A Smartphone-based Tool for Assessing Parkinsonian Hand Tremor

N. Kostikis, *St. Member, IEEE*, D. Hristu-Varsakelis, *Senior Member, IEEE*, M. Arnaoutoglou, and C. Kotsavasiloglou

***Abstract*— The aim of this study is to propose a practical smartphone-based tool to accurately assess upper limb tremor in Parkinson's Disease (PD) patients. The tool uses signals from the phone's accelerometer and gyroscope (as the phone is held or mounted on a subject's hand) to compute a set of metrics which can be used to quantify a patient's tremor symptoms. In a small-scale clinical study with 25 PD patients and 20 age-matched healthy volunteers, we combined our metrics with machine learning techniques to correctly classify 82% of the patients and 90% of the healthy volunteers, which is high compared to similar studies. The proposed method could be effective in assisting physicians in the clinic, or to remotely evaluate the patient's condition and communicate the results to the physician. Our tool is low-cost, platform independent, non-invasive, and requires no expertise to use. It is also well-matched to the standard clinical examination for PD and can keep the patient "connected" to his physician on a daily basis. Finally, it can facilitate the creation of anonymous profiles for PD patients, aiding in further research on the effectiveness of medication or other overlooked aspects of patients' lives.**

***Index Terms*— Accelerometer, Bootstrap Aggregation Decision Tree, Gyroscope, Machine Learning, Parkinson's Disease, Smartphone, Tremor Quantification.**

I. Introduction

PARKINSONS' Disease (PD) is a chronic neurodegenerative disorder, affecting more than 1% of people over 55 and more than 3% of those over 75 years of age [1]. It is caused by low and falling dopamine levels. Its symptoms are extensively documented, including tremor, bradykinesia, rigidity, postural instability and impaired cognitive function [2], however the reason why the dopaminergic neurons die (resulting in abnormally low dopamine levels) is unknown. Thus, current treatments only focus on alleviating the symptoms and improving the patients' daily life.

The typical clinical examination of a PD patient involves a standardized procedure where the physician evaluates the performance of the patients on a scale of 0-4, observing them in postures and tasks described in the Unified Parkinson's Disease Rating Scale (UPDRS) [3]. When motor impairment is present, physicians may use the results of Electromyography (EMG) or imaging scans such as Positron Emission Tomography (PET) and Single-Photon Emission Computed Tomography (SPECT) to exclude other reasons for the symptoms before diagnosing PD, but these techniques do not constitute PD assessment tools per se.

Tremor is not the main quality of life constraining-aspect of the disease [4], but responds well to dopaminergic medication; as such, its objective quantification can provide useful feedback on the efficacy of the treatment regimen. Because of that, researchers have turned to platforms involving wearable accelerometers to help them assess PD tremor, with relative success [5],[6],[7],[8],[9].

When it comes to PD, it is important to be able to assess patients accurately and frequently in order to adjust their treatment as necessary when variations in the severity of the symptoms occur. Unfortunately, in addition to the time and costs involved, it is difficult for most patients (even more so for those in rural areas), to be monitored frequently by a specialized physician. This creates the need for tools that can aid the physician by assessing a patient's condition remotely, and that are practical and easy to use.

In the last few years, researchers have explored the ability of smartphones to quantify PD tremor accurately [10],[11],[12]. These devices are now ubiquitous with a suite of on-board sensors (including accelerometers) and wireless connectivity. They are also easy to use compared with most wearable accelerometer solutions previously developed. As such, they have been making their way into medical research aimed at developing mobile tools for aiding the physician [13].

Although scientifically important, none of the accelerometer- or smartphone-based methods proposed to date have made their way into mainstream clinical practice. The last decade there have been efforts to create platforms incorporating inertial measurement sensors into scalable body sensor networks, with wireless connectivity and real-time on-node signal processing. Approaches like SHIMMER, Kinesia and Xsens do provide researchers with valuable information, however they have high complexity in terms of hardware and software, high cost, and lack standardized, approved and widely accepted protocols. Additionally, they can hardly be considered ubiquitous.

The goal of this work is to investigate the use of a smartphone-based tool for assessing PD induced hand tremor. Our approach involves using the phone's embedded

N. Kostikis is with the Department of Applied Informatics, University of Macedonia, Thessaloniki, 54006, Greece (corresponding author; telephone: +306932588797; fax: +302310891290; nkostikis@uom.gr).

D. Hristu-Varsakelis is with the Department of Applied Informatics, University of Macedonia, Thessaloniki, 54006, Greece (dcv@uom.gr).

M. Arnaoutoglou, M.D., is with the School of Medicine, Aristotle University of Thessaloniki, 54124, Greece (marnaout@med.auth.gr).

C. Kotsavasiloglou, M.D., Thessaloniki, Greece (chkot@cteam.gr).

accelerometer and gyroscope sensors to quantify PD hand tremor. Operationally, the patient simply visits a web site [14] and takes up simple postures much like those used in standardized clinical exams, with the smartphone mounted on their hand. The data thus recorded can then be used to classify a subject as healthy or not, and to track the severity of the tremor in PD patients.

