Robust Detection of Parkinson's Disease Using Harvested Smartphone Voice Data: A Telemedicine Approach

Sanjana Singh[1] and Wenyao Xu, PhD[2]

[1]McLean High School, McLean, Virginia.
[2]Department of Computer Science and Engineering, State University of New York at Buffalo, Buffalo, New York.

Abstract

Introduction: Parkinson's disease affects over 10 million people globally, and ∼20% of patients with Parkinson's disease have not been diagnosed as such. The clinical diagnosis is costly: there are no specific tests or biomarkers and it can take days to diagnose as it relies on a holistic evaluation of the individual's symptoms. Existing research either predicts a Unified Parkinson Disease Rating Scale rating, uses other key Parkinsonian features such as tapping, gait, and tremor to diagnose an individual, or focuses on different audio features.
Methods: In this article, we present a classification approach implemented as an iOS App to detect whether an individual has Parkinson's using 10-s audio clips of the individual saying "aaah" into a smartphone.
Results: The 1,000 voice samples analyzed were obtained from the mPower (mobile Parkinson Disease) study, which collected 65,022 voice samples from 5,826 unique participants.
Conclusions: The experimental results comparing 12 different methods indicate that our approach achieves 99.0% accuracy in under a second, which significantly outperforms both prior diagnosis methods in the accuracy achieved and the efficiency of clinical diagnoses.

Keywords: telemedicine, m-Health, home health monitoring, sensor technology, Parkinson's disease

Introduction

Parkinson's disease affects over 10 million people globally that progressively debilitates an individual.[1] Approximately 20% of patients with Parkinson's have not been diagnosed as such.[2] Early symptoms often go unnoticed, but as the disease progresses, a tremor, impaired balance, changes in writing, and changes in speech[3] contribute to the high cost of Parkinson's, which is over $25 billion each year in the United States alone.[1] A diagnosis is vital to improve symptoms. In recent years, researchers have tried to improve the accuracy of predicting Parkinson's, and many have tried smartphone-based approaches.

There is currently no objective test to diagnose Parkinson's disease. Doctors perform a costly examination to rule out other neurological conditions and to accurately diagnose the patient. It can take a few hours to a few days for a diagnosis. Early in the progression of the disease, pathological signs are difficult to detect but become bilateral as the disease progresses. There is no "cure" for Parkinson's, but there are ways to slow the progression. Early detection of Parkinson's disease helps patients manage their symptoms and slow the degeneration. Studies have shown that the costs of managing Parkinson's disease increase significantly with the disease progression as the offsetting of complications from drugs increases the costs. In the first stage of Parkinson's disease, indirect costs make up the bulk of the ∼12,000 Euros a year in Germany the disease costs the afflicted individual, but as the disease progresses, the cost of health insurance and care make up the majority of the ∼47,000 Euros a year that the disease costs.[4] Slowing the progression of the disease would ultimately reduce the financial burden a diagnosis would impose, and early detection is the key to managing symptoms and slowing the progression.

We present a method that is capable of efficiently running on a smartphone that accurately predicts Parkinson's from a 10-s nonspeech recording taken using a smartphone. This method is far more cost-effective and efficient than traditional clinical methods and could revolutionize the industry. Our method utilizes Mel-Frequency Cepstral Coefficients (MFCCs) feature extraction method, L1-based feature selection, and a Support Vector Machine (SVM) classifier to classify recordings of an individual saying "aaah" for 10 s when sampled at a uniform frequency. Previous research has shown that "aaah" can be used to assess vocal features.[5] Lots of existing literatures have also sought to predict Parkinson's by detecting other key features of the disease. Some researchers have also used vocal features to monitor the progression of Parkinson's disease and estimate the Unified Parkinson Disease Rating Scale (UPDRS) score with limited success.

Related Work

There exists some research regarding the quantification of Parkinson's disease. Existing research either relies on different key features to diagnose Parkinson's to predict the severity of the disease or uses a significantly smaller dataset to design a method to diagnose Parkinson's. Our work differs from existing work in many ways, as detailed below.

DIAGNOSING PARKINSON'S DISEASE FROM MIXED COHORT

Asgari and Shafran[6] introduced a method to predict the severity of Parkinson's disease on the UPDRS scale using three speech tasks solely from Parkinson's patients. Similarly, speech signal processing is used to estimate UPDRS scores with 42 speech tasks divided into subsets of words, monologue, and diadochokinetic tasks in work done by Galaz et al.[7] Recent work by Zhan et al. utilized several features of voice, finger tapping, gait, balance, and reaction time to calculate a mobile Parkinson disease score (not a UPDRS score).[8] Noninvasive speech tests were utilized by Tsanas et al. to predict the UPDRS score with a mean difference of 7.5 from the clinical diagnosis.[9] These works all differ from our research as they quantify the severity of the disease of Parkinson's patients, using data only from Parkinson's patients, while we are diagnosing Parkinson's, using data from both Parkinson's disease and control individuals.

