s = \frac{1}{n} \sum_{i=1}^n \frac{T_{Di}}{T_{Ii}} \quad (10)$$

where  $T_I$  is a tap interval,  $T_D$  is the distance covered by the index-finger within that interval, and  $n$  is the total number of tap intervals in the tapping signal  $S_i$ .

De-acceleration in tapping speed is a symptom of fatigue. The VCs in tapping speed ( $VT_s$ ) and tapping acceleration ( $A_c$ ) can be computed to estimate fatigue symptoms (Eqs. (11) and (12)).

$$VT_s = \frac{\sqrt{(1/n) \sum_{i=1}^n (T_{Di}/T_{Ii} - T_s)^2}}{T_s} \quad (11)$$

$$A_c = \frac{1}{n-1} \sum_{i=1}^{n-1} \frac{|T_{vi+1} - T_{vi}|}{T_{Ii+1}} \quad (12)$$

where  $T_v$  is the speed of index finger within a tap interval  $T_I$ .

Other pace-related tapping features such as the opening velocity of index-finger can be computed by summing up the peak and valley magnitudes (adjacent left to the peak) and dividing the summed magnitude by the time interval between peak and valley locations. The average opening velocity ( $O_T$ ) can be computed using Eq. (13).

$$O_T = \frac{1}{n} \sum_{i=1}^n \frac{(p_i + |v_i^*|)}{T_{Oli}} \quad (13)$$

where  $p$  is the peak magnitude,  $v^*$  is the magnitude of valley adjacent left to the peak,  $n$  is the total number of tap intervals, and  $T_{Oli}$  is the time interval between the peak and valley locations. Similarly, the closing velocity can be computed by summing up the valley magnitude with the magnitude of adjacent peak to the left and dividing the summed magnitude by the time interval between their locations. The average closing velocity ( $C_T$ ) can be computed using Eq. (14).

$$C_T = \frac{1}{n} \sum_{i=1}^n \frac{(|v_i| + p_i^*)}{T_{Cii}} \quad (14)$$

where  $v$  is the valley magnitude,  $p^*$  is the magnitude of peak adjacent left to the valley,  $n$  is the total number of tap intervals, and  $T_{Cii}$  is the time interval between the peak and valley locations.

**Table 1**

The tapping features. Features incorporate clinical symptoms (i.e. slow paced, amplitude reduction between index-finger and thumb, fatigue and arrhythmia) visually inspected by clinicians to rate between the UPDRS-FT levels.

Symbol	Feature description	Related symptom (based on UPDRS-FT <sup>a</sup> )
1. $T_n$	Total number of taps	Slowed pace
2. $\Delta T_n$	Difference between number of taps in time slots 1 and 2	Fatigue
3. $T_s$	Average tapping speed	Slowed pace
4. $VT_s$	Variation coefficient (VC) in tapping speed	Fatigue
5. $\Delta A$	Difference between the average maximum amplitude of finger taps in time slots 1 and 2	Fatigue and amplitude reduction
6. $VA$	VC in the maximum amplitude of finger taps	Fatigue and amplitude reduction
7. $\Delta VA$	Difference between the VCs in maximum amplitude of finger taps in time slots 1 and 2	Fatigue and amplitude reduction
8. $O_T$	Average opening velocity of index-finger	Slowed pace
9. $C_T$	Average closing velocity of index-finger	Slowed pace
10. $A_c$	Tapping acceleration	Fatigue
11. $T_z$	Average zero-crossing rate	Slowed pace
12. $T_E$	Signal energy	Slow pace and amplitude reduction
13. $Avg_{CCNP}$	Mean of cross-correlation between the normalized peaks	Arrhythmia
14. $Avg_{CCNV}$	Mean of cross-correlation between the normalized valleys	Arrhythmia
15. $FS$	Standard deviation of face-rectangle centroid during tapping	–

<sup>a</sup> UPDRS-FT [16] (patient taps thumb with index finger in rapid succession). 0, normal; 1, mild slowing (in pace) and/or reduction in amplitude; 2, moderately impaired. Definite and early fatiguing. May have occasional arrests in movement (i.e. mild arrhythmia). 3, severely impaired. Frequent hesitation in initiating movements or arrests in ongoing movement (i.e. severe arrhythmia).

Further assessments are made on spectral variability of tapping signal using the signal energy ( $T_E$ ) and average zero-crossing rate ( $T_z$ ) given in Eqs. (15) and (16), respectively.

$$T_E = \frac{\sum_{i=1}^N S_i^2}{N} \quad (15)$$

where  $S_1, \dots, S_N$  is the calibrated tapping signal of length  $N$ .

$$T_z = \sum_{m=1}^M Z(m)/M; \quad Z(m) = \frac{f_s}{2N} \left( \sum_{i=1}^n |\text{sign}(S_m(i)) - \text{sign}(S_m(i-1))| \right) \quad (16)$$

where  $M$  is the total number of signal blocks,  $f_s$  is the sampling rate (i.e. 25 fps),  $n$  is the window size (kept as  $n=5$ ),  $S_m$  is the windowed part of the signal  $S_1, \dots, S_N$ , and  $Z(m)$  is the zero-crossing rate in the windowed signal  $S_m$  of block  $m$ . For a signal length of 150 frames, where  $n$  is kept as 5, the number of signal blocks  $M$  is  $150/5=30$ .

An important consideration in feature extraction is the movement of the head during RFT which is motivated from the fact that head mobilization correlates well with the dominant hand index-finger tapping [24]. In a previous study, it was revealed that head drifts up to 1.6 mm during tapping in normal controls. Since face-rectangle is readily computed during face detection step, variations in head position are estimated by computing the Euclidean distance  $f_d$  between the face-rectangle center in the first video frame and the face-rectangle center in the other video frames. The standard deviation between  $f_d$  ( $FS$ ) in the total video frames  $N$  (Eq. (17)) is computed and used in the feature analysis.

$$FS = \sqrt{\frac{1}{N-1} \sum_{d=1}^{N-1} (f_{d+1} - f_d)^2} \quad (17)$$

### 2.7. Feature selection and analysis

The UPDRS based finger-tapping assessments are subjective and the raters may differ in opinion when identifying the symptom level. For instance, rater-1 identified 123 videos as level '0', 163 videos as '1', 78 videos as '2' and 23 videos as '3', whereas rater-2 identified 77 videos as '0', 184 videos as '1', 101 videos as '2' and 25 videos as '3'. The inter-rater agreement between the two raters is analyzed using Spearman pair-wise correlation and is computed as  $\rho = 0.74$ .