Preliminary versions of this work can be found in [15] and [16], where we presented proof-of-concept results for the tool discussed here. This paper substantially differs from and extends our previous work by i) including additional samples from healthy subjects which are *age-matched* to the PD patients used in our study, ii) including data on a small sample of patients off medication in order to quantitatively track the severity of their hand tremor, iii) exploring the correlation between our quantitative metrics results and the patients' (subjective) clinical examination for all supported postures, and iv) using a machine learning feature classification approach to choose those metrics which are better at distinguishing between pathological and healthy signals, thus increase our method's accuracy. The accuracy in separating healthy from PD subjects attained in this work is on par with other works using smartphones' accelerometers and short-duration data [12]. It is also very close to the sensitivity and specificity achieved in [17], where the authors used sensors of the SHIMMER platform to perform mobile gait analysis. There are works that achieve near 100% accuracy but do so using day-long signals [18] which may not be practical in our setting. High accuracy (98.5% sensitivity, 97.5% specificity) is also achieved using smartphone accelerometer-based gait analysis, with a combination of tests and metrics [19].

The proposed method has been implemented in the form of an app and is available for use on any smartphone with iOS or Android installed, without the need for any downloads or memory-consuming installations [14]. It can be offered as a web-service, so that developers can build their own applications around it and extend its functionality. Our approach does not require the presence of an expert or any kind of special equipment to conduct measurements. It transmits data in real time via TCP/IP, connecting the patient to his physician with no delays, and can benefit the research community by providing anonymized information on PD hand tremor profiles.

The remainder of this paper is structured as follows. In Section II we describe our experimental setup, the subjects used in our study and the metrics calculated based on the signal(s) collected from each subject. Section III contains the main results, including statistical analyses of the subjects' scores under each metric, correlation with the subjects' clinical "picture", and a machine-learning-based discrimination scheme for separating healthy from PD subjects. Section IV summarizes our findings and discusses their implications for medical practice, as well as future work.

II. Experimental Setup

A. Volunteers

We recruited twenty-five PD patients from the outpatient clinic of the 1st Department of Neurology at the Aristotle University of Thessaloniki. They all agreed to participate after they were offered a detailed explanation of the study's procedure and goals. All of them were right-handed, under L-DOPA treatment and suffering from PD for more than two years. During the study, two of them were hospitalized overnight so that they could be tested in the morning before they received their medication, to approximate de novo PD patients. Those two will be referred as our PDDN group, while the PD group comprises the other twenty-three (see Table I for patients' information).

TABLE I
Information on the PD and PDDN Subjects

Groups	*Num*	*Age*	*Sex*	*UPDRS Upper Limbs Tremor Components*			
				Rest Right	*Rest Left*	*Extended Right*	*Extended Left*
PD	*1*	76	F	0	1	0	1
	2	77	M	0	0	0	1
	3	80	F	2	0	0	0
	4	76	M	1	0	1	0
	5	74	M	0	0	0	1
	6	76	M	0	2	0	0
	7	62	F	0	1	1	1
	8	82	M	0	0	1	1
	9	69	F	0	0	0	1
	10	76	F	1	2	1	2
	11	81	M	1	0	2	1
	12	39	M	2	1	2	2
	13	65	F	0	2	0	1
	14	78	M	2	0	1	1
	15	50	M	1	0	0	0
	16	75	F	0	0	0	1
	17	70	M	0	0	1	2
	18	80	F	4	4	2	1
	19	76	M	0	0	0	0
	20	75	F	2	0	2	1
	21	43	F	0	0	0	0
	22	78	F	1	0	2	1
	23	73	F	0	0	0	0
*PDDN**	*A*	83	M	1	0	0	0
	B	77	F	0	1	2	1

*These patients' UPDRS scores are the ones evaluated while the patients were on medication.

The control group for the study, labeled as group H, contains twenty healthy volunteers, none of whom suffered from a movement disorder, hypertension or diabetes. They were screened for several health conditions which could exclude them from the study, such as hypertension or any movement disorder. They were also notified of the procedure and the purpose of the study before agreeing to participate. Grouping information on all participants of the study is provided in Table II.

TABLE II
Information on Subjects' Sex, Age and Grouping

Group	*Group size*	*Sex Statistics*		*Age Statistics*	
		Females	*Males*	*Mean±SD*	*SE*
H	20	10	10	67.20±6.25	1.39
PD	23	12	11	70.91±11.78	2.45
PDDN	2	1	1	80.00±4.24	3.00
Total	45	23	22		

The ages of the two main groups were mean-tested with the non-parametric Mann-Whitney test and were found not to be statistically different at the 1% significance level, therefore the groups can be considered age-matched.

B. Hardware and Software

The UPDRS scores of the PD volunteers were assessed by the same physician (our third author), just before data collection. We attached an iPhone on our volunteers' hands using the same custom-made mounting glove (Fig. 1) from [15] and [16]. It consists of a perforated case into which the phone "locks", and a wrist-supporting glove, both commercially available. The glove fits tightly on the volunteer's hand and the case is tightly sewn on the glove using non-elastic thread, ensuring the stability of the device on top of the hand. With the device attached, each participant had to maintain each of two prescribed postures for 30 seconds, while acceleration and gyroscope data was recorded by the phone. The two postures we used were the same ones used during the clinical evaluation: a) "Extended", i.e., seated with both hands extended in front of the torso (Postural Tremor of the Hands, component 3.15 of the MDS-UPDRS) and b) "Rest", i.e., seated with both hands placed on the arms of the chair (Rest Hand Tremor, component 3.17 of the MDS-UPDRS). The procedure was then repeated for the subject's other hand, in the same two postures. In the following, we will specify the combination of a patient's hand (Right of Left upper extremity) during each position as rR, rL, eR, and eL for rest-right, rest-left, extended-right, and extended-left, respectively.