USING VOCAL DATA TO REDUCE MISDIAGNOSIS

Other distinct features of Parkinson's such as gait, olfactory senses, finger tapping, and hand tremor are used to diagnose Parkinson's disease. Existing literature on gait has been used to detect and monitor Parkinson's symptoms on smartphones.[10–12] Challa et al. designed a novel approach to predicting Parkinson's disease using Sleep Behavior Disorder, olfactory loss, and nonmotor features with 97.2% accuracy.[13] Hand tremors are a key feature of Parkinson's disease, and a smartphone-based tool was created to quantify tremor symptoms and classify individuals with Parkinson's (82% accuracy).[14] However, a hand tremor is a common characteristic of several other diseases and can be a source of misdiagnosis.

Another way to observe the movement hindrance inflicted by Parkinson's is through rapid finger tapping evaluations. Khan et al. were able to distinguish healthy controls from Parkinson's disease patients with 95% accuracy, whereas UPDRS-finger tapping (UPDRS-FT) levels were discriminated with 88% accuracy.[15] Using gait as the main feature from which Parkinson's is diagnosed makes it easy to employ in daily use, but the nature of the data makes it noise prone and, thus, less accurate. These works utilize different prominent symptoms of Parkinson's such as hindrances in movement while we focus on auditory features. A recent study by Tolosa et al. found that only 74% of patients diagnosed with Parkinson's truly had the disease, placing the misdiagnosis rate even higher than initially presumed.[16] Nonetheless, the misdiagnosis rate of Parkinson's is incredibly high, and tremor has been shown to be a common source of the misdiagnosis. Jain et al. studied the misdiagnosis of essential tremor (ET) and found that only 11% of individuals diagnosed with ET truly had Parkinson's disease.[17] Meara et al. found ET and Alzheimer's disease to be the most common causes of misdiagnosis, which reinforces the findings of Jain et al.[18] Thus, to avoid misdiagnosis from confounding factors, we focus on quantifying vocal data.

MUCH LARGER DATASET OF VOCAL DATA ANALYZED

Tsanas et al. created a tool to detect Parkinson's in speech patterns. Their dataset consisted of 263 phonations where they sought to distinguish Parkinson's patients from control subjects. Their method used several dysphonia measures and ultimately achieved 98.6% accuracy.[19] Unlike their method, we implement a solution, which extracts MFCC features instead of dysphonia features on a much larger dataset (1,000 samples compared to the 263 phonations).

Much of the literature focuses on discovering symptoms that are characteristic of this disease. A survey conducted by Hartelius and Svensson found that 70% of Parkinson's disease patients had speech impediments and 41% had chewing or swallowing impairment.[20] Gamboa et al. used 2 s sustained "aaah" audio, as well as a sentence to analyze voice qualities of both control and Parkinson's disease individuals. Ultimately, they concluded that Parkinson's disease patients had voice arrests, struggle, and higher jitter that were unaffected by Parkinson's disease severity, measured by the UPDRS.[21] Ramig et al. concluded that noninvasive acoustic analysis could hold clinical value to diagnose and track the progression of neurological diseases, including Parkinson's disease.[22] These research works affirm that Parkinson's individuals have notable vocal impairment, and our work discriminates individuals on vocal alone to avoid confounding factors.

Other existing work relevant to our problem details the high proportion of individuals afflicted with Parkinson's who lack a diagnosis. Schrag et al. estimated that 20% of afflicted patients lacked a professional diagnosis.[2] This astounding proportion can be attributed to the extremely costly process for a clinical diagnosis, often costing thousands of dollars and hours of time, so we designed a method, which has neither of those caveats.

Framework

In this section, we present our framework with details on the feature extraction, feature selection, and classification.

FEATURE EXTRACTION

Using the MFCC transforms the data format, from a wave file to a numpy array, which is portable to use in an SVM classifier. We first divide the signal into 20 ms frames with a frame step of 10 ms to allow for overlaps. Thus, for a 10 s sample, there are 1,000 frames, on which the following procedure will be applied.

