Smartphone Application for Classification of Motor Impairment Severity in Parkinson's Disease

Blake P. Printy*, Lindsey M. Renken*, John P. Herrmann, Isac Lee, Bryant Johnson, Emily Knight, Georgeta Varga and Diane Whitmer

***Abstract*— Advanced hardware components embedded in modern smartphones have the potential to serve as widely available medical diagnostic devices, particularly when used in conjunction with custom software and tested algorithms. The goal of the present pilot study was to develop a smartphone application that could quantify the severity of Parkinson's disease (PD) motor symptoms, and in particular, bradykinesia. We developed an iPhone application that collected kinematic data from a small cohort of PD patients during guided movement tasks and extracted quantitative features using signal processing techniques. These features were used in a classification model trained to differentiate between overall motor impairment of greater and lesser severity using standard clinical scores provided by a trained neurologist. Using a support vector machine classifier, a classification accuracy of 0.945 was achieved under 6-fold cross validation, and several features were shown to be highly discriminatory between more severe and less severe motor impairment by area under the receiver operating characteristic curve (AUC $>$ 0.85). Accurate classification for discriminating between more severe and less severe bradykinesia was not achieved with these methods. We discuss future directions of this work and suggest that this platform is a first step toward development of a smartphone application that has the potential to provide clinicians with a method for monitoring patients between clinical appointments.**

I. INTRODUCTION

Parkinson's disease (PD) is a complicated, chronic, and debilitating neurodegenerative disease estimated to affect 1-2% of the population over the age of 60 [1]. It is characterized by a heterogeneous set of motor deficits including tremor, bradykinesia, rigidity, freezing, and postural instability. Bradykinesia, or the slowness of voluntary movement, is a cardinal symptom of PD and correlates with overall motor impairment [2]. It is therefore of interest to accurately quantify the severity of bradykinesia in PD patients, not only to improve diagnostics, but also to aid clinicians in devising maximally effective treatment strategies.

The clinical "gold standard" for evaluating the severity of PD symptoms is an assessment called the Unified Parkinson's Disease Rating Scale, which includes an evaluation of motor symptoms referred to as the motor subscore (UPDRS III), along with questionnaires about quality of life, mentation and mood, and complications of drug therapy [3]. The evaluation of upper extremity bradykinesia within the UPDRS III involves a trained clinician's observing and scoring of a patient during three kinematic tasks: 1) repetitive wrist pronation/supination for 15 seconds, 2) rhythmic tapping of the thumb and index finger together for 15 seconds, and 3) opening and closing of the hand for 15 seconds. Speed, amplitude, and rhythmicity are qualitatively assessed for each task and used to score patients on a scale from 0 (normal motor activity) to 4 (the most severe impairment). As with any subjective assessment, this way of scoring patients results in some inconsistency across clinicians; several studies have demonstrated low inter-rater reliability [4].

Movement disorders researchers have recently invested in developing automated, quantitative methods for assessing motor symptoms of PD. Specialized equipment including electronic MIDI keyboards [2], accelerometers [5] and gyroscope sensors [5,6,7] have been used to capture kinematics of PD patients' performance of specific voluntary motor tasks. Data from patient-worn gyroscope and accelerometer sensors have also been used to effectively predict global dyskinesia severity during arm resting or extension [8], the severity of rest, postural and kinetic tremors [9], and speed, amplitude and rhythm components of UPDRS scores [4]. These devices require specialized equipment that is either expensive or custom built in a lab setting and usually require technical expertise. With health care practices evolving toward personalized treatments and telemedicine, there is potential utility in developing diagnostic systems that could be easily used by patients at home [8].

The ubiquity of motion sensors embedded in smartphones and open access to software development toolkits suggest smartphones as viable devices for automated kinematic assessment. The objective of the present study was to develop a platform that uses smartphone hardware features, custom software and algorithms, and clinically tested kinematic tasks to automatically and accurately assess overall disease severity, motor impairment, and bradykinesia in PD.

II. METHODS

A. Kinematic Data Acquisition

A total of 26 adult patients with idiopathic PD gave informed consent to participate in this IRB-approved research study in a movement disorders clinic on the day of an

Funding for this project was provided by the St. David's Foundation Impact Fund.

B. Printy, L. M. Renken, J. P. Herman, I. Lee, and B. Johnson are with the Dept. of Biomedical Engineering, The University of Texas at Austin, Austin, TX 78712, blakeprinty@utexas.edu.