Fig. 1. Photo of the custom made glove-case with an iPhone mounted on.

The hardware setup was the same as the one used in our earlier work, [15] and [16]:

1. An iPhone 4S with the latest iOS, with Internet access enabled, and screen orientation locked in vertical,
2. A web application to collect data from the smartphone's sensors,
3. A web server to host the site and store the signals, and
4. A MATLAB application for processing the signals received at the server.

As per the protocol described earlier, the volunteers were asked to maintain certain postures for 30 seconds, during which the application automatically collected the accelerometer and gyroscope data and sent them to our server. We then used the data to extract features which quantify and characterize the subjects' tremor levels. The signals were sampled at 20Hz, which is sufficient to identify events occurring at 9Hz or less [20], such as PD-induced tremor.

Our web application, being written in PHP and JavaScript, is entirely independent of the client's hardware or software platform. It only demands basic prerequisites such as an embedded accelerometer and gyroscope and one of the most popular smartphone operating systems, iOS or Android. We successfully tested it on a Samsung Galaxy S4 and a Google Nexus 5, both running Android 4.4.2.

C. Signals and Metrics

For each volunteer's session we obtained two signals from the phone's sensors, the acceleration vector $\alpha(i) = [\alpha_x(i), \alpha_y(i), \alpha_z(i)]^T$ (in m/s^2) and the rotational velocity vector $\omega(i) = [\omega_x(i), \omega_y(i), \omega_z(i)]^T$ (in *deg/s*), with i denoting discrete time. The rotational velocity in practice should be more information-rich because it is constructed using both accelerometer and gyro data and is expected to capture more of the characteristics of the tremor.

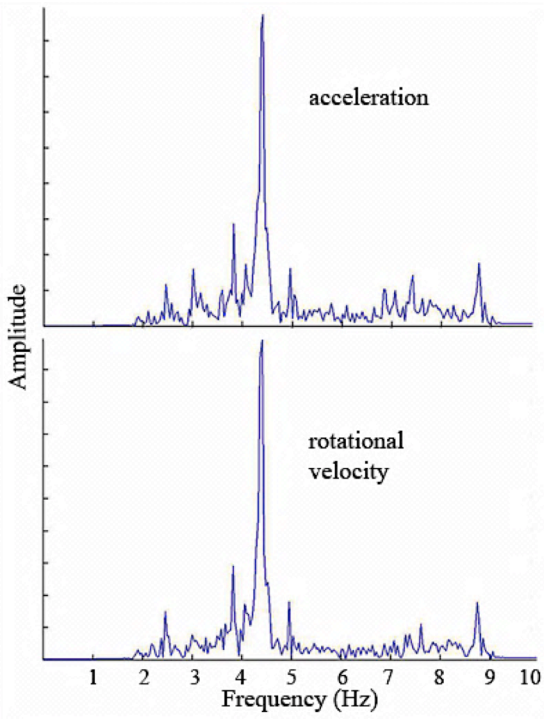

Fig. 2. Y-axis acceleration ($\alpha_y(i)$) and rotational velocity ($\omega_y(i)$) spectral analyses for a PD patient 3 from Table I. These are signals from a Rest Right recording.

We applied a band-pass filter with cutoff frequencies of 1.5Hz and 9.5Hz, in order to exclude noise due to breathing, pulse, or any high-frequency sudden movements during the recordings. The spectral analysis of $\alpha(i)$ and $\omega(i)$ of a PD volunteer with typical Parkinsonian tremor is shown in figure 2. As expected, her acceleration and rotational velocity signal amplitude peaks at about 3-5Hz, which is consistent with the literature [20].

We used the acceleration and rotational velocity signals as in [16], to compute the following four metrics for each session:

$$mag_\alpha = \sum_1^N \|\alpha(i)\|^2 \quad (1)$$

$$mag_\omega = \sum_1^N \|\omega(i)\|^2 \,, \quad (2)$$

$$sd_\alpha = \sum_{i=1}^{N-1} \sum_{\kappa\in\{x,y,z\}} |\alpha_\kappa(i) - \alpha_\kappa(i+1)| \,, \quad (3)$$

$$mAmp_\omega = \sum_{\kappa=\{x,y,z\}} \max_{4\leqq\xi\leqq7} \widehat{\omega_\kappa}(\xi) \,, \quad (4)$$

where: mag_α and mag_ω are the sums of squared magnitudes of the acceleration, and the rotation rate vector respectively,

and sd_α, is the sum of absolute differences in the acceleration vector, summed over each of the three axes, x, y, and z. To compute the $mAmp_\omega$ metric (4) we initially obtained the magnitude of the Fourier transform of each of the three axial components of the rotation vector $\omega(i)$, defined as $\widehat{\omega_\kappa}(\xi)$, $\kappa \in \{x, y, z\}$. We then determined each component's maximum in the $4 \leqq \xi \leqq 7\ Hz$ range (that range being consistent with the frequency of Parkinsonian tremor) and summed the three maxima [16].

