Multimodal Detection of Parkinson Disease based on Vocal and Improved Spiral Test

Hung N. Pham[1], Binh P. Nguyen[2], Chan Yi Jie Kelvin[3], Gopa Sen[3],
Han Yuen Kwang Andy[3], Lim Pier[3], Teresa Cheng Siew Loon[3], Quang H. Nguyen[1] and Matthew C. H. Chua[3]
[1]School of Information and Communication Technology, Hanoi University of Science and Technology, Hanoi, Vietnam
[2]School of Mathematics and Statistics, Victoria University of Wellington, Wellington, New Zealand
[3]Insitute of System Science, National University of Singapore, Singapore, Singapore
Email: [1]hungpn@soict.hust.edu.vn, [2]b.nguyen@vuw.ac.nz, [3]mattchua@nus.edu.sg

***Abstract*—Previous studies have used data related to either voice or spiral drawing to detect Parkinson disease (PD). However, different people experience different symptoms and different levels of severity of PD. Hence, in this study, we propose a multimodal approach combining voice and image test to enhance the reliability of detecting PD patients. To substantiate this idea, we have evaluated both voice and spiral test data using various machine learning models. The results based on the two types of dataset demonstrate an excellent level of accuracy for PD identification.**

Pairwise correlation and k-means clustering techniques are used to extract features from the vocal dataset. In this classification problem, the highest accuracy of 95.89% is obtained using an ensemble of 3 classification models. The Pearson's correlation is used to extract features from the image dataset. The best accuracy of 99.6% is achieved using the k-Nearest Neighbors classifier in the Dynamic Spiral Test (DST). An accuracy of 98.8% and 94.9% are achieved using the Logistic Regression classifier and the Adaptive Boosting classifier on the Static Spiral Test (SST) and Stability Test on Certain Point (STCP) respectively. A second ensemble making use of results from DST, SST, and STCP will provide the overall result of the spiral test.

Additionally, we propose to implement the multi-modal approach into a touch-enabled smartphone-based application to carry out preliminary PD tests conveniently, without the need of supervision of additional medical personnel or any specialized equipment. The final ensemble for the application makes use of the results of the respective ensemble from the vocal and spiral test.

***Index Terms*—Parkinson disease detection, pairwise correlation, k-mean, ensemble, vocal test, sprial test**

I. Introduction

According to statistics, there are about 60,000 Americans diagnosed with Parkinson disease (PD) each year while there are more than 10 million people worldwide with PD. The probability of getting Parkinson's disease increases with age with approximately four percent of the PD patients worldwide are under the age of 50. It is estimated, that the number of Parkinson's disease patients will increase from 4.1 million in 2005 to 8.7 million in 2030 [1].

There are a few measurement techniques for PD [2], such as Unified Parkinson's Disease Rating Scale (UPDRS), the HoehnYahr Scale, the Schwab and England Scale of Activities of Daily Living and the Parkinson's Disease Questionnaire 39. Among them, UPDRS is the most common technique used to follow the PD progression and evaluate the results of surgical, medical, and other interventions of the disease. Speech characteristics are key components for early-stage detection where vocal impairment was detected in approximately 90% of the patients who are in their earlier stages of the disease [3]. Hence, there is an increasing interest in building PD diagnostic and tele-monitoring systems based on vocal features.

Lately, researchers found out that diagnostic tools such as biological and genetic biomarkers as well as imaging techniques have high accuracies in predicting PD [2]. In 2017, a group of Australian scientists developed a new diagnostic method called the Spiral test. The test could detect signs of the disease by analyzing the amount of time used to draw spirals, amount of pressure exerted and the characteristics of the lines.

This study proposes a multi-modal approach to detect early onset of PD utilizing two of the early symptoms displayed by People with Parkinson's (PWP): vocal test and spiral test. Two datasets have been used in this research – one vocal test dataset and one spiral test dataset to substantiate the suitability of these tests in the application.

II. Related Work

A. Vocal Data Set

Studies in the area of using vocal test to detect Parkinson disease can be categorized into two main groups: to ascertain which are the most effective vocal features [4]–[6] to detect Parkinson's disease, while some studies focus on enhancing classification accuracy (to distinguish between healthy and unhealthy subjects) [7], [8].