An appropriate feature validation approach would be to correlate the tapping features with the target ratings provided by the two raters. The choice of correlation model however is complicated because a human rater has to follow a range of tapping symptoms in discriminating between the severity levels. Sometimes two RFT samples are rated similar by a clinician yet have a different type of anomaly. For example, strong symptoms of fatigue in one sample may exist weakly in another sample rated similar but which has severe symptoms of arrhythmia. Another important consideration for choosing an appropriate correlation model is that the target (UPDRS-FT) classes exhibit a monotonic rank-order, i.e. '0' (normal) to '3' ('severe impairment'). The situation demands a highly structured and hierarchical correlation model for feature analysis.

The Guttman correlation model [25] fits the ranked nature of the tapping dataset where an expert examines a range of different symptoms to rate between the severity levels. The Guttman scale is appropriate for non-parametric analysis of a clinically rated dataset as it can provide a hierarchical rank-order structure for a set of symptoms representing a single severity level. In the Guttman scale, a variable  $y$  (i.e. a human rater) with  $m$  distinct ordered values (i.e. UPDRS-FT classes) is said to be a function of variable  $x$  (a computed tapping feature) with  $n$  distinct ordered values, if for each value of  $x$  there is one and only one target value of  $y$ . The converse needs not to hold, i.e. for the same target value of  $y$ , there may be two or more values of  $x$ . Moreover, on a Guttman scale, an agreement to one target class implies an agreement with the lower-order target classes. The clinical ratings from the two raters have been used as the target variables in Guttman correlation analysis. The Guttman monotonicity coefficient ( $\mu_2$ ) is utilized to map the tapping features with target variables. The formula for computing  $\mu_2$  is given in Eq. (18). Jackknifing has been used as a cross-validation method to estimate the precision of  $\mu_2$ . The jackknifing estimates of  $\mu_2$  are listed in Table 2.

$$\mu_2 = \frac{\sum_{h=1}^n \sum_{i=1}^n (x_h - x_i)(y_h - y_i)}{\sum_{h=1}^n \sum_{i=1}^n |x_h - x_i| |y_h - y_i|} \quad (18)$$

Nearly all the tapping features showed strong monotonic correlations (or anti-correlations) with the clinical ratings provided by rater-1 (Table 2). The features representing arrhythmia and fatigue symptoms (i.e.  $Avg_{CCNP}$ ,  $VT_s$ ,  $\Delta VA$  and  $VA$ ) were very strongly anti-correlated ( $\mu_2 < -0.8$ ) and they decreased monotonically with increasing symptom severity. These correlations were statistically significant ( $p < 0.0001$ ). The features representing pace and amplitude reduction showed comparatively moderate correlations (or

**Table 2**

Jackknifing estimates of Guttman monotonicity coefficient ( $\mu_2$ ). Estimates in bold represent strong absolute correlation ( $\mu_2 > 0.5$  or  $\mu_2 < -0.5$ ) between the clinical ratings and the respective tapping feature ( $p < 0.0001$ ).

	$T_n$	$\Delta T_n$	$T_s$	$VT_s$	$\Delta A$	$\Delta VA$	$VA$	$O_T$	$C_T$	$A_c$	$T_Z$	$T_E$	$AVG_{CCNP}$	$AVG_{CCNV}$	$FS$
$T_n$	1														
$\Delta T_n$	0.70	1													
$T_s$	0.29	0.05	1												
$VT_s$	-0.39	-0.49	0.83	1											
$\Delta A$	-0.69	-0.54	0.92	0.87	1										
$\Delta VA$	-0.62	-0.59	0.77	<b>1</b>	0.91	1									
$VA$	-0.38	-0.47	0.82	<b>1</b>	0.87	<b>1</b>	1								
$O_T$	0.01	-0.26	<b>1</b>	0.83	0.96	0.81	0.83	1							
$C_T$	-0.01	-0.28	<b>1</b>	0.84	0.95	0.82	0.84	<b>1</b>	1						
$A_c$	0.61	0.05	0.93	0.57	0.60	0.45	0.57	0.91	0.96	1					
$T_Z$	0.91	0.71	0.05	-0.72	-0.69	-0.83	-0.71	-0.11	-0.11	0.49	1				
$T_E$	-0.11	-0.39	-0.06	-0.02	-0.04	-0.01	-0.02	-0.02	-0.01	0.01	-0.42	1			
$AVG_{CCNP}$	-0.51	-0.56	0.93	0.97	0.94	0.97	0.97	0.93	0.92	0.62	-0.68	-0.11	1		
$AVG_{CCNV}$	-0.16	-0.24	0.94	0.53	0.95	0.56	0.53	0.94	0.93	0.75	-0.13	0.24	0.83	1	
$FS$	0.28	0.36	0.58	-0.08	0.36	-0.01	0.08	0.54	0.54	0.56	0.51	-0.43	0.28	0.56	1
Target 1 <sup>a</sup>	0.38	<b>0.62</b>	<b>-0.54</b>	<b>-0.84</b>	<b>-0.68</b>	<b>-0.86</b>	<b>-0.84</b>	<b>-0.59</b>	<b>-0.63</b>	<b>-0.44</b>	<b>0.69</b>	<b>-0.18</b>	<b>-0.80</b>	<b>-0.39</b>	0.17
Target 2 <sup>b</sup>	0.16	0.44	-0.38	-0.41	-0.38	-0.41	-0.41	-0.38	-0.38	-0.31	0.37	-0.31	-0.40	-0.26	0.42

Bold represents strong correlation.

<sup>a</sup> Clinical ratings from rater-1.

<sup>b</sup> Clinical ratings from rater-2.

anti-correlations) than the features representing arrhythmia and fatigue. The head movements during tapping (FS) was weakly correlated ( $\mu_2 = 0.17$ ).

In comparison to rater-1, the representative feature of arrhythmia and fatigue were moderately anti-correlated with the clinical ratings provided by rater-2. The features representing pace and amplitude reduction were weakly correlated (or anti-correlated). It can be noticed that the head movements (FS), which was moderately correlated ( $\mu_2 = 0.42$ ) with Target-2 (Table 2), may have created bias in assessing tapping disabilities resulting in coarser scores by rater 2.

Despite the differences between the raters, the fatigue and arrhythmia representative features were either very strongly anti-correlated with Target-1 or moderately anti-correlated with Target-2 (Table 2). Some of these features were perfectly correlated between each other. For example, tapping speed ( $T_s$ ) was perfectly correlated with opening ( $O_T$ ) and closing ( $C_T$ ) velocities. The representative features of fatigue  $VT_s$  and  $\Delta VA$  were perfectly correlated between each other. Redundancies are eliminated by choosing that feature which has the higher sum of absolute correlation values between Target-1 and Target-2. The step eliminates  $T_s$ ,  $O_T$ ,  $VT_s$  and  $VA$  from the list of tapping features and preserves  $C_T$  and  $\Delta VA$ . The remaining 11 features are utilized in tapping classification.