III. Results

A. Means Testing

Since our goal is to facilitate monitoring and diagnosis of PD-induced tremor, it is essential to establish that the metrics described in the previous section can be useful in differentiating the PD vs H populations. We used the non-parametric Mann-Whitney test to establish that the two populations have statistically different means in all four metrics on all four postures, rR, rL, eR, eL. As shown in Table III, all between-groups tests found significant differences between the mean scores of the metrics of the H volunteers compared to the mean of the PD volunteers. This suggests that the two populations (H and PD) have statistically different scores under every one of the signal metrics computed, and one may attempt to differentiate H vs PD subjects based on one or more of those metrics.

TABLE III
METRICS MEANS TESTING BETWEEN THE H AND PD GROUPS

	Metric	H mean vs PD mean statistical difference	
		Decision	p
Rest Right	mag_α	Yes	0,0018
	mag_ω	Yes	<0.001
	sd_α	Yes	0.0014
	$mAmp_\omega$	Yes	<0.001
Rest Left	mag_α	Yes	<0.001
	mag_ω	Yes	<0.001
	sd_α	Yes	<0.001
	$mAmp_\omega$	Yes	<0.001
Extended Right	mag_α	Yes	<0.001
	mag_ω	Yes	<0.001
	sd_α	Yes	<0.001
	$mAmp_\omega$	Yes	<0.001
Extended Left	mag_α	Yes	<0.001
	mag_ω	Yes	<0.001
	sd_α	Yes	<0.001
	$mAmp_\omega$	Yes	<0.001
Sum of all metrics over all positions and sides	mag_α	Yes	<0.001
	mag_ω	Yes	<0.001
	sd_α	Yes	<0.001
	$mAmp_\omega$	Yes	<0.001
Absolute differences of summed right hand vs left hand scores	mag_α	Yes	<0.001
	mag_ω	Yes	0.0012
	sd_α	Yes	<0.001
	$mAmp_\omega$	Yes	<0.001

Means testing was applied to the volunteers' metrics. Numerical values indicate p-values for the Mann-Whitney test at the 1% significance level.

There was no statistical difference in the subjects' left vs. right mean scores *within* each group (see Table IX in Appendix). It is typical for PD patients to manifest the disease's symptoms with some laterality, i.e., to a greater degree on one side, right or left. That is indeed the case with our PD volunteers because 19 of the 23 have differences between the sums of the UPDRS components concerning right vs left hand tremor indicating laterality of motor impairment. Although clinically observable, the Mann-Whitney test for the summed UPDRS scores for right vs left hands yields no statistical difference, with p=0.7327. In order to identify the laterality statistically, for each metric we summed the scores of both postures for right and left hand separately and calculated the absolute differences between them. As shown in the last four rows of Table III, for each metric, the absolute differences between hands for the PD group is statistically different from those of the H group. That means that the amount of difference between hands is not the same for PD and H, presumably due to the disease's laterality.

B. Correlation with Clinical Examination

In previous work [16] we attempted to establish the validity of our smartphone-based method of upper limb parkinsonian tremor quantification by running a Pearson product-moment correlation analysis between the UPDRS scores of the PD volunteers and their respective signal metrics. Table IV contains the results of the correlation analysis for the Rest posture scores. The numbers are slightly different from our previous study because the signals are now band pass filtered, as previously explained.

TABLE IV
CORRELATION COEFFICIENTS BETWEEN EACH METRIC AND THE UPDRS SCORES – REST POSTURE

Metrics	Corr. Coefficients			
	Right Hand		Left Hand	
	r	p	r	p
mag_α	0.70	<0.001	0.70	<0.001
mag_ω	0.69	<0.001	0.79	0
sd_α	0.77	0	0.87	0
$mAmp_\omega$	0.75	0	0.85	0

In [16] we used the Rest posture data only, whereas here we were also interested in the Extension posture data for both hands. The correlation analysis of the Extension posture data (Table V) yields low coefficients (r<0.6) with low confidence (p>0.01). The results are better for the right hand but generally do not suggest good correlation between the UPDRS scores and the smartphone metrics.

TABLE V
CORRELATION COEFFICIENTS BETWEEN EACH METRIC AND THE UPDRS SCORES – EXTENDED POSTURE

Metrics	Coefficients			
	Right Hand		Left Hand	
	r	p	r	p
mag_α	0.46	0.02	0.4	0.057
mag_ω	0.53	0.008	0.38	0.06
sd_α	0.63	0.001	0.42	0.04
$mAmp_\omega$	0.58	0.003	0.36	0.08

These findings show a connection between the manifestation of the action hand tremor and the hardware experimental setup. The fact that the resting tremor is identified consistently, whereas in the extended posture the measured tremor correlates weakly with the clinical assessment, is probably related to the effect of the mass of the smartphone on the dynamics of the hand/arm system.

C. Trials on Patients Off Medication

The 23 volunteers of the PD group who underwent the smartphone-based tremor measuring procedure were under

medication, but the timespan from their last dose of L-DOPA was anywhere from 1 to 4 hours when they were tested. That means that some of them were at the peak of the drug's effect while for others this was not the case. With an eye towards tracking the progression of PD, we wanted to see how our metrics would "react" to an alteration in a patient's condition, such as that brought on by medication intake. We observed two PD volunteers, referred to as the PDDN group, who stayed in the clinic overnight and followed our experimental protocol both "off" and "on" medication (i.e., right before taking their medication in the morning and one hour after that). In Table VI we present the percent differences "on"-"off" in the four metrics, along with the UPDRS scores ("off" and "on") for both PDDN subjects. We would expect those differences to be negative because we expect higher metric scores (more pronounced tremor) while off-medication or in a de novo state, and lower after the drug ingestion ("on").