Tsanas *et al.* [9] focused on features extraction using signal processing techniques applied on voice samples taken from 42 patients with early-stage PD to estimate the unified Parkinson's disease rating scale (UPDRS) using linear and non-linear regression. Their results show an accuracy difference of about 7-point from clinical UPDRS estimations.

Tsanas *et al.* [7] applied feature selection techniques using 132 dysphonia measurements on 263 samples. The team obtained 99% overall classification accuracy. In another work, Sakar *et al.* [6] mutual information-based selection algorithm was applied using a permutation test to select features and rank them based on maximum-relevance-minimum-

TABLE I
DESCRIPTION OF SPEECH SAMPLES COLLECTED

Sample No.	Description
1	Sustained vowel ”a”
2	Sustained vowel ”o”
3	Sustained vowel ”u”
4 -13	Numbers from 1 10
14 - 17	Rhymed Short sentences
18 - 26	Words in Turkish language

redundancy (mRMR) into an support vector machine (SVM) classifier. Using leave-one-subject-out (LOSO) as the cross-validation technique of their model in order to avoid bias, their approach achieved 92.75% of classification accuracy.

B. *Spiral Test Data Set*

This area of study is relatively new, it is often used to ascertain if subjects have PD [10]. In [11], the analysis was done on drawn spirals, capturing kinematic, dynamic, and spatial abnormalities. These were used to calculate indices that quantify motor performance and disability. Linear mixed effect models adjusting for age, gender and handedness were used to compare the different indices between cases and controls. Results show that spiral indices yielded good accuracy in diagnosing early PD from cross-validation studies.

In [10], [11], seven characteristic parameters were extracted from hand movement in spiral drawing experiment. The characteristic value was linearly interpolated. The team identified a few good predictors of PD in this paper.

III. MATERIALS AND METHODS

A. *Data*

Two data sets, the Parkinson Speech Dataset with Multiple Types of Sound Recordings [6] and Improved Spiral Test Using Digitized Graphics Tablet for Monitoring Parkinson's Disease, were obtained from the UCI Machine Learning Repository. The findings from these data sets will substantiate the proposed application for mobile detection and monitoring of Parkinson's disease.

1) Vocal Data Set: The PD dataset [6] used in this study, comprising of 26 speech samples of 20 PD patients and 20 healthy individuals, was collected by the Department of Neurology in Cerrahpasa, Faculty of Medicine of Istanbul University. The 26 speech samples consist of sustained vowels, numbers, words and short sentences which are recorded using a Trust MC-1500 microphone with a frequency range between 50Hz and 13kHz. For the collection of the speech samples, the microphone was placed at a distance of 10cm from the subject. A description of the samples is described in Table I.

From each of the samples collected, the Praat acoustic analysis software is used to extract 26 linear and time-frequency based features with reference to previous works in this field of study [5], [12]. The features of the PD dataset used in this study are as seen in Table II.

In addition to the data collected from the 40 individuals stated above, a separate dataset was collected from a test group

TABLE II
TIME-FREQUENCY-BASED FEATURES EXTRACTED FROM SPEECH SAMPLES

Features	Group
Jitter (local) Jitter (local, absolute) Jitter (rap) Jitter (ppq5) Jitter (ddp)	Frequency Parameters
Number of pulses Number of periods Mean period Standard dev. of period	Pulse Parameters
Shimmer (local) Shimmer (local, dB) Shimmer (apq3) Shimmer (apq5) Shimmer (apq11) Shimmer (dda)	Amplitude Parameters
Fraction of locally unvoiced frames Number of voice breaks Degree of voice breaks	Voicing Parameters
Median pitch Mean pitch Standard deviation Minimum pitch Maximum pitch	Pitch Parameters
Autocorrelation Noise-to-harmonic Harmonic-to-noise	Harmonicity Parameters

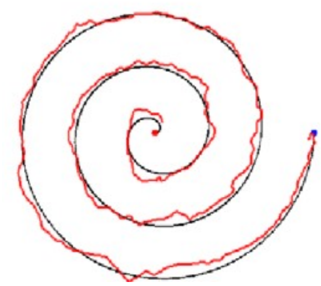

Fig. 1. Static Spiral Test

made up of a separate 28 PD patients. For this test group, only the sustained vowels “a” and “o” samples are collected from each patient instead of the 26 samples type collected from each of the 40 individuals.