First, the Discrete Fourier Transform is applied where $h(n)$ is an analysis window and K is the length of the frame, 20 ms. We then calculate the periodogram based power spectral estimate, $S_i(k)$, for each speech frame, $s_i(n)$, as follows:

$$S_i(k) = \sum_{n=1}^{N} s_i(n)h(n)e^{-\frac{2\pi jkn}{N}} 1 \le k \le K \qquad (1)$$

The Mel-Spaced filterbank, $P_i(k)$, is calculated using 40 filters by taking the summation of the product of each filterbank with a power spectrum[23] as seen in Equation (2):

$$P_i(k) = \frac{1}{N} |S_i(k)|^2 \qquad (2)$$

The data at this step are 40 energies per 1,000 frames per sample. We then take the log of each filterbank energy calculated. The Discrete Cosine Transform is applied to the 40 log energies, and the first 13 coefficients are saved. Thus, each frame has 13 dimensions, outputting an array 1,000 by 13 per sample.

L1-BASED FEATURE SELECTION

The original MFCC dimensions are 1,000 by 13 per sample. Reducing the dimensionality not only makes it more compact and efficient to train a SVM on but may also improve accuracy. An L1-based selection model is a Linear Support Vector Classification model penalized with L1-norm. This model entails the parameter C, which controls the sparsity by trading off variance and bias. With a larger C, more features are selected, and a smaller C leads to fewer features selected. The parameter C greatly influenced the accuracy, and ultimately, a C of 1 was selected which reduced the individual dimension per sample from 1,000 by 13 to a mere 995 by 1. Further insight in the selection of C is detailed in the Comparison Between Tree-Based, L1-Norm, and L2-Norm section.

SVM CLASSIFIER

The features selected in the aforementioned step are fed into a SVM, a supervised learning method to classify the data into Parkinson's and Control groups. SVM classifiers use hyperplanes to separate the data into distinct groups.[24] Before the SVM step, further preprocessing is applied using the Scikit learn MinMaxScaler. This method translates the features from minimum to maximum on a scale from 0 to 1, and the same transformation is applied to the dataset to test upon.

Several different types of SVM Classifiers were run, of which the Radial Basis Kernel SVMs achieved the highest accuracy. Unlike a traditional support vector classification (SVC), which has a linear kernel:

$$\langle x, x' \rangle, \qquad (3)$$

the radial basis function (RBF) Kernel,

$$\exp(-\gamma ||x - x'||^2), \qquad (4)$$

includes a gamma parameter. The decision boundary for a RBF-SVC is more circular and fluid than a linear SVC, which has linear divisions separating the classes.

An RBF-SVM classifier must be fine-tuned using two parameters, C and gamma. The parameter C trades off variance and bias, while gamma alters the influence each individual sample has on the decision boundary. Changes, as small as 0.001, to either parameter can have significant ramifications on the accuracy achieved by the classifier.

Figure 1 shows our method. The raw audio data were divided into fixed length segments. These segments were then randomly placed into training and testing sets. MFCCs were extracted for each segment. The most important features were selected in L1-based feature selection, which reduced the feature dimension from 13,000 per sample to 995. An SVM classifier was trained on the selected features, and the test accuracy was noted. If the accuracy was unacceptable, the MFCC feature extraction, L1-based feature selection, and SVM parameters were altered.

Optimization and Evaluation

In this section, we describe details on the optimization of the method, comparisons of various feature extraction methods, and comparison with eight other methods implemented.

DATASET

The data are from mPower: Mobile Parkinson Disease Study, which is directed by Sage Bionetworks.[25] The data were collected from the Parkinson mPower study app, using Research Kit, and the dataset consists of 65,022 ten-second voice samples of 5,826 unique participants. The app was made available on the Apple App store to users with iPhone 4S's and above with at least software of iOS 8. The ease of submitting data to this study, from the comfort of one's own smartphone, contributed to the unprecedented scale of this study. Each participant involved in the study first downloaded the app,
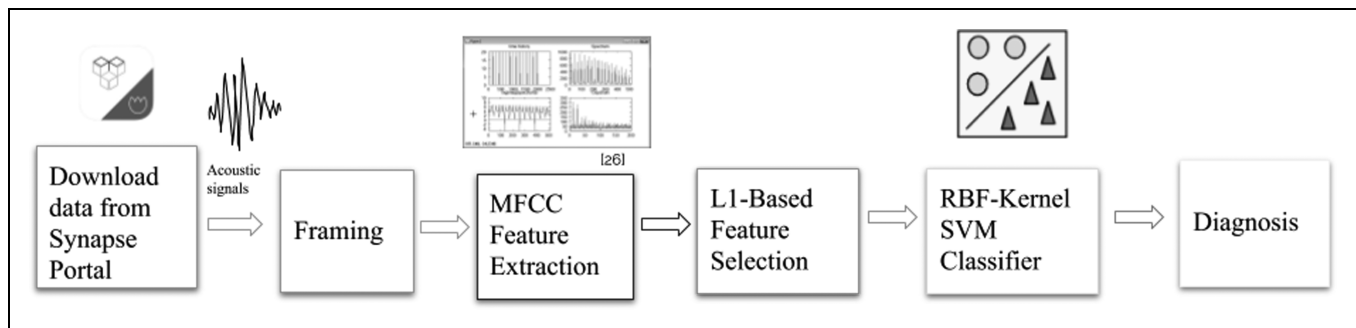
Fig. 1. Flowchart of our method. First, we downloaded data from the Synapse portal. These audio signals were divided into frames, and features were extracted utilizing MFCC feature extraction. The most predictive MFCC features were selected utilizing L1-based feature selection, and these selected features were classified into Parkinson's and control groups using an RBF-Kernel SVM classifier to output a diagnosis. MFCC, mel-frequency cepstral coefficient; RBF, radial basis function; SVM, support vector machine.