From the UPDRS scores of the two volunteers it is clear that these two patients did not suffer from severe hand tremor.

TABLE VI
UPDRS SCORES AND PERCENT DIFFERENCES IN SIGNAL METRICS ("OFF" VS "ON" MEDICATION) FOR TWO SUBJECTS (PDDN)

Metric	*PDDN Subject*	*Rest*				*Extended*			
		Right		*Left*		*Right*		*Left*	
		Off	*On*	*Off*	*On*	*Off*	*On*	*Off*	*On*
UPDRS	*A*	2	1	0	0	1	0	1	0
	B	1	0	1	1	2	2	1	1
$\%\Delta mag_\alpha$	*A*	-4%		-20%		-58%		-31%	
	B	-22%		-29%		6%		-25%	
$\%\Delta mag_\omega$	*A*	-40%		-35%		-85%		-55%	
	B	-75%		-64%		5%		-38%	
$\%\Delta sd_\alpha$	*A*	-10%		-23%		-43%		-37%	
	B	-26%		-32%		5%		-14%	
$\%\Delta mAmp_\omega$	*A*	-31%		-23%		-58%		-48%	
	B	-74%		-52%		1%		-21%	

Their physician observed that the medication improved mostly the patients' rigidity (which is not measured by our tool) and less so their tremor. However, it is encouraging to note that the readings of the smartphone's sensors respond well and follow the expected negative trend of the changes in the UPDRS scores in the "on" state. The only discrepancy is observed in the eR position of volunteer *B* for all metrics, however for that position there were also no observed clinical changes in the UPDRS "on"-"off" as well.

D. Supervised Machine Learning To Establish Discriminating Criteria

To arrive at a practical and ubiquitous tool which will be able to accurately assess upper limb tremor in PD patients, it is essential to be able to distinguish correctly between healthy vs PD signals. In [15] we used short signals and simpler metrics but were able to classify accurately 9 out of 10 PD volunteers. In this study we have a larger group of volunteers, signals of greater duration and four metrics which can be used as classification features.

1) Fit of the Metrics as Discriminating Criteria

In Table VII we present the results of an ROC curve fitting analysis between the H and PD groups. For each metric and posture, the percentages of the confusion matrix (True Positive as TP and True Negative as TN), the Area Under the Curve (AUC) value and the cutoff point are recorded. The percentages suggest that no one metric or posture alone can be used with confidence as the sole discriminating criterion in the current dataset because the highest sensitivity (TP) is 87% and the highest specificity (TN) is 95% but they do not appear concurrently. In fact, the metrics which are based on the gyroscope signal, i.e., mag_ω and $mAmp_\omega$, rank higher in combined sensitivity and specificity, whereas the acceleration magnitude ranks high in specificity.

TABLE VII
TP, TN, AUC, AND CUTOFF VALUES CALCULATED FROM AN ROC ANALYSIS FOR ALL METRICS AND POSTURES BETWEEN H AND PD GROUPS

Metric	*Rest*				*Extended*			
	Right		*Left*		*Right*		*Left*	
	TP	*TN*	*TP*	*TN*	*TP*	*TN*	*TP*	*TN*
	AUC		*AUC*		*AUC*		*AUC*	
	CutOff		*CutOff*		*CutOff*		*CutOff*	
mag_α	65%	85%	69%	80%	65%	90%	65%	95%
	0.7804		0.8413		0.8217		0.8217	
	1.8989		1.3418		15.02		13.31	
mag_ω	73%	85%	82%	85%	65%	85%	78%	75%
	0.8435		0.8935		0.7978		0.8435	
	213.96		166.17		2622.21		2175.75	
sd_α	73%	75%	65%	80%	82%	70%	65%	85%
	0.7870		0.8217		0.8348		0.8022	
	58.88		55.34		128.80		149.32	
$mAmp_\omega$	87%	85%	87%	85%	82%	80%	82%	85%
	0.9065		0.9022		0.8717		0.8957	
	0.2366		0.2393		0.7740		0.85	

2) Feature Selection for a Classification Model

To investigate our metrics' potential as classifiers on new data, we decided to use a non-parametric decision tree-based ensemble machine learning algorithm for multivariate classification, as per [19], [21], [22]. All four metrics (1)-(4) for each posture were used, resulting in a set of 16 classification features of a bootstrap aggregation for a random forest of decision trees (BagDT). We used MATLAB's TreeBagger class for the construction of the ensemble of trees and its method oobPredict to calculate the trained model's prediction accuracy. The ensemble had 500 grown trees.