2) Spiral Test Image Data Set: The image data set [13], [14] consists of 15 healthy individuals (thereafter known as the control set) and 62 individuals diagnosed with Parkinson's disease (thereafter known as People-with-Parkinson's (PWP)). For all subjects, three types of handwriting recordings, the Static Spiral Test (SST), Dynamic Spiral Test (DST) and Stability Test on Certain Point (STCP) are taken.

The Static Spiral Test is frequently used for clinical research for purposes like determining motor performance, measuring tremor, and diagnosing PD. In this test, three wound Archimedean spirals appear on the graphics tablet and patients are asked to retrace the same spiral as much as they can with a digital pen. An example image of the Static Spiral Test is illustrated as Figure 1.

The second test is the Dynamic Spiral Test which is similar

TABLE III
ATTRIBUTES FROM SPIRAL TEST

Attributes	Collected
X	X position (left/right) of the screen
Y	Y position (up/down) of the screen
X	Z position of the pen (perpendicular to the screen)
Pressure	Pressure on the screen with the digital pen
Grip Angle	Individual user grip angle of the pen
Time	System time sample is recorded
Test ID	0 : Static Spiral Test (SST) 1 : Dynamic Spiral Test (DST) 2: Stability Test on Certain Point (STCP)

to the Static Spiral Test. However, in this test, the spiral disappears and appears in certain time intervals, forcing the patient to keep the pattern in mind and continue to draw.

The third test is the Stability Test on Certain Point, in which subjects are asked to hold the digital pen on the point without the rest of the hand touching the screen in a certain time. The purpose of this test is to determine the patient's hand stability or hand tremor level. The attributes from spiral test are shown as Table III.

B. Data Preprocessing

1) Vocal Data Set: Exploratory data analysis and specifically pairwise correlation were carried out to investigate if there were any features which are highly correlated to one another. From the investigation, it was found that there were certain features which have correlation coefficients of above 0.85 with other feature(s). The highly correlated features were thus not used in this study. In addition to the features with correlation above 0.85, Jitter (local, absolute) and Shimmer (apq11) were both removed during feature selection even though each only has a correlation of 0.74 with Jitter (ppq5) and 0.73 with Shimmer (local) respectively. It was tested and verified that the removal of Jitter (local, absolute) and Shimmer (apq11) did not have a significant impact on the results of the model.

Table IV shows the features remaining after feature selection which are used in this study. Therefore, out of the initial 26 features processed, only 13 features are used for the development of the classifier using the vocal data set.

Previous studies have shown that the sustained vowels tests carry more PD discriminant information. K-means clustering, a unsupervised learning without heuristic basis, was carried out to investigate if there was a possible difference of information gain in the speech samples. K-means clustering was performed on a range of 2 to 10 clusters. A silhouette score was produced, indicating how close the features and voice samples are within each cluster. The higher the silhouette score, the more distinguished the clusters are from each other. Two clusters produce the highest silhouette scores. Further studies using the 2 clusters showed that among the 26 speech samples, samples number 1, 2 and 3 which are the sustained vowels "a", "o" and "u", displayed a relatively high frequency in one cluster compared to the other, which suggest they may have a significant impact on the classification of PD.

TABLE IV
FEATURES REMAINING AFTER FEATURE SELECTION

Feature	Group
Jitter (ppq5)	Frequency Parameters
Number of pulses Standard dev. of period	Pulse Parameters
Shimmer (local)	Amplitude Parameters
Fraction of locally unvoiced frames Number of voice breaks Degree of voice breaks	Voicing Parameters
Median pitch Standard deviation Minimum pitch Maximum pitch	Pitch Parameters
Noise-to-harmonic Harmonic-to-noise	Harmonicity Parameters

The above exploratory study was supported by earlier work done in [5] which has shown that using only sustained vowels is sufficient to make a prediction of PD occurrence. The speech sample 1, 2 and 3 was proved to show a high information. Then, we decided to only use the sustained vowels "a" and "o" samples for the study to conform to the UCI test data set. The UCI test data set only contained the "a" and "o" samples.

2) Spiral Test Image Dataset: The data consists of the individual samples of each test that each user of the system did. The first step was to collate all samples into one table where each sample was labeled if it belonged to a control subject or a PWP subject.