Apart from eliminating feature redundancies, the feature set was further pruned using a chi-squared feature selection algorithm [26]. The chi-squared statistic calculates the goodness of fit of how well an input is correlated with the target class. The method returns a ranking of each feature in decreasing order by the value of chi-squared statistic in relation to the class label. By this way the uncorrelated data can be discarded prior to classification, resulting in improved classification accuracy.

## 2.8. Tapping classification

The utilization of SVM classifiers in biomedical decision support systems has gained immense popularity for their ability to implement flexible decision boundaries in high dimensional feature space. The SVM provides a fast training algorithm that guarantees the optimality of training results [27]. It requires only a little a priori knowledge, i.e. only a labeled dataset. The implicit regularization of classifier's complexity avoids over-fitting and leads to good generalizations. A brief introduction of SVM is given. Details can be found elsewhere [28].

Considering an  $n$ -class classification problem and a set of training vectors  $\{V_i\}_{i=1, \dots, M}$  with corresponding label  $S_i$ , the SVM classifier assigns a new label  $\hat{S}$  to a test vector  $T$  by evaluating

$$\hat{S} = \sum_i \alpha_i S_i K(T, V_i) + b \quad (19)$$

where the weights  $\alpha_i$  and bias  $b$  are SVM parameters which are maximized during SVM training using Eq. (20).

$$\sum_i \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j S_i S_j K(V_i, V_j) \quad (20)$$

Under the constraints

$$0 \leq \alpha_i \leq C \quad \text{and} \quad \sum_i \alpha_i S_i = 0 \quad (21)$$

where  $C$  is a positive constant called the SVM complexity parameter that weights the influence of training errors.  $K(\cdot)$  is the SVM kernel. A universal kernel function based on Pearson VII function (PUK) [29] was utilized in training the SVM. PUK is generally used for the curve fitting purposes and has a general form given in Eq. (22).

$$f(x) = \frac{H}{[1 + ((2(x - x_0)\sqrt{2^{1/\omega} - 1})/\sigma)^2]^\omega} \quad (22)$$

Here  $H$  is the peak height at the center  $x_0$  of the peak and  $x$  is an independent variable. The variables  $\sigma$  and  $\omega$  control the half-width and the tailing factor of peak, respectively. Importantly a curve with  $\omega$  equals to 3 and  $\sigma$  equals to 1, is comparable to a sigmoid function used in the neural network modeling [29].

For a given set of training vectors  $\{V_i\}_{i=1, \dots, M}$ , Ustun et al. [29] modified Eq. (22) to formulate a kernel function given in Eq. (23).

$$K(V_i, V_j) = \frac{1}{[1 + ((2\sqrt{|V_i - V_j|^2} \sqrt{2^{1/\omega} - 1})/\sigma)]^{2\omega}} \quad (23)$$

As can be seen, the single variable  $x$  in Eq. (22) is replaced by two training vectors  $V_i$  and  $V_j$ . A Euclidean distance between these vectors is introduced so that the two identical training vectors would have a zero distance. The peak height  $H$  is replaced by 1 and the peak-offset  $x_0$  is removed.

If  $K(V_i, V_j)$  is a positive definite, then Eqs. (19) and (20) may lead to a convex quadratic programming (QP) optimization problem,

for which the convergence toward global optimum can be guaranteed. This optimization problem can be solved using the sequential minimal optimization (SMO) algorithm [30]. SMO decomposes the overall QP problem into QP sub-problems. This decomposition is performed by solving the smallest possible QP optimization problem at every step involving two Lagrange multipliers satisfying the linear equality constraint to find local optima. At each decomposition step, SMO finds the optimal values for these multipliers and updates the SVM cost function to reflect a new optimal separation between the training vectors holding one class label and the training vectors holding the other class labels.

The SVM classifier based on SMO and configured with PUK kernel function, with  $\omega$  equals to 3 and  $\sigma$  equals to 1, was utilized for tapping classification. A 10-fold cross validation strategy [31] was adopted to obtain unbiased generalization estimates. The confusion matrices (Table 4.1a and 2a) were used to characterize the classification performance. Each row in the confusion matrix represents the actual class instances while each column represents the instances in a predicted class. The diagonals represent the correctly predicted samples or the true positives in each class. In order to improve the classification accuracy, the chi-squared algorithm stratified by 10-fold cross validation was used for feature selection. Further, the receiver operating characteristic (ROC) curve was utilized to analyze the feasibility of classification model independent of the class distribution [32].

Three different classification experiments were performed on the tapping features selected using the Guttman correlation analysis and chi-squared test. In the first experiment the features were used to classify between UPDRS-FT levels. Due to a low agreement between the two raters (Spearman's  $\rho = 0.74$ ), classification tests were performed separately for rater-1 and rater-2 using their ratings as targets. Out of the total 387 video samples, rater-1 categorized 123 videos as '0', 163 as '1', 78 as '2' and 23 videos as '3'. On the other hand, rater-2 categorized 77 videos as '0', 184 as '1', 101 as '2' and 25 videos as '3'. As there were very few samples categorized by the two raters as '3' (severely impaired); in order to avoid standard error in classification, these samples were merged into class '2'. This left behind 3 levels of symptom severities (i.e. '0: normal', '1: mild' and '2: moderate-severe') for classification. In the second experiment, classification test was performed on the tapping features to separate between PD and HC samples. Further tests were made in the third experiment to compare between the proposed CV-based SVM classification scheme and a sensor-based log-linearized Gaussian mixture networks (LLGMN) classification scheme reported by Shima et al. [33].

### 2.8.1. Experiment 1: Classification between UPDRS-FT levels

Two classification matrices of dimensions 11 (selected features)  $\times$  387 (samples) were formed based on the targets provided by rater-1 and rater-2, respectively. The features were normalized on a scale of 0 to 1. A chi-squared value is yielded for each tapping feature fitting the test, and 0 for features failing the test. It was noticeable that the features that showed strong correlations with the clinical ratings in Guttman correlation analysis also produced high chi-square values (Table 3). It was further noticed that FS was ranked second-last for the targets provided by rater-1, whereas it was ranked 5th for the targets provided by rater-2.  $T_E$  failed the chi-squared test and was discarded. FS was removed from the list of features used in tapping classification as it was not directly related to the index-finger movements. The remaining features were used to classify between the levels of symptom severities using the SVM algorithm. Also, the SVM complexity parameter  $C$  was set to 1 for classification experiments. The results are discussed in Section 3.1.