To overcome the problem of possible over-fitting we used each feature's out-of-bag error estimate, as defined in [22]. That is a measure of importance calculated by the TreeBagger class, defined as the increase in prediction error if the variable's values are permuted across the out-of-bag

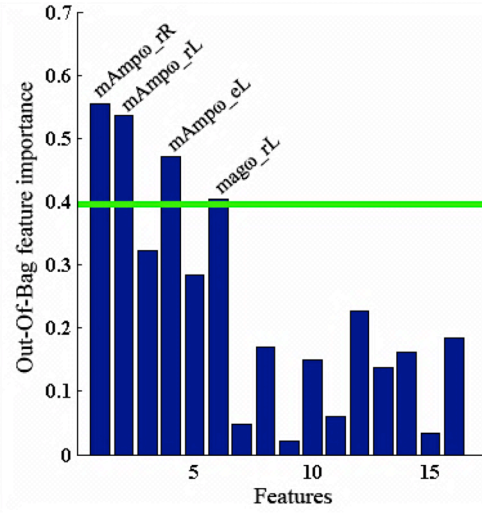

Fig. 3. Out-of-bag feature importance for 16 features comprising each metric for each posture. Four of them score significantly higher than all others (importance>0.4).

observations. The higher the increase, the higher the importance of the given feature (see Fig. 3). We also used three other approaches to evaluate the importance of our model's features, namely Information Gain, One-R and Chi-Squared, as described in [23], all tested with 10-fold stratified cross-validation (see Table X in Appendix). What we found, after multiple runs with various feature subsets is that the model with the best classification potential, i.e., highest average accuracy and lowest dimensionality, uses the four most important features as calculated by the Out-Of-Bag Error and the One-R approach, namely $mAmp_{\omega}$ for rR, rL and eL, and mag_{ω} for rL.

The fact that three out of the four most relevant features involve the volunteers' left hand could be related to the fact that all of them were right-handed. Data from left-handed volunteers should also be collected to test whether the dexterity of the dominant hand plays a significant role in that regard.

3) The Classification Model

We used MATLAB's TreeBagger class, which constructs a bagged ensemble of decision trees (BagDT), i.e., a Breiman's random forest [22], to build a classification model that would accurately identify unknown data. Our preference for that approach was guided by its good performance relative to a total of six machine learning algorithms tested, and whose performance in terms of sensitivity (TP), specificity (TN) and AUC is shown in Table VIII. All classifiers used the same subset of four features deemed to be most important by the feature selection procedure described above. Apart from BagDT, the algorithms were tested with 10-fold stratified cross-validation.

The accuracy of the bagged ensemble of decision trees was better than any of the other classifiers, as it classified correctly 90% of the H and 82% of the PD volunteers, with AUC=0.9435 (Fig 4).

TABLE VIII
PERFORMANCE OF SIX DIFFERENT MACHINE LEARNING APPROACHES

Algorithms	*Sensitivity (TP)*	*Specificity (TN)*	*AUC*
Naïve Bayes	56%	100%	0.80
Logistic Regression	74%	100%	0.88
SVM	56%	100%	0.78
AdaBoost	83%	85%	0.91
C4.5	83%	75%	0.88
BagDT	82%	90%	0.94

IV. DISCUSSION AND CONCLUSIONS

In this work we proposed a method to quantify PD induced hand tremor based on a smartphone mobile platform. Using an ensemble of decision trees as a machine learning model we were able to identify positively 82% of PD volunteers and 90% of the H volunteers. Although the clinical trial was relatively small, it is larger than others on the same topic, such as [5], [7], [10], [12], [19], it offers encouraging results and serves as proof of concept; the statistical analyses we conducted show that our proposed method for quantifying upper limb Parkinsonian tremor can be used both in a clinical setting to facilitate the physician's work by offering her an accurate assessment tool, and at home by the patients themselves to self-monitor their progress.

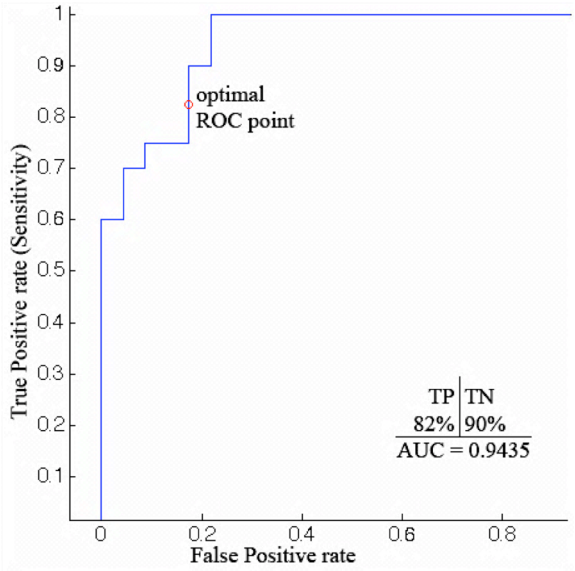

Fig. 4. ROC curve for the classification from the minimized BagDT model.

Our web application is readily available [14], with a more advanced version underway. We plan to extend our work in four directions: additional samples, data security, signal processing and user/data management. Collecting samples from left-handed volunteers will allow us to examine the influence of the volunteers' dominant hand dexterity on the feature selection procedure and enlarge our overall sample. In addition, we will be applying cryptography to all communications from the smartphone to our servers as the signals are considered sensitive data. We plan to explore the use of new metrics and keep training our model so that we improve its sensitivity and specificity. Finally, we intend to create a scalar backend application which will accept subscriptions from physicians and individual patients who wish to use our tool, and provide them with personalized information on their profiles.