Running Pearson's Correlation on the dataset singled out the attributes "Z" and "Pressure" as very highly correlated. As more information (graduation of 1024 steps) was available in the attribute "Pressure", the attribute "Z" was removed. The dataset was then split into the 3 tests, labeled Test 0, Test 1 and Test 2. The attributes "Subject", "Timestamp" and "Test ID" were also removed for each of these 3 test datasets. Finally, as the intended platform of the app does not have the grip angle variable in its most basic mode using a standard stylus, we left the variable "grip angle" out of the model building and analysis. Prior to the models, the attributes were scaled according to Min-Max scaling.

C. Methods

1) Vocal Data Set: As previously mentioned, data was collected from a total of 68 individuals out of which 28 came from a test group. For this study, data from all 68 individuals were mixed to form the overall dataset from which the train and test set were split by the 80:20 ratio. In addition, it was also found that the dataset was highly skewed towards the **PD class ("P" class)** as a result of the sampling. Thus, oversampling of the **healthy class ("H" class)** was carried out to achieve an overall "P" class to "H" class ratio of 1.3 : 1 in the overall data set.

The model for the binary classifier was trained using the k-fold cross-validation method and k = 5 is used in this study.

TABLE V
CLASSIFIERS FOR VOCAL DATA SET

Classifier	Description
Random Forest	The model was built using the grid search method to determine the number of variables randomly sampled as candidates at each split.
k-Nearest Neighbours	Normalization of predictors was carried out and 10 different values of k were tested and k = 9 was found to produce the best accuracy.
Support Vector Machine (SVM)	Normalization of predictors was carried out and 10 different values of regularization parameter C were tested. Radial basis kernel used.

TABLE VI
CLASSIFICATION METHODS ATTEMPTED FOR SPIRAL IMAGE DATA

Classification Methods Used	
Adaboost	Estimators: 100, Algorithm: SAMME.R, Linear
k-Nearest Neighbours	4 neighbours, Mahalanobis metric, Uniform Weight
Neural Network	Layers (100,100,50), AdamOptimizer, Activation: ReLu, Alpha: 0.0001, Max iterations: 200
Simple Decision Tree	Do not split subset smaller than 5, max tree depth: 100
Nave Bayes	-
Logistic Regression	Lasso L1 regularization
Random Forest	50 trees, limit depth of individual trees to 3
Stochastic Gradient Descent	Loss function: Hinge, L2 regularization, constant learning rate: 0.01
Support Vector Machines	Cost: 1, loss epsilon : 0.1, Cubic polynomial kernel

The k-fold cross-validation method is chosen over having a single hold-out set due to the small amount of data which may result in larger variations for performance evaluation. Another advantage of k-fold cross-validation is that it requires lower training time compared to the repeated k-fold cross-validation and leave-one-out cross-validation methods while producing sufficiently accurate results in this study.

Three types of modeling techniques were built and tested for the study involving the vocal dataset as seen in Table V. The results from each of the classification model are further discussed in the subsequent sections.

2) Spiral Test Image Dataset: Previously, the data was regrouped by the tests. Each test data was then split using a "Leave One Subject" out method. In our case, we left 2 subjects out per test, a PWP and a control subject to provide a balanced test analysis. Data balancing was done on the training data. For each model, 5-fold cross-validation was also performed make sure that the model did not over-fit. Finally, the model was used against the samples of the 2 test subjects to get the final classification accuracy of the model. We made use of the following models in Orange, which is based on the Python Scikit-Learn libraries for the classification of this dataset.

Interestingly, the SST, DST and STCP tests responded well to different models. For brevity, for the next section, we will only focus on the models that gave a good cross-validation and test accuracy for each spiral test.

TABLE VII
CONFUSION MATRIX FOR RANDOM FOREST

	Predicted "H"	Predicted "P"	Sum
Actual "H"	32	0	32
Actual "P"	3	38	41
Sum	35	38	73

TABLE VIII
CONFUSION MATRIX FOR K-NEAREST NEIGHBORS

	Predicted "H"	Predicted "P"	Sum
Actual "H"	27	5	32
Actual "P"	7	34	41
Sum	34	39	73

IV. MODELING

A. Vocal Data Set

1) Evaluation Metrics: The evaluation metrics used for the classification model are accuracy, sensitivity, and specificity. These metrics are commonly used in classification models and they are also one of the most comprehensible metrics.