G. Varga, M.D. is with the Parkinson Disease - Movement Disorders Clinic, Austin, TX 78746, marigeta@yahoo.com.

E. Knight is with the NeuroTexas Institute Research Foundation at St. David's Healthcare, Austin, TX 78705.

D. Whitmer, Ph.D. (corresponding author) is jointly affiliated with the NeuroTexas Institute Research Foundation at St. David's Healthcare, Austin, TX 78705, and the Dept. of Neuroscience, The University of Texas at Austin, Austin, TX 78712, dwhitmer@austin.utexas.edu.

*These authors contributed equally to the work.

978-1-4244-7929-0/14/$26.00 ©2014 IEEE

outpatient appointment. A total of 18 patients (8 female) participated in the complete set of kinematic and clinical tasks with each hand. The cohort of patients participating were 68.5 +/- 12.1 (mean +/- std) years of age and had a disease duration of 5.5 +/- 2.5 (mean +/- std) years.

Research participants first underwent a clinical exam and either the full UPDRS assessment (N = 12) or the UPDRS III (N = 6), performed by a board-certified movement disorders neurologist (GV). We defined Bradykinesia Subscore (BSS) for each arm as the sum of questions 23-25 from the UPDRS [3], a value that can range from 0 in the case of no motor impairment to 12 with maximal impairment. The UPDRS total, UPDRS III, and BSS for the left and right arms (separately) were stored in a secure database to serve as clinical ground truth for offline data analyses. Left and right hand measurements were treated as independent because movement disorders clinicians assess the right and left side of the body independently.

Participants next performed a battery of kinematic tasks using an iPhone 5C (Apple, Inc.) and our custom software application. These kinematic tasks were proctored by members of the research team who were blinded to the clinical UPDRS assessments. Participants performed a series of four kinematic tasks, three trials each, with each upper extremity. The order of tasks and hands tested were randomized in order to eliminate any potential bias introduced by the sequence of experiments. For each task, the patient was instructed to make the movements as quickly and as large as possible. The four tasks were: 1) hand opening and closing for 15 seconds over the screen of the smartphone while the phone was lying flat on a table, 2) repetitive tapping of the index finger for 15s on the screen of the smartphone, 3) alternating tapping between the index and middle finger for 15s on the screen of the smartphone, and 4) repetitive wrist pronation/supination for 15s with arms extended straight in the front of the patient and the smartphone in a holder strapped to the dorsal side of the patient's hand. The alternating finger-tapping task was also utilized because previous research has demonstrated that alternating finger tapping on a MIDI keyboard correlates highly with UPDRS III scores and with bradykinesia [2]. Although three trials of each task were measured, our data analyses used only the first trial in order to eliminate variability resulting from varying rest periods between trials across tasks and patients.

B. Custom Software Application

We developed a custom iPhone (Apple, Inc.) application to capture kinematic data from tasks imitating those performed during the UPDRS III [3] and from an alternating finger tapping task that had been tested previously with PD patients [2]. During application use, users first entered the system by providing a unique identifier via a login screen (Fig. 1) and then navigated through a series of views corresponding to specific tasks performed throughout the data collection process. All patients were assigned unique identifiers in order to keep proctected health information strictly confidential throughout the study. For each session, the software automatically presented the four movement tasks in a randomized task order.

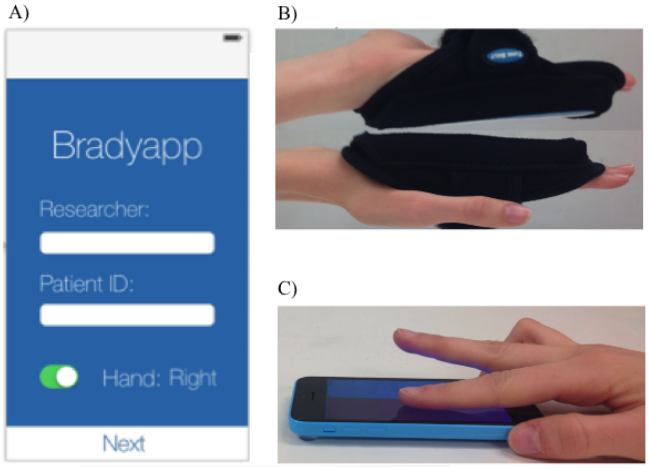

Fig. 1. Diagrams of application use. A) login screen, B) positioning during wrist pronation/supination task, C) positioning during finger tapping task.