Implications for Medical Practice and Research

The availability of a ubiquitous assessment tool of an otherwise subjectively rated symptom, such as hand tremor, primarily assists the physician in planning an effective treatment regimen. However, the ability to easily measure hand tremor daily at home by means of the method proposed here and storing the relevant metrics in order to later track and visualize the progression of the disease and the effectiveness of the medication is more than welcome as a benefit for the patient himself. Finally, biobanking large data sets with motor information collected according to a specific protocol is something the scientific community could harvest and conduct valuable research on. Apple's new framework, Research Kit is based on the same principle [24]. Our motion monitoring apparatus, combined with signal processing and machine learning can detect motor abnormalities in a non-invasive manner and assess the diagnosis and long-term quality of life of PD patients.

V. ETHICAL STANDARDS AND CONFLICT OF INTEREST STATEMENT

This study was performed by approval of the appropriate ethics committee and in accordance with the ethical standards

laid down in the 1964 Declaration of Helsinki and its later amendments.

The authors also declare that they have no conflict of interest.

APPENDIX

TABLE IX
METRICS MEANS TESTING WITHIN THE H AND PD GROUPS

	Signal	*Metric*	*Right mean vs Left mean statistical difference*	
			Decision	*p*
H	*Rest*	mag_α	No	0.1806
		mag_ω	No	0.4094
		sd_α	No	0.1806
		$mAmp_\omega$	No	0.3369
	Extended	mag_α	No	0.7557
		mag_ω	No	0.9461
		sd_α	No	0.8604
		$mAmp_\omega$	No	0.9461
	Summed	mag_α	No	0.6554
		mag_ω	No	0.9461
		sd_α	No	0.7353
		$mAmp_\omega$	No	0.9461
PD	*Rest*	mag_α	No	0.5098
		mag_ω	No	0.7417
		sd_α	No	0.3795
		$mAmp_\omega$	No	0.7417
	Extended	mag_α	No	0.9475
		mag_ω	No	0.6764
		sd_α	No	0.9125
		$mAmp_\omega$	No	0.8090
	Summed	mag_α	No	0.8433
		mag_ω	No	0.6134
		sd_α	No	0.6445
		$mAmp_\omega$	No	0.7584

Means testing was applied to the volunteers' metrics. Numerical values indicate p-values for the Mann-Whittney test at the 1% significance level.

TABLE X
FOUR MOST IMPORTANT FEATURES OUT OF SIXTEEN ACCORDING TO THREE ATTRIBUTE SELECTION APPROACHES

Information Gain		*One-R*		*Chi-Squared*	
Merit	*Features*	*Merit*	*Features*	*Merit*	*Features*
0.48	$mAmp_\omega$_eL	81.39	$mAmp_\omega$_rL	20.04	$mAmp_\omega$_eL
0.42	$mAmp_\omega$_rR	80.11	mag_ω_rL	19.99	$mAmp_\omega$_rL
0.42	$mAmp_\omega$_rL	79.33	$mAmp_\omega$_eL	19.91	$mAmp_\omega$_rR
0.41	$mAmp_\omega$_eR	78.81	$mAmp_\omega$_rR	18.85	$mAmp_\omega$_eR

For the Information Gain, the One-R and the Chi-Squared approach the merit is a correlation-based metric as described in [25]. Higher values indicate more important features.

REFERENCES

[1] S. K. Van Den Eeden, , C. M. Tanner, A.L. Bernstein, R. D. Fross, A. Leimpeter, D. A. Bloch, and L. M. Nelson, "Incidence of Parkinson's disease: variation by age, gender, and race/ethnicity," *American Journal of Epidemiology,* no. 157, pp. 1015-1022, 2003.

[2] E. R. Kandel, J. H. Schwartz, and T. M. Jessell, *Principles of neural science*. McGraw-Hill, New York, 2000.

[3] C. G. Goetz, S. Fahn, P. Martinez-Martin, W. Poewe, C. Sampaio, G. T. Stebbins, M. B. Stern, B. C. Tilley, R. Dodel, B. Dubois, R. Holloway, J. Jankovic, J. Kulisevsky, A. E. Lang, A. Lees, S. Leurgans, P. A. LeWitt, D. Nyenhuis, C. W. Olanow, O. Rascol, A. Schrag, J. A. Teresi, J. J. Van Hilten, and N. LaPelle, "Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): Process, format, and clinimetric testing plan," *Movement Disorders,* vol. 22, no. 1, pp. 41–47, Jan. 2007.

[4] T. Ellis, J. T. Cavanaugh, G. M Earhart, M. P. Ford, K. B. Foreman, and L.E. Dibble, "Which measures of physical function and motor impairment best predict quality of life in Parkinson's disease?," *Parkinsonism & related disorders*, vol. 17, pp. 693-697, 2011.

[5] S. Patel, K. Lorincz, R. Hughes, N. Huggins, J. Growdon, D. Standaert, M. Akay, J. Dy, M. Welsh, and P. Bonato, "Monitoring motor fluctuations in patients with Parkinson's Disease using wearable sensors," *Transactions on Information Technology in Biomedicine, IEEE*, vol. 13, no. 6, pp. 864-873, 2009.