The confusion matrix, which offers a simple comparison between the predicted values and their actual values, will be used to present the results of the classification models tested. The three confusion matrices seen in Tables VII, VIII, IX are the results of the test set which was not involved in the development of the training model.

Based on the results above, it is observed that the k-Nearest Neighbours and SVM models were not able to perform as well as the Random Forest model. However, it can be observed that both the Random Forest and SVM models are able to achieve zero false negative predictions while all models exhibited false positive predictions ranging from 3 to 7 cases.

Table X gives a summary of the performance metrics of each classification model and their ensemble based on majority voting.

In addition to the three evaluation metrics, the confidence interval for the ensemble error is calculated as follows:

$$error \pm const * \sqrt{\frac{error * (1 - error)}{n}} \quad (1)$$

TABLE IX
CONFUSION MATRIX FOR SVM

	Predicted "H"	Predicted "P"	Sum
Actual "H"	32	0	32
Actual "P"	4	37	41
Sum	36	37	73

TABLE X
PERFORMANCE OF ALL MODELS

	Accuracy	Specificity	Sensitivity
Random Forest	95.89%	100%	91.43%
k-Nearest Neighbours	83.56%	87.18%	79.41%
SVM	93.15%	100%	86.49%
Ensemble	95.89%	100%	91.43%

TABLE XI
CONFUSION MATRIX FOR SST USING LOGISTIC REGRESSION

	Predicted "H"	Predicted "P"	Sum
Actual "H"	2945	169	3114
Actual "P"	0	1246	1246
Sum	2945	1415	4360

TABLE XII
CONFUSION MATRIX FOR DST USING k-NEAREST NEIGHBORS

	Predicted "H"	Predicted "P"	Sum
Actual "H"	2363	10	2373
Actual "P"	3	1112	1115
Sum	2366	1122	3488

where the constant is dependent on the level of confidence and n is the number of instances used for the evaluation of the model. The error translates to 0.041 +/- 0.046 with 95% confidence interval.

B. Spiral Test Image Data Set

1) Static Spiral Test: For SST, the best accuracy came from Logistic Regression classifier, giving an accuracy of 98.8%, sensitivity of 96.1% and specificity of 94.6%. The error translates to 0.012 +/- 0.003 with 95% confidence interval. Table XI shows the confusion matrix for SST using Logistic Regression.

2) Dynamic Spiral Test: For DST, the k-Nearest Neighbors classifier scored at best, giving an accuracy of 99.6%, sensitivity of 99.6% and a specificity of 99.6%. The error translates to 0.004 +/- 0.002 with 95% confidence interval. Table XII shows the confusion matrix for DST using k-Nearest Neighbors.

3) Stability Test on Certain Point: For STCP, the best accuracy came from the Adaptive Boosting Classifier.

$$F(x) = sign(\sum_{m=1}^{M} \theta_m f_m(x)), \tag{2}$$

where f_m stands for the m-th weak classifier and θ_m is the corresponding weight. AdaBoost is adaptive in the sense that subsequent weak learners are tweaked in favor of those instances misclassified by previous classifiers. Adaptive boosting was able to give an accuracy of 94.9%, sensitivity of 99.9% and specificity of 86.9%. The error translates to 0.051 +/- 0.007 with 95% confidence interval. Table XIII shows the confusion matrix for STCP using Adaptive Boosting.

C. Proposed Final Ensemble Method

For the vocal test data set, a voting ensemble which makes use of the prediction from the 3 classifier models will determine whether the patient is at risk of Parkinson's disease. As illustrated in Table XIV, if at least 2 of the 3 models give a positive prediction, the final ensemble will then give a positive prediction.

For the spiral tests, a voting ensemble model will be used. The prediction for a single test is taken from the majority prediction of all the samples for this particular test. In this way, we would be able to use the accuracy as the percentage of confidence we have in the prediction for this particular test. The output of the ensemble is a single prediction stating whether the patient is at risk, along with the accuracy obtained for that user for test 1, 2 and 3. As illustrated in Table XV, if at least 2 of 3 tests give a positive prediction, the final ensemble will then give a positive prediction.