### 2.8.2. Experiment 2: Classification between PD and HC samples

A further classification test was performed on the tapping features to separate between 387 PD and 84 HC samples. The same tapping features which were previously used in classifying the UPDRS-FT levels (given in Table 3) were used to classify between the HC and PD samples. The SVM complexity parameter  $C$  was tuned to improve the generalization performance. The best classification results were achieved at  $C = 13$ . The kernel configuration was kept the same, i.e.  $\omega = 3$  and  $\sigma = 1$ . For feature selection, the goodness of fit of each feature was evaluated using the chi-squared test. According to the test (Table 5a), the representative features of 'slowed pace' and 'amplitude reduction',  $T_Z$ ,  $T_E$  and  $T_n$ , were the top ranked features, followed by the arrhythmia features  $AVG_{CCNV}$  and  $AVG_{CCNP}$ . The fatigue features  $\Delta VA$ ,  $\Delta A$  and  $A_c$  were ranked low.  $\Delta T_n$  and  $C_T$  could not pass the fitness test and were discarded. The remaining 8 features were used in classification. The results are discussed in Section 3.2.

### 2.8.3. Experiment 3: Comparison between SVM and LLGMN schemes

A comparative analysis was made between the proposed CV-based SVM classification scheme and a sensor-based log-linearized Gaussian mixture networks (LLGMN) classification scheme reported by Shima et al. [33]. This LLGMN scheme utilized magnetic sensors. The magnetic coils were attached to the distal parts of index-finger and thumb. A total of 65 RFT samples, 33 PD patient and 32 HC, were acquired. The patient samples were categorized by a clinician between mild (UPDRS-FT '1') and moderate (UPDRS-FT '2') symptom levels, but only with an aim to perform a correlation analysis between the symptom severities and feature quantities. The features were extracted from tapping samples to estimate pace, amplitude and fatigue. Out of the 11 tapping features, 5 features ( $VA$ ,  $O_T$ ,  $C_T$ ,  $T_Z$  and  $T_E$ ) were similar to the features used in the CV-based system (Table 1). Other 6 features were (1) total tapping distance, (2) average maximum amplitude, (3) average finger tapping interval, (4) VC of finger tapping interval, (5) VC of opening velocity, and (6) VC of closing velocity. These 11 features categorized between 33 patients and 32 HC samples with 93.1% accuracy. It can be noticed that this sensor-based system did not incorporate any feature to estimate tapping rhythm.

In order to compare between the CV-based and sensor-based classification schemes, we performed three further experiments to discriminate between HC and PD samples. Since the sensor-based scheme utilized the samples rated 'mild' and 'moderate' to represent the patient groups, in the same way, we selected patient samples rated 'mild' (UPDRS-FT '1') and 'moderate' (UPDRS-FT '2') in our dataset to form three different patient groups. In the first experiment, the samples rated '1' and '2' were combined to represent the first patient group. In the second experiment, the samples rated '1' were selected to represent the second patient group. Similarly, in the third experiment, the samples rated '2' were chosen to represent the third patient group. Due to the difference in opinion between the two raters in identifying UPDRS-FT levels, each classification experiment was performed separately for rater-1 and rater-2, respectively. The SVM complexity parameter and the kernel configuration was kept the same, i.e.  $C = 13$ ,  $\omega = 3$  and  $\sigma = 1$ . The tapping features listed in Table 5a were used. The results are given in Section 3.3.

## 3. Results

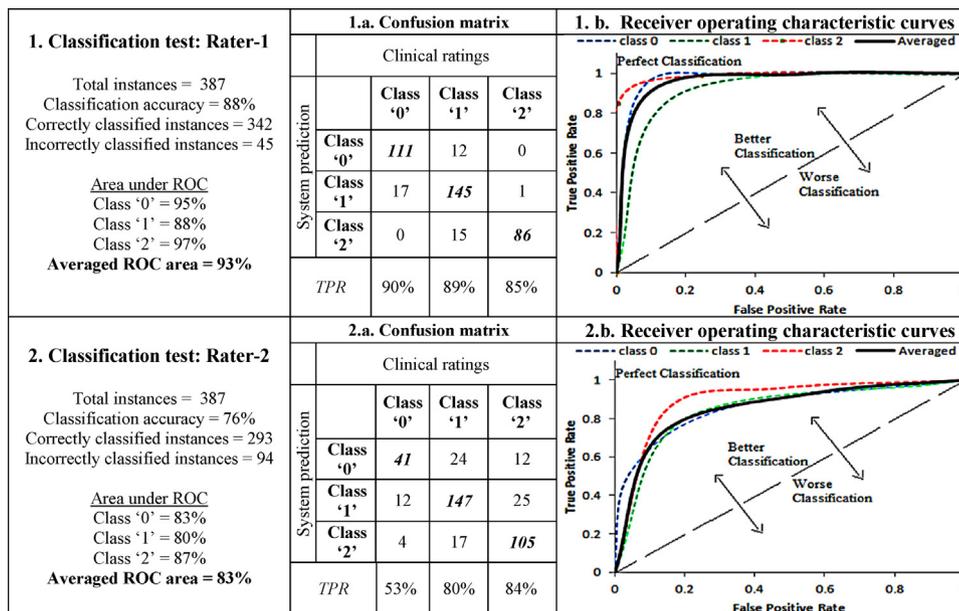
### 3.1. Experiment 1: Classification between UPDRS-FT levels

Two different classification tests were performed on the tapping features using the clinical ratings from rater-1 and rater-2 as targets

**Table 3**  
Chi-squared evaluation of tapping features.

Rater 1			Rater 2		
Rank	Feature	Average merit	Rank	Feature	Average merit
1	$\Delta VA$	199.4 ± 0.92	1	AVGCCNP	84.12 ± 8.25
2	AVGCCNP	169.44 ± 9.13	2	$\Delta VA$	77.56 ± 14.55
3	$\Delta A$	147.64 ± 10.8	3	$T_Z$	77.84 ± 5.98
4	$C_T$	134.89 ± 10.3	4	$\Delta A$	66.06 ± 4.73
5	$T_Z$	131.45 ± 4.93	5	FS	59.61 ± 6.27
6	$T_n$	62.56 ± 11.33	6	$\Delta T_n$	50.25 ± 6.50
7	AVGCCNV	64.73 ± 11.79	7	$C_T$	46.48 ± 3.47
8	$\Delta T_n$	40.15 ± 3.925	8	AVGCCNV	38.16 ± 6.71
9	$A_c$	36.416 ± 2.71	9	$T_n$	31.26 ± 8.11
10	FS	34.716 ± 5.91	10	$A_c$	28.07 ± 2.82
11	$T_E$	0	11	$T_E$	22.03 ± 2.5