Five separate built-in hardware components were accessed throughout a typical usage session: the gyroscope, accelerometer, capacitive touch screen, microphone, and the front-facing camera. Gyroscope and accelerometer data (task 4) were sampled from each orthogonal coordinate at 100 Hz. Data from the capacitive touch screen were acquired at 100 Hz (task 2, 3) in the form of boolean values describing the contact of a user's finger with the screen (i.e. 1 when the patient has their finger on the screen, and 0 otherwise). Sound data were captured from the microphone and initially stored in Core Audio Format (.caf), before conversion into Waveform Audio (.wav) format for external processing (task 2, 3). Video data from the front-facing camera were acquired at 30 Hz (task 1) and written to Quicktime Movie (.mov) format for external processing. During execution, only hardware components associated with movement for each task were accessed, and data specific to each task were stored within the application sandbox throughout application use. These data were later extracted and packaged for analysis using the Organizer window of Xcode 5 (Apple, Inc.). Acquired data were visually inspected for correspondence with movement features during application development and throughout clinical data collection for noted instances of significant fatigue (slowing of movement and decreases in amplitude [2]) and pauses; rigorous quantitative reliability and validity testing were not performed.

C. Feature Quantification

We quantified features from our kinematic task data that we hypothesized would be informative for describing frequency, amplitude and rhythmicity, which are the movement features visually assessed by neurologists during the repetitive movements of the UPDRS III [4]. For the hand opening/closing task, movement signals were extracted from video as the sum of pixel brightness values for each frame given that recordings were acquired under constant

lighting conditions and fixed hand distances within sessions. Average frequencies of repetitive hand movements (task 4) and wrist movements (task 1) were quantified as the peak frequency from the power spectral density (PSD), computed using Welch's method with 4 second windows with 50% overlap. Average frequency from tap data (tasks 2, 3) was calculated by averaging the number of taps per second over 5s windows with 50% overlap. A related quantity, speed of movement, was quantified as the root mean square (RMS) angular velocity from gyroscope signals (task 4) [6]. Similarly, a surrogate for speed of movement was quantified as the average peak value of the microphone signal (tasks 2,3), calculated with 5s windows with 50% overlap, under the assumption that the force of the tap is correlated with velocity of movement. Amplitude of the dominant rhythm was quantified as the power of the peak frequency from the PSD normalized to the total power of the PSD (tasks 1,4). Rhythmicity was quantified as the coefficient of variation (CV) of the intervals between successive strikes of the same finger and the CV of the durations of contact between each finger and a specified region of the touchscreen (tasks 2,3), which were preprocessed using a log10 transformation to remove skewness in the distribution of values [2]. To assess the slowing of movement as a function of time, we computed the PSD from consecutive 4s segments of the data trace (using 1.5 s windows and 50% overlap with Welch's method), and extracted the slope of a least-squares fit of the peaks over each window (tasks 1,4). As patients fatigue during the wrist rotation task, their movements may become more imprecise. To capture the amount of signal that came from "off-axis" gyroscope signals, we computed the RMS value from the cross-correlation of each of the X- and Z-axis gyroscope signals with the Y-axis signal, and normalized these with the RMS value from the Y-axis auto-correlation.

D. Classification Methods

Due to limitations introduced by the number of samples in the analysis, a rating scheme was implemented to simplify the multi-class problem into a binary one by dividing the data into two groups based on the midpoint of possible UPDRS III scores: patients with scores less than 35 (N = 7) were labelled as having "less severe" symptoms, and those with scores greater than or equal to 35 (N = 11) were labelled as having "more severe" symptoms. Similary, a separate evaluation was performed with the BSS, where scores less than 6 were labelled as "less severe" (N = 17) and those greater than or equal to 6 were "more severe" (N = 19). UPDRS totals were not used in the analyses because of limitations in sample size. From the feature data, machine learning models were developed to classify patients according to two independent response variables: the UPDRS III for the most affected hand/arm and the BSS for each hand/arm.