[6] K. Lorincz, B. R. Chen, G. W. Challen, At. R. Chowdhury, S. Patel, P. Bonato, and M. Welsh, "Mercury: A Wearable Sensor Network Platform for High-fidelity Motion Analysis," in *Proc. 7th ACM Conf. Embedded Networked Sensor Systems (SenSys'09),* November 2009.

[7] A. Weiss, S. Sharifi, M. Plotnik, J. P. Van Vugt, N. Giladi, and J. M. Hausdorff, "Toward automated, at-home assessment of mobility among patients with Parkinson disease, using a body-worn accelerometer," *Neuro-rehabilitation and Neural Repair*, vol. 25, pp. 810-818, 2011.

[8] R. LeMoyne, T. Mastroianni, and W. Grundfest, "Wireless accelerometer configuration for monitoring Parkinson's disease hand tremor," *Advances in Parkinson's Disease*, vol.2, no.2, pp. 62-67, 2013.

[9] J. Klucken, J. Barth, P. Kugler, J. Schlachetzki, T. Henze, F. Marxreiter, Z. Kohl, R. Steidl, J. Hornegger, B. Eskofier, and J. Winkler "Unbiased and Mobile Gait Analysis Detects Motor Impairment in Parkinson's Disease," *PLoS ONE*, vol. 8, issue 2, 2013.

[10] R. LeMoyne, T. Mastroianni, M. Cozza, C. Coroian, and W. Grundfest, "Implementation of an iPhone for characterizing Parkinson's disease tremor through a wireless accelerometer application," in *Proc. 32nd Int'l Conf. of the IEEE Engineering in Medicine and Biology Society*, Buenos Aires, 2010.

[11] R. Delano, B. Parise, and L. West, "iTrem", in *Proc. Int'l Conf. on Health Informatics,* Rome, pp. 26-29, 2011.

[12] J. F. Daneault, B. Carignan, C. É. Codère, A. F. Sadikot, and C. Duval, "Using a smart phone as a standalone platform for detection and monitoring of pathological tremors," *Frontiers in Human Neuroscience*, vol. 6, p. 357, 2013.

[13] E. Ozdalga, A. Ozdalga, and N. Ahuja, "The Smartphone in Medicine: A Review of Current and Potential Use Among Physicians and Students," *J Med Internet Res*, vol 14, no 5, 2012.

[14] http://afroditi.uom.gr/itremorsense

[15] N. Kostikis, D. Hristu-Varsakelis, M. Arnaoutoglou, C. Kotsavasiloglou, and S. Baloyiannis, "Towards remote evaluation of movement disorders via smartphones," in *Proc. 33rd Int'l Conf. of the IEEE Engineering in Medicine and Biology Society,* Boston, 2011, pp. 5240–5243.

[16] N. Kostikis, D. Hristu-Varsakelis, M. Arnaoutoglou, and C. Kotsavasiloglou, "Smartphone-based evaluation of parkinsonian hand tremor: quantitative measurements vs clinical assessment scores," in *Proc. 36th Int'l Conf. of the IEEE Engineering in Medicine and Biology Society,* Chicago, 2014, pp. 906–909.

[17] J. Barth, J. Klucken, P. Kugler, T. Kammerer, R. Steidl, J. Winkler, J. Hornegger, and B. Eskofier, "Biometric and Mobile Gait Analysis for Early Diagnosis and Therapy Monitoring in Parkinson's Disease," in *Proc. 33rd Int'l Conf. of the IEEE Engineering in Medicine and Biology Society,* Boston, 2011, pp. 868-871.

[18] M. Brunato, R. Battiti, D. Pruitt, and E. Sar, "Supervised and unsupervised machine learning for the detection, monitoring and management of Parkinson's disease from passive mobile phone data," *Submission for the M.J. Fox Foundation Competition: Predicting Parkinson's Disease Progression with Smartphone Data*, 2013, available: https://kaggle2.blob.core.windows.net/prospector-files/1117/958625cf-3514-4e64-b0e7-13ebd3cf9791/kaggle.pdf

[19] S. Arora, V. Venkataraman, S. Donohue, K.M. Biglan, E.R. Dorsey, and M.A. Little, "High accuracy discrimination of Parkinson's disease participants from healthy controls using smartphones," in *Proc. 2014 IEEE International Conference on Acoustics, Speech and Signal Processing*, Florence, Italy, 2014.

[20] S. Fahn, "Classification of movement disorders", *Mov Disord,* vol. 26, pp. 947-957, 2011.

[21] X. Wu · V. Kumar, J. R. Quinlan J. Ghosh Q.Yang H. Motoda G. J. McLachlan, A. Ng, B. Liu, P. S. Yu, Z. H. Zhou, M. Steinbach, D. J. Hand, and D. Steinberg, "Top 10 algorithms in data mining", *Knowl Inf Syst*, vol. 14, pp. 1-37, 2007.

[22] L. Breiman, "Random forests," *Machine learning*, vol. 45, no.1, pp. 5-32, 2001.

[23] J. Novakovic, "The impact of feature selection on the accuracy of Naïve Bayes classifier," *18th Telecommunications forum TELFOR* 2010, Belgrade, 2010.

[24] Apple - ResearchKit, https://www.apple.com/researchkit/

[25] M. A. Hall, "Correlation-based Feature Selection for Machine Learning," Ph.D. thesis, Dept. Comp. Sc., Univ. of Waikato, Hamilton, New Zealand, 1999.