**Table 4**  
Post classification analysis.



in test-1 and test-2, respectively. An accuracy of 88% was achieved in test-1 for classification of tapping features between 3-symptom severity levels using 10-fold cross validation (Table 4.1). High true positive rates (TPR) were achieved for classes '0' (90%), '1' (89%) and '2' (85%), respectively. It can be observed in Table 4.1b that the ROC curves for all symptom severity levels are protruding upwards from the diagonal threshold showing that the samples were efficiently distinguished in each symptom class. The area under the ROC curves in symptom class '0', '1' and '2' was 95%, 88% and 97%, respectively. The averaged ROC area was 93%.

In contrast to test-1, the classification accuracy was low in test-2 (76%). Despite of the fact that the averaged area under the ROC curves was high (83%), a low TPR (53%) was observed for classification in the symptom class '0'. One possible reason for a low TPR in '0': normal' class could be that rater-2 might have been influenced by the head movements of patients during tapping, supported by the fact that rater-2 ratings were moderately correlated ( $\mu_2 = 0.42$ ) with FS and that FS was ranked 5th in the chi-squared test for the ratings provided by rater-2. The other possible reason could be that rater-2 rated the samples in the knowledge that the patients were

**Table 5**  
Classification between PD and HC samples.

<b>a. Chi-squared test</b>			<b>b. Classification results</b> Total instances = 471 Classification accuracy = 95.8% Correctly classified instances = 451 Incorrectly classified instances = 20  Average ROC area = 90%	<b>c. Confusion matrix</b>	
Rank	Feature	Average merit		Actual	
1	$T_Z$	225.3+20.8		HC	Patient
2	$T_E$	170.7+7.11		70	14
3	$T_n$	140.9+8.44			
4	AVGCCNV	22.41+2.41		6	381
5	AVGCCNP	15.79+0.57			
6	$\Delta VA$	15.29+1.64		TPR	83%
7	$\Delta A$	15.52+12.7			
8	$\Delta T_n$	0 + - 0		98%	
9	$A_c$	2.005+6.01			
10	$C_T$	0 + - 0			

**Table 6**  
Comparative analysis between LLGMN [33] and SVM.

a. Rater-1						Comparison between LLGMN and SVM 
Subjects	Classification rate	Total instances	Correctly classified	HC TPR	Patients TPR	
HC vs. patient group-1 <sup>a</sup>	93.54%	325 HC=84 Patient=241	304 HC=69 Patient=235	82%	97%	
HC vs. patient group-2 <sup>b</sup>	95.14%	247 HC=84 Patient=163	235 HC=75 Patient=160	90%	98%	
HC vs. patient group-3 <sup>c</sup>	95.06%	162 HC=84 Patient=78	154 HC=80 Patient=74	95%	95%	
b. Rater-2						Comparison between LLGMN and SVM 
Subjects	Classification rate	Total instances	Correctly classified	HC TPR	Patients TPR	
HC vs. patient group-1 <sup>a</sup>	94.31%	369 HC=84 Patient=285	348 HC=72 Patient=276	86%	97%	
HC vs. patient group-2 <sup>b</sup>	91.42%	268 HC=84 Patient=184	245 HC=70 Patient=175	84%	95%	
HC vs. patient group-3 <sup>c</sup>	97.83%	185 HC=84 Patient=101	181 HC=80 Patient=101	95%	100%	

<sup>a</sup> Samples rated UPDRS-FT '1' and '2' by the rater.  
<sup>b</sup> Samples rated UPDRS-FT '1' by the rater.  
<sup>c</sup> Samples rated UPDRS-FT '2' by the rater.

in the advanced stage of PD and with a low expectation that they could perform well in the tapping test.

3.2. Experiment 2: Classification between PD and HC samples

An accuracy of 95.8% was achieved for classification between PD and HC samples (Table 5b). The TPRs were high for both HC (83%) and patient group (98%) (Table 5c). There were a total of 14 false positive samples in the HC group. This could be due to the fact that some of the samples in the patient group were rated '0: normal' by both the clinicians, and that the features computed from the HC samples matched with the features computed from the 'normal' samples in the patient group.

It is important to observe the variation between chi-squared test results for the features used in discriminating between HC and PD samples (Table 5a) and the features used in discriminating between UPDRS-FT levels (Table 3). In case of UPDRS-FT, the representative features of fatigue and arrhythmia were ranked higher by the chi-squared test whereas the features representing pace and amplitude reduction were placed in the lower ranks (Table 3). On the other hand, when classifying between HC and PD patients, the pace and amplitude reduction features were ranked higher comparatively to the arrhythmia and fatigue features (Table 5a). One reason for this difference is that the arrhythmia and fatigue symptoms are evaluated by the clinicians in the 2nd (moderate) and 3rd (severe) symptom levels in the UPDRS-FT (see Table 1) [16]. A clinician, following the UPDRS-FT, would mark a sample as 'moderately' or 'severely' impaired only if he detects fatigue and arrhythmia symptoms in tapping. Besides, the pace is only confined to the 1st (mild)

symptom level of the UPDRS-FT and is ignored in other symptom severity levels.

On the other hand, when separating between the HC and patient group, the representative features of pace,  $T_z$ ,  $T_E$  and  $T_n$  were ranked top by the chi-squared test. The Guttman correlation analysis (Table 2) showed that  $T_z$  increases with the increasing symptom severity. Moreover, the group means for  $T_z$  were  $0.14 \pm 0.005$  for HC and  $0.23 \pm 0.004$  for the patient group within the 95% confidence interval. These mean values were different with statistical significance ( $p < 0.0001$ ). One reason for the increasing value of  $T_z$  is that the muscular rigidity in PD restricts the patients to stretch their index-finger away from the thumb. When trying to tap with lower amplitude and with an intention to tap as fast as possible, the zero-crossings  $T_z$  as well as the number of taps  $T_n$  increases. Besides, the signal energy  $T_E$  reduces due to low amplitude in quick tap successions. This is supported by the fact that the group means for  $T_E$  were higher for HC ( $0.24 \pm 0.007$ ) and lower for the patient group ( $0.15 \pm 0.002$ ) within the 95% confidence interval. Importantly, the mean values were different with statistical significance ( $p < 0.0001$ ). Moreover, the Guttman correlates showed a decreasing trend for  $T_E$  (Table 2) with an increasing symptom severity.