Performance of models for both the UPDRS III and the BSS were evaluated (separately) using 6-fold cross-validation, which was chosen to most evenly distribute data in testing folds. Both a support vector machine (SVM), built under the C-SVC formulation with a radial basis function (RBF) kernel, and a random forest model with 500 trees, were used (separately) as the classification engine in the model. The RBF kernel was chosen for its ability to identify non-linear relationships in the feature space. During each fold of cross validation, the following algorithm was used to test the model: 1) Use stratified sampling to split the data into training $\left(\frac{5}{6}\right)$, and testing $\left(\frac{1}{6}\right)$ subsets, 2) Train the classifier with the training set, 3) Predict UPDRS III or BSS classes for patients within the testing set using the SVM classifier, 4) Calculate performance metrics. In order to assess the accuracy of each model, classification error (percentage of wrongly-classified subjects) and area under the non-parametric receiver operating characteristic curve (AUC) were respectively calculated from classification likelihoods after cross-validation. In addition to this, several of the most discriminatory features were tested for their association with PD motor impairment in a multiple linear regression model using UPDRS III scores as the model response variable. A logistic regression model with multiple explanatory variables was also tested for the more severe/less severe UPDRS III rating scheme.

III. RESULTS

A. Clinical Data

The distributions of UPDRS III scores from the most affected side and Bradykinesia Subscores per side are depicted in Fig. 2. Fig. 3 details two types of signals captured from the iPhone during a typical usage session. Examples are given for a patient with less severe symptoms (L) and a patient with more severe symptoms (R).

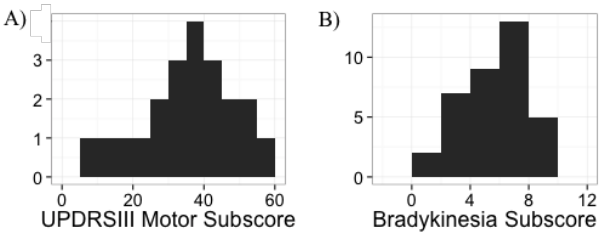

Fig. 2. Histograms of of clinically assessed scores of motor severity: A) UPDRS III, B) Bradykinesia Subscores.

B. Classification Results

Among all of the quantified features, the top 10% most discriminatory in the less severe/more severe UPDRS IIII classification problem include: CV log10 duration of single finger tap (CvDurSingle), CV log10 duration of finger tapping for both fingers during alternating tapping (CvDurAlt), mean peak loudness for the index finger from alternating finger tapping (AvgLoudnessIndex), CV peak loudness for the middle finger from alternating finger tapping, and mean peak loudness for the middle finger from alternating finger tapping. These features were all found to have an area under the ROC curve of greater than 0.85, where the less severe/more severe UPDRS III label was used as the response variable in calculation. Fig. 4 shows correlation plots for the

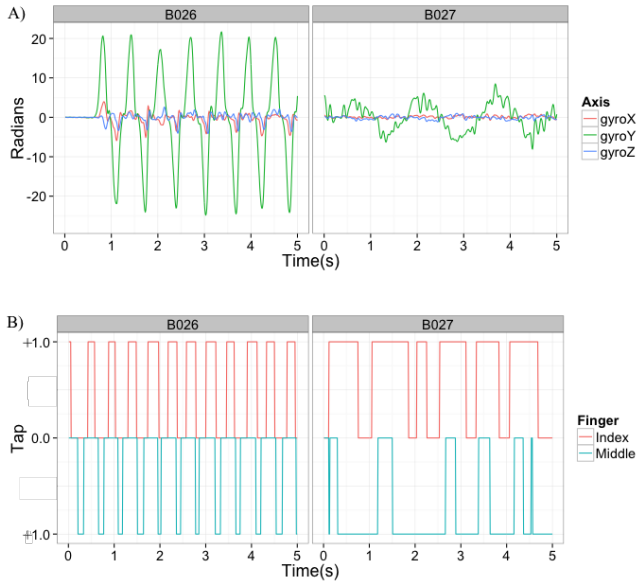

Fig. 3. Examples of raw signals collected from the iPhone during different movement tasks. (A) depicts gyroscope signals for a patient with less severe (left) and more severe (right) motor symptoms, and (B) depicts alternating tapping touch-screen signals (inverted for visual clarity) for a patient with less severe (left) and more severe (right) motor symptoms.

CvDurAlt and AvgLoudnessIndex features. Combinations of the top performing features were tested in each of the classification models. Significant classification results with UPDRS III scores as response variables included models using the CvDurSingle, CvDurAlt, and AvgLoudnessIndex features within each of the classifiers (separately). The SVM classifier performed with 0.055 average error rate (AUC = 0.9166), and the random forest classifier performed with 0.111 error rate (AUC = 0.9166).

Among the regression models constructed with combinations of top-performing features as predictor variables, one showed significant association between predictor variables and UPDRS III scores. A multiple linear regression model with CvDurSingle and the coefficient of variation of the interval of alternating finger tapping (CvIntAlt) showed that CvIntAlt was significantly associated with UPDRS III scores ($p = 0.0412$, multiple $R^2 = 0.232$).

No significant classification accuracy was achieved using BSS scores as response variables in either classifier for any tested feature set (0.35 minimum error rate for both classifiers).

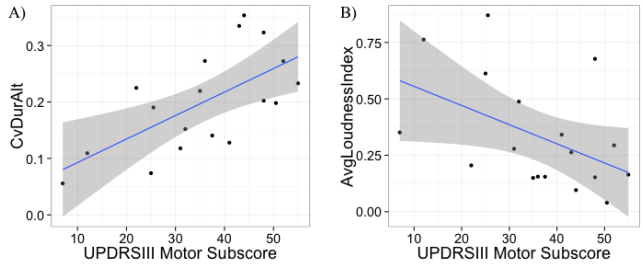

Fig. 4. Scatterplot with regression line (blue) and LOESS 95% confidence bands (gray) of data from A) CvDurAlt and B) AvgLoudnessIndex features.

IV. DISCUSSION

This pilot study demonstrates a system that correctly classifies overall motor impairment of greater and lesser severity with an accuracy of 94.5% from a relatively small cohort of patients with PD. Our system includes software that guides participants to perform a set of kinematic tasks that use embedded hardware features of a smartphone and an algorithm that automates classification of motor impairment with machine learning techniques trained to clinical neurology ratings as ground truth. At this stage, the system is not able to classify the severity of bradykinesia. Further work will seek to determine whether poor BSS classification was a result of inadequate capture of some kinematic features, poor feature selection, limitations of the small sample size or other factors.

In contrast to several systems that place multiple sensors on the patient, our study uses a widely available consumer device. The iPhone application and analysis software developed for this project provide proof-of-principle demonstration that consumer smartphone devices have the potential to help assess motor symptoms of PD. Severity of resting tremor already has been successfully quantified based on data from accelerometers and gyroscopes within smartphones [5]. To the best of our knowledge, this is the first time that a smartphone touch-screen and microphone have been used for assessment of motor symptoms for PD. Next steps with this project will involve experiments to test the validity and reliability of kinematic measurements with the chosen hardware components, testing of longer trials over a larger cohort of patients, and inclusion of multiple blinded rater UPDRS assessments to extend applicability of results beyond a single neurology clinic. The development of a more consistent, mobile way of evaluating patients could facilitate more appropriate and personalized treatments.

REFERENCES

[1] L. DeLau et al., "Epidemiology of Parkinson's Disease," The Lancet Neurology, vol. 5, no. 6, pp. 525-535, June 2006.

[2] A.L.T. Tavares et al., "Quantitative measurements of alternating finger tapping in Parkinson's disease correlate with UPDRS motor disability and reveal the improvement in fine motor control from medication and deep brain stimulation," Mov. Disord., vol. 20, no. 10, pp. 1286-1298, Oct. 2005.

[3] Movement Disorders Society Task Force, "The Unified Parkinson's Disease Rating Scale (UPDRS): Status and recommendations," Mov. Disord., vol. 18, no. 7, pp. 738750, Jul. 2003.

[4] D.A. Heldman et al., "The modified bradykinesia rating scale for Parkinson's disease: reliability and comparison with kinematic measures," Mov. Disord., vol. 26, no. 10, pp. 1859-1863, Aug. 2011.

[5] N. Kostikis et al., "Towards remote evaluation of movement disorders via smartphones," IEEE EMBS, pp. 5240-5243, 2011.

[6] M.M. Koop et al., "Quantitative measures of fine motor, limb, and postural bradykinesia in very early stage, untreated Parkinson's disease," Mov. Disord., vol. 23, no. 9, pp. 1262-1268, Jul. 2008.

[7] J.W. Kim et al., "Quantification of bradykinesia during clinical finger taps using a gyrosensor in patients with Parkinson's disease," Med. Biol. Eng. Comput., vol 49., no. 3, pp. 365-371, Mar. 2011.

[8] T.O. Mera et al., "Feasibility of home-based automated Parkinson's disease motor assessment," J. Neurosci. Methods, vol. 203, no. 1, pp. 152-156, Jan. 2012.

[9] J.P. Guiffrida et al., "Clinically deployable Kinesia technology for automated tremor assessment," Mov. Disord., vol. 24, no. 5, pp. 723-730, Apr. 2009.