filled out the informed consent forms, and performed Parkinson's disease activities. Individuals could selectively participate in the following five main studies: demographics, gait, memory, finger tapping, and voice. The smartphone is the most ubiquitous mobile device, and its ability to handle ambient noise makes it a very good recording device for such a large study. To ensure quality recordings, the user must first record background noise before they are approved to record them saying "aaah," as seen in *Figure 2*. All data were sampled at a wavelength of 44.1 kHz to maintain uniformity. We randomly chose 1,000 samples from mPower. These 1,000 samples were randomly split into a training set consisting of 800 samples and a test set of 200 samples. Fourteen percent of the participants in the overall study were professionally diagnosed with Parkinson's, whereas the remaining 86% consisted of control individuals.[25] Age is the largest risk to developing Parkinson's disease as advancing age reduces neurons and may cause this disease.[26] The prevalence in the U.S. population is estimated at 0.3%, but increases to 4% to 5% for individuals over 85 years old.[27] However, the ages of participants were overwhelmingly slanted toward younger individuals, as individuals in the age range of 18 to 34 made up 57% of participants, whereas only 18% were at least 55 years old.[25] To the best of our knowledge, this is the largest dataset of Parkinson's audio data.

CLASSIFICATION OF PERFORMANCE EVALUATION AND FURTHER STEPS

This subsection evaluates the performance of the framework. Previous studies have sought to detect Parkinson's disease, citing the high misdiagnosis rates. Our work ultimately yields a Parkinson's diagnosis rate of 99.00%, which is notably higher than prior work, which attained an accuracy rate of 98.6%.[19] Further evaluation is in action to investigate other possible confounding factors such as gender, smoking, and other diseases (Alzheimer's, ET) in the diagnosis. Beyond speech data, tapping, gait, and memory data are available in the mPower dataset and can be used to synthesize a comprehensive diagnostic method, which can attain even lower misdiagnosis rates. Furthermore, a more powerful supercomputer would be capable of processing the methods on the full dataset in a timely manner to tune the parameters of each method and determine the most accurate method. Analyzing
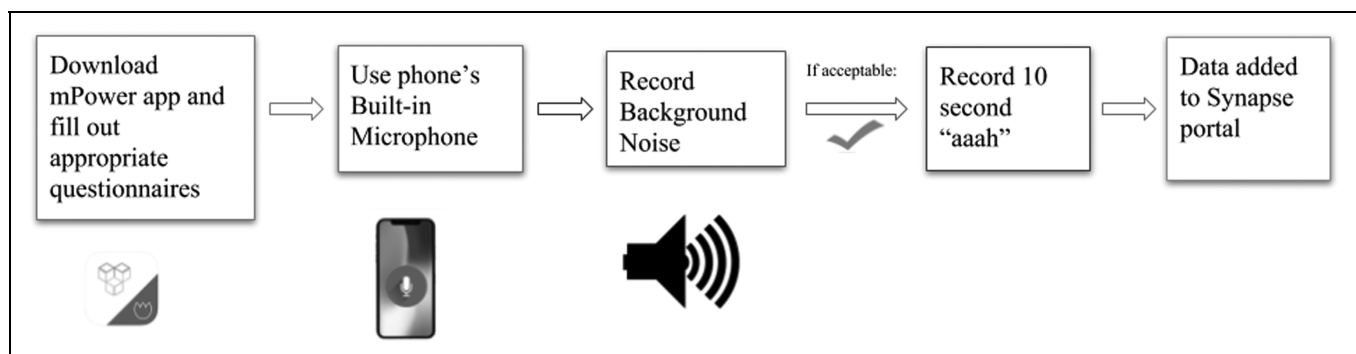

Fig. 2. Flowchart detailing the data collection process. The individual must first download the mPower app and fill out consent forms and questionnaires. For the vocal task, the individual first had to record background noise to determine whether they had a sufficiently low level of background noise. If it was deemed acceptable, the