3.3. Experiment 3: Comparison between SVM and LLGMN schemes

In the first experiment to discriminate between HC and the first patient group, for rater-1, the CV-based tapping features categorized between 241 patients (TPR=97%) and 84 HC (TPR=82%) samples with an accuracy of 93.5% in 10-fold cross-validation (Table 6a). For rater-2, the same classification scheme categorized

between 285 patients (TPR=97%) and 84 HC (TPR=86%) samples with an accuracy of 94.3% (Table 6b).

In the second experiment to discriminate between HC and the second patient group, for rater-1, the SVM classification scheme categorized between 163 patients (TPR=98%) and 84 HC (TPR=90%) with 95.1% accuracy (Table 6a). For rater-2, the same scheme categorized between 184 patients (TPR=95%) and 84 HC (TPR=84%) with 91.42% accuracy (Table 6b).

Lastly, in the third experiment to discriminate between HC and the third patient group, for rater-1, the SVM classification scheme categorized between 78 patients (TPR=95%) and 84 HC (TPR=95%) samples with 95.1% accuracy (Table 6a). For rater-2, the same scheme categorized between 101 patients (TPR=100%) and 84 HC (TPR=95%) with 97.8% accuracy (Table 6b).

In all these classification experiments (Table 6), the proposed CV-based SVM scheme discriminated between HC and patient group with an average 94.5% accuracy which is comparable to the 93.1% classification accuracy of sensor-based LLGMN system [33]. Since both these classification schemes, CV-based and sensor-based, utilized more or less the same tapping features, the experiments determine the efficacy of features in producing high classification rates irrespective of the type of method being used to produce the tapping signals.

#### 4. Discussion

In this paper we have introduced a novel CV framework for objective assessment of rapid finger tapping. The uniqueness of this method is the utilization of facial features to normalize the tapping signal and to cope with camera calibration. Besides, the method is able to adjust and quantify tapping signals that are affected by involuntary movements of hands. Importantly, the extracted tapping features showed strong statistically significant correlation with the experts' opinion on tapping impairment. The described framework exhibited good classification performance in separating symptom severity levels based on clinical ratings, as well as in identifying PD patients and HC.

The method has advantages compared to other technologies (magnetic sensors [10,11,33], accelerometers [2,5,8,9,12]) that have been developed for PD evaluation to support subjective judgment of raters. It does not require expensive setups and expertise in operating the software. Rather it only requires a computer with a webcam, which are low cost and are normally used in the household. The subjects can perform the tapping task naturally in the same manner as instructed by the movement disorder expert.

It may be further argued if it is necessary to perform face detection for tapping signal normalization since the present day RGB-D (Microsoft Kinect [34]) sensor allows computing the real distance between index-finger tip position and camera. However, the current studies on motion detection [35–42] agree that kinect sensor is not capable to detect fine body movements, although it can be used to detect gross motion of the body. It is primarily due to the reason that the depth data provided by kinect presents several noise-related problems such as holes in the depth image [39], misalignment in the image and the lack of sharp object boundaries [40], which suggests that kinect is unfeasible to be used for tracking fine motion details such as finger movements. To comprehend the problem of holes, studies have used morphological dilation and erosion filters [41] that are again not suitable to be used for tracking a moving index-finger, which is thin. These morphological filters can cause incorrect readings due to the dilated and eroded pixels of index-finger top position.

Instead, the proposed methodology can be optimized with high resolution video recordings which may allow constructing a hand model to measure biomechanics such as joint angles and flexions of fingers and thumb. Another alternative is to use a recently

introduced technology called leap motion [43], which is capable to compute exoskeleton of human hand. Current research is now combining leap motion and kinect sensors to perform finger tapping analysis [44]. However, there are drawbacks in leap that needs further work [45]. For instance, one of the leap's problems is inconsistency in the quality and behavior of its application 'Airscape Home', which uses a 3d vertical virtual plane to register input. According to users [46], this framework works ineffectively in practice because there is no physical feedback for when the finger has crossed the plane. In case if leap is utilized to track finger motion in PD, this will require repeated experiments only because the patients are unaware of the placement of their hands and the feedback about their hands' position is not given. This will further complicate the symptom analysis because the repeated experiments will cause exaggeration and fatigue for the patients that will affect the accuracy of measuring symptom severity. Another problem reported by the leap users is that, it is difficult to control drawings on computer screen and it is very frequent that things may get clicked accidentally. This suggests that leap is very sensitive to motion and this lack of precision can produce great amount of noise if it is used for finger motion analysis. Given that these problems are solved in the future leap updates, the technology can be effective in estimating finger tapping symptoms in PD.

There were some limitations associated with this work. First, the agreement between the clinical ratings provided by the two raters was not very strong (Spearman's  $\rho=0.74$ ). One possible extension would be to utilize clinical ratings from another expert to validate the classification performance. The second limitation was the unavailability of enough video recordings which were marked '3' (severely impaired) by the clinicians. Further, there were no cases which were marked '4' (unable to perform tapping) by the clinicians based on the UPDRS-FT.

The plan now is to accommodate the described CV framework into a recently developed test-battery system for PD symptom assessment [47]. This system has the potential to generate a comprehensive symptom profile by quantifying a variety of fine motor tests (spiral drawing and self-assessed symptoms) through statistical and machine learning algorithms. The inclusion of RFT into this system would enhance the coverage in symptom profiling. The system would be able to process patient-recorded tapping videos uploaded to a central server through a web interface. It could provide a feasible solution for clinicians to track disease progression and evaluate treatment interventions as well as benefit the patients having physical restrictions by allowing at-home monitoring.

In summary, the high classification accuracies achieved in separating between the UPDRS-FT symptom severity levels as well as in separating between the HC and patient group support the feasibility of tapping features and SVM classification model to be used for tapping symptom classification. Importantly, the tapping features which were previously estimated by the sensor-based methods can now be estimated by a non-invasive CV-based algorithm.

#### Acknowledgments

This research is a module of project 'PAULINA' which has been running in Dalarna University in collaboration with Animech and Nordforce Technology, and is funded by a grant from the Swedish Knowledge Foundation.

#### References

- [1] Collyer CE, Broadbent HA, Church RM. Preferred rates of repetitive tapping and categorical time production. *Perception & Psychophysics* 1994;55(4):443–53. <http://dx.doi.org/10.3758/BF03205301>.