The final diagnosis will be determined by either the vocal or spiral ensemble having a positive diagnosis, i.e., having PD. This is done with the consideration that our analysis has shown that being tested positive on either the vocal or spiral test is sufficient to diagnose the subject as suffering from PD. For clarity, Table XVI shows the possible combinations of results from the two tests.

TABLE XIII
CONFUSION MATRIX FOR STCP USING ADAPTIVE BOOSTING

	Predicted "H"	Predicted "P"	Sum
Actual "H"	1336	200	1536
Actual "P"	2	2401	2403
Sum	1338	2601	3939

TABLE XIV
VOCAL TESTS ENSEMBLE

User	Random Forest	k-Nearest Neighbors	SVM	Ensemble (Final Prediction)
PWP	PWP	PWP	Healthy	PWP

TABLE XV
SPIRAL TESTS ENSEMBLE

User	SST (k-Nearest Neighbors)	DST (k-Nearest Neighbors)	STCP (Adaptive Boosting)	Final Prediction
PWP	PWP	Healthy	PWP	PWP

V. PROPOSED SMARTPHONE APPLICATION

The proposed App currently will only be on the Apple iOS platform with a 3D-Touch enabled Apple iPhone. The reason is that only the iPhone has a screen pressure sensor, which is crucial for the classification of the spiral tests. Moreover, the iOS Software Development Kit has built-in functionality to do audio recording and simple drawings, making it fast to come up with a prototype. The same App can be used on other mobile phones when the mobile phone makers build in pressure sensors into their hardware. For the vocal test, signal processing routines will have to be programmed to extract the relevant attributes from the audio clips. This can be programmed efficiently through the use of the Accelerate

TABLE XVI
OVERALL ENSEMBLE

Vocal Tests	Spiral Tests	Final Diagnosis
Negative	Negative	Negative
Positive	Negative	Positive
Negative	Positive	Positive
Positive	Positive	Positive

Framework in the iOS SDK. Apple has also recently introduced CoreML, which allows one to load Scikit-Learn based models right into the phone, this enables users to use the app offline, if required. It is not advisable to use a finger as input for tracing the spiral. We would highly recommend the use of a standard stylus for the spiral test.

We have conducted our spiral test analysis with the assumption that the attribute "grip angle" will not be attainable by the phone hardware.

VI. Conclusion

The analysis of the datasets has proven that both speech and spiral test can be relied upon for Parkinson's disease detection. Early detection of PD is critical to administer early treatment and to aid them with necessary changes to adapt to the disease. This paper suggests the possibility of an early stage Parkinson's Disease mobile app using the Vocal and Spiral test. Several variables were dropped compared to existing tests and our models are have shown to have high accuracy in detecting Parkinson's disease. Furthermore, the mobile app will enable collection of data remotely from PWP, which can serve as an improved dataset for PD researchers.

The reduction from the 26 voice samplings for the Vocal test to using only the sustained vowel "a" and "o" recording will be convenient and reassuring to apprehensive users.

Previous studies have shown early symptoms of Parkinson's disease is different for every patient - given this, using multimodal test data involving both voice and image would provide a more effective way to detect PD for patients who may display different types of predominant symptoms. As such, we believe that having a 2-prong approach (using the two indicated test) would be more reliable in detecting early-stage Parkinsons Disease.

The models were built using small Vocal test and Spiral test data set. Since both studies were conducted in Istanbul University, it is reasonable to deduce the subjects were from Turkey. In addition, both data sets did not indicate the severity of the patients' Parkinson's Disease. More data is needed from different regions around the globe with patients having an early stage of Parkinson to improve the reliability and the robustness of the models.

For the proposed mobile App, the voice recording quality would likely be different compared to the Trust MC-1500 microphone employed in the test. Similarly, for the spiral test, the test done on the iPhone may vary from the graphics tablet as the iPhone screen is comparably smaller than the tablet used in the original dataset test. Data collection has to be re-done using the App on the iPhone itself. This would improve the accuracy and sensitivity of the App. Another way that may improve the performance of the system is applying ensemble learning with more classification algorithms, such as, the support vector machines, random forests, enhanced k-nearest neighbours [15], and kernel dictionary learning [16].