- [2] Shimoyama I, Ninchoji T, Uemura K. The finger-tapping test: a quantitative analysis. *Archives of Neurology* 1990;47(6):681–4, <http://dx.doi.org/10.1001/archneur.1990.00530060095025>.
- [3] Volkow ND, Gur RC, Wang GJ, Fowler JS, Moberg PJ, Ding YS, et al. Association between decline in brain dopamine activity with age and cognitive and motor impairment in healthy individuals. *American Journal of Psychiatry* 1998;155(3):344–9.
- [4] Findley LJ. The economic impact of Parkinson's disease. *Parkinsonism & Related Disorders* 2007;13(Suppl.):S8–12, <http://dx.doi.org/10.1016/j.parkreldis.2007.06.003>.
- [5] Shimoyama I, Hinokuma K, Ninchoji T, Uemura K. Microcomputer analysis of finger-tapping as a measure of cerebellar dysfunction. *Neurologia Medico-Chirurgica* 1983;23(6):437–40, <http://dx.doi.org/10.2176/nmc.23.437>.
- [6] Konczak J, Ackermann H, Hertrich I, Spieker S, Dichgans J. Control of repetitive lip and finger movements in Parkinson's disease. *Movement Disorders* 1997;12(5):665–76, <http://dx.doi.org/10.1002/mds.870120507>.
- [7] Agostino R, Currà A, Giovannelli M, Modugno N, Manfredi M, Berardelli A. Impairment of individual finger movements in Parkinson's disease. *Movement Disorders* 2003;18(5):560–5, <http://dx.doi.org/10.1002/mds.10313>.
- [8] Okuno R, Yokoe M, Fukawa K, Sakoda S, Akazawa K. Measurement system of finger-tapping contact force for quantitative diagnosis of Parkinson's disease. In: Clark J, editor. *Proceedings of the 29th annual IEEE conference on engineering in Medicine and Biology Society (EMBS)*. Lyon, France: IEEE Press; 2007. p. 1354–7, <http://dx.doi.org/10.1109/IEMBS.2007.4352549>.
- [9] Okuno R, Yokoe M, Akazawa K, Abe K, Sakoda S. Finger taps movement acceleration measurement system for quantitative diagnosis of Parkinson's disease. In: Dhawan AP, Akay M, editors. *Proceeding of the 28th annual IEEE conference on engineering in Medicine and Biology Society (EMBS)*. New York, USA: IEEE Press; 2006. p. 6623–6, <http://dx.doi.org/10.1109/IEMBS.2006.2609904>.
- [10] Kandori A, Yokoe M, Sakoda S, Abe K, Miyashita T, Oe H, et al. Quantitative magnetic detection of finger movements in patients with Parkinson's disease. *Neuroscience Research* 2004;49(2):253–60, <http://dx.doi.org/10.1016/j.neures.2004.03.004>.
- [11] Shima K, Tsuji T, Kan E, Kandori A, Yokoe M, Sakoda S. Measurement and evaluation of finger-tapping movements using magnetic sensors. In: Dumont G, Galiana H, editors. *Proceeding of the 30th annual IEEE conference on engineering in Medicine and Biology Society (EMBS)*. Vancouver, BC, Canada: IEEE Press; 2008. p. 5628–31, <http://dx.doi.org/10.1109/IEMBS.2008.4650490>.
- [12] Elble RJ. Gravitational artifact in accelerometric measurements of tremor. *Clinical Neurophysiology* 2005;116(7):1638–43, <http://dx.doi.org/10.1016/j.clinph.2005.03.014>.
- [13] Goetz CG, Stebbins GT, Wolff D, DeLeeuw W, Bronte HS, Elble R, et al. Testing objective measures of motor impairment in early Parkinson's disease: feasibility study of an at-home testing device. *Movement Disorders* 2009;24(4):551–6, <http://dx.doi.org/10.1002/mds.22379>.
- [14] Kupryjanow A, Kunka B, Kostek B. UPDRS tests for diagnosis of Parkinson's disease employing virtual-touchpad. In: Tjoa AM, Wagner RR, editors. *Proceedings of the IEEE workshop on Database and Expert Systems Applications (DEXA)*. 2010. p. 132–6, <http://dx.doi.org/10.1109/DEXA.2010.87>.
- [15] Nyholm D, Remahl AN, Dizdar N, Constantinescu R, Holmberg B, Jansson R, et al. Duodenal levodopa infusion monotherapy vs. oral polypharmacy in advanced Parkinson disease. *Neurology* 2005;64(2):216–23, <http://dx.doi.org/10.1212/01.WNL.0000149637.70961.4C>.
- [16] Fahn S, Elton RL, Members of the UPDRS Committee. Unified Parkinson's Disease Rating Scale. In: Fahn S, Marsden CD, Goldstein M, Calne DB, editors. *Recent developments in Parkinson's disease*. New Jersey: McMillan Health Care; 1987. p. 153–63.
- [17] Livio M. *The golden ratio: the story of phi, the world's most astonishing number*. New York: Random House Digital Inc; 2008.
- [18] Criss K, McNames J. Video assessment of finger-tapping for Parkinson's disease and other movement disorders. In: Bonato P, Brennan C, editors. *Proceedings of the 31st annual IEEE conference on engineering in Medicine and Biology Society (EMBS)*. Boston, MA, USA: IEEE Press; 2011. p. 7123–6, <http://dx.doi.org/10.1109/IEMBS.2011.6091800>.
- [19] Turk MA, Alex P. Face recognition using Eigenfaces. In: *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'91)*. Maui, HI: IEEE Press; 1991. p. 586–91, <http://dx.doi.org/10.1109/CVPR.1991.139758>.
- [20] Viola P, Jones MJ. Robust real-time face detection. *International Journal of Computer Vision* 2004;57(2):137–54, <http://dx.doi.org/10.1023/B:VISI.0000013087.49260.fb>.
- [21] Bradski GR, Davis JW. Motion segmentation and pose recognition with motion history gradients. *Machine Vision and Applications* 2002;13(3):174–84, <http://dx.doi.org/10.1007/s001380100064>.
- [22] Ledley RS, Buas M, Golab TJ. Fundamentals of true-color image processing. In: *Proceedings of the 10th IEEE international conference on pattern recognition*. Atlantic City, NJ: IEEE Press; 1990. p. 791–5, <http://dx.doi.org/10.1109/ICPR.1990.118218>.
- [23] Yoshinaga H, Miyazima S, Mitake S. Fluctuation of biological rhythm in finger-tapping. *Physica A: Statistical Mechanics and its Applications* 2000;280(3):582–6, [http://dx.doi.org/10.1016/S0378-4371\(99\)00594-4](http://dx.doi.org/10.1016/S0378-4371(99)00594-4).
- [24] Edward V, Windischberger C, Cunnington R, Erdler M, Lanzerberger R, Mayer D, et al. Quantification of fMRI artifact reduction by a novel plaster cast head holder. *Human Brain Mapping* 2000;11(3):207–13, [http://dx.doi.org/10.1002/1097-0193\(200011\)11:3<207::AID-HBM60>3.0.CO;2-J](http://dx.doi.org/10.1002/1097-0193(200011)11:3<207::AID-HBM60>3.0.CO;2-J).
- [25] Guttman L. A basis for scaling qualitative data. *American Sociological Review* 1944;9(2):139–50.
- [26] Geng X, Liu TY, Qin T, Li H. Feature selection for ranking. In: Clarke C, Fuhr N, Kando N, editors. *Proceedings of the 30th annual international ACM SIGIR conference on research and development in information retrieval*. New York, NY, USA: ACM Press; 2007. p. 407–14, <http://dx.doi.org/10.1145/1277741.1277811>.
- [27] Guan W. *New support vector machine formulations and algorithms with application to biomedical data analysis*. Atlanta, GA, USA: Ph.D. thesis, Georgia Institute of Technology; 2011.
- [28] Decoste D, Scholkopf B. Training invariant support vector machines. *Machine Learning* 2002;46(1–3):161–90, <http://dx.doi.org/10.1023/A:1012454411458>.
- [29] Ustun B, Melssen WJ, Buydens LMC. Facilitating the application of Support Vector Regression by using a universal Pearson VII function based kernel. *Chemometrics and Intelligent Laboratory Systems* 2006;81(1):29–40, <http://dx.doi.org/10.1016/j.chemolab.2005.09.003>.
- [30] Scholkopf B, Platt JC, Shawe TJ, Smola AJ, Williamson RC. Estimating the support of a high-dimensional distribution. *Neural Computation* 2001;13(7):1443–71, <http://dx.doi.org/10.1162/089976601750264965>.
- [31] Stone M. Cross-validated choice and assessment of statistical predictions. *Journal of the Royal Statistical Society Series B (Methodological)* 1974;36(2):111–47.
- [32] Metz CE. Basic principles of ROC analysis. *Seminars in Nuclear Medicine* 1978;8(4):283–98, [http://dx.doi.org/10.1016/S0001-2998\(78\)80014-2](http://dx.doi.org/10.1016/S0001-2998(78)80014-2).
- [33] Shima K, Tsuji T, Kandori A, Yokoe M, Sakoda S. Measurement and evaluation of finger tapping movements using log-linearized Gaussian mixture networks. *Sensors* 2009;9(3):2187–201, <http://dx.doi.org/10.3390/s90302187>.
- [34] Han J, Shao L, Xu D, Shotton J. Enhanced computer vision with Microsoft Kinect sensor: a review. *IEEE Transactions on Cybernetics* 2013;43(5):1318–34, <http://dx.doi.org/10.1109/TCYB.2013.2265378>.
- [35] Calderita LV, Bandera JP, Bustos P, Skiadopoulos A. Model-based reinforcement of Kinect depth data for human motion capture applications. *Sensors* 2013;13(7):8835–55, <http://dx.doi.org/10.3390/s130708835>.
- [36] Dong F, Ieng SH, Savatier X, Etienne-Cummings R, Benosman R. Plenoptic cameras in real-time robotics. *International Journal of Robotics Research* 2013;32(2):206–17, <http://dx.doi.org/10.1177/0278364912469420>.
- [37] Ren G, O'Neill E. 3D selection with freehand gesture. *Computers & Graphics* 2013;37(3):101–20, <http://dx.doi.org/10.1016/j.cag.2012.12.006>.
- [38] Pedrocchi A, Ferrante S, Ambrosini E, Gandolla M, Casellato C, Schauer T, et al. MUNDUS project: Multimodal Neuroprosthesis for daily upper limb support. *Journal of Neuroengineering and Rehabilitation* 2013;10(1):66, <http://dx.doi.org/10.1186/1743-0003-10-66>.
- [39] Hwang J, Lee K, Kim J, Lee S. A novel hole filling method using image segmentation-based image inpainting. In: Moeller R, editor. *IEEE international conference on consumer electronics (ICCE)*. Berlin, Germany: IEEE Press; 2013. p. 470–1, <http://dx.doi.org/10.1109/ICCE.2013.6486980>.
- [40] Gallego J, Pardàs M. Region based foreground segmentation combining color and depth sensors via logarithmic opinion pool decision. *Journal of Visual Communication and Image Representation* 2013, <http://dx.doi.org/10.1016/j.jvcir.2013.03.019> (in press).
- [41] Camplani M, Mantecón T, Salgado L. Depth-color fusion strategy for 3-d scene modeling with Kinect. *IEEE Transactions on Cybernetics* 2013, <http://dx.doi.org/10.1109/TCYB.2013.2271112> (in press).
- [42] Vijayanagar KR, Loghman M, Kim J. Real-time refinement of Kinect depth maps using multi-resolution anisotropic diffusion. *Mobile Networks and Applications* 2013, <http://dx.doi.org/10.1007/s11036-013-0458-7> (in press).
- [43] Hodson H. Leap motion hacks show potential of new gesture tech. *New Scientist* 2013;218(2911):21, [http://dx.doi.org/10.1016/S0262-4079\(13\)60864-7](http://dx.doi.org/10.1016/S0262-4079(13)60864-7).
- [44] Tarnanas I, Schlee W, Tsolaki M, Müri R, Mosimann U, Nef T. Ecological validity of virtual reality daily living activities screening for early dementia: longitudinal study. *JMIR Serious Games* 2013;1(1), <http://dx.doi.org/10.2196/games.2778>, e1, 1–13.
- [45] Weichert F, Bachmann D, Rudak B, Fissler D. Analysis of the accuracy and robustness of the leap motion controller. *Sensors* 2013;13(5):6380–93, <http://dx.doi.org/10.3390/s130506380>.
- [46] Newman J. Leap motion first impressions: there's work to do. <http://www.pcworld.com/article/2044909/leap-motion-first-impressions-theres-work-to-do.html> (accessed 11.10.2013).
- [47] Memedi M, Westin J, Nyholm D, Dougherty M, Groth T. A web application for follow-up of results from a mobile device test battery for Parkinson's disease patients. *Computer Methods and Programs in Biomedicine* 2011;104(2):219–26, <http://dx.doi.org/10.1016/j.cmpb.2011.07.017>.