References

[1] ERl Dorsey, R Constantinescu, JP Thompson, KM Biglan, RG Holloway, K Kieburtz, FJ Marshall, BM Ravina, G Schifitto, A Siderowf, et al. Projected number of people with Parkinson disease in the most populous nations, 2005 through 2030. *Neurology*, 68(5):384–386, 2007.

[2] Michela Tinelli, Panos Kanavos, and Federico Grimaccia. The value of early diagnosis and treatment in Parkinson's disease: a literature review of the potential clinical and socioeconomic impact of targeting unmet needs in Parkinsons disease. 2016.

[3] Aileen K Ho, Robert Iansek, Caterina Marigliani, John L Bradshaw, and Sandra Gates. Speech impairment in a large sample of patients with Parkinson's disease. *Behavioural Neurology*, 11(3):131–137, 1999.

[4] Mahnaz Behroozi and Ashkan Sami. A multiple-classifier framework for Parkinson's disease detection based on various vocal tests. *International Journal of Telemedicine and Applications*, 2016, 2016.

[5] Max A Little, Patrick E McSharry, Eric J Hunter, Jennifer Spielman, Lorraine O Ramig, et al. Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. *IEEE Transactions on Biomedical Engineering*, 56(4):1015–1022, 2009.

[6] Betul Erdogdu Sakar, M Erdem Isenkul, C Okan Sakar, Ahmet Sertbas, Fikret Gurgen, Sakir Delil, Hulya Apaydin, and Olcay Kursun. Collection and analysis of a Parkinson speech dataset with multiple types of sound recordings. *IEEE Journal of Biomedical and Health Informatics*, 17(4):828–834, 2013.

[7] Athanasios Tsanas, Max A Little, Patrick E McSharry, Jennifer Spielman, and Lorraine O Ramig. Novel speech signal processing algorithms for high-accuracy classification of Parkinson's disease. *IEEE Transactions on Biomedical Engineering*, 59(5):1264–1271, 2012.

[8] Pei-Fang Guo, Prabir Bhattacharya, and Nawwaf Kharma. Advances in detecting Parkinson's disease. In *International Conference on Medical Biometrics*, pages 306–314. Springer, 2010.

[9] Athanasios Tsanas, Max A Little, Patrick E McSharry, and Lorraine O Ramig. Accurate telemonitoring of Parkinson's disease progression by noninvasive speech tests. *IEEE Transactions on Biomedical Engineering*, 57(4):884–893, 2010.

[10] Min Wang, Bei Wang, Junzhong Zou, and Masatoshi Nakamura. A new quantitative evaluation method of spiral drawing for patients with Parkinson's disease based on a polar coordinate system with varying origin. *Physica A: Statistical Mechanics and Its Applications*, 391(18):4377–4388, 2012.

[11] Marta San Luciano, Cuiling Wang, Roberto A Ortega, Qiping Yu, Sarah Boschung, Jeannie Soto-Valencia, Susan B Bressman, Richard B Lipton, Seth Pullman, and Rachel Saunders-Pullman. Digitized spiral drawing: A possible biomarker for early Parkinson's disease. *PloS ONE*, 11(10):e0162799, 2016.

[12] C Okan Sakar and Olcay Kursun. Telediagnosis of Parkinson's disease using measurements of dysphonia. *Journal of Medical Systems*, 34(4):591–599, 2010.

[13] M Isenkul, B Sakar, and O Kursun. Improved spiral test using digitized graphics tablet for monitoring Parkinson's disease. In *Proceedings of the International Conference on e-Health and Telemedicine*, pages 171–5, 2014.

[14] Parkinson speech dataset with multiple types of sound recordings data set. https://archive.ics.uci.edu/ml/datasets/Parkinson+Disease+Spiral+Drawings+Using+Digitized+Graphics+Tablet, 2014.

[15] Binh P. Nguyen, Wei-Liang Tay, and Chee-Kong Chui. Robust biometric recognition from palm depth images for gloved hands. *IEEE Transactions on Human-Machine Systems*, 45(6):799–804, Dec 2015.

[16] Xuan Chen, Binh P. Nguyen, Chee-Kong Chui, and Sim-Heng Ong. Automated brain tumor segmentation using kernel dictionary learning and superpixel-level features. In *Proceedings of the 2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC 2016)*, pages 2547–2552. IEEE, Oct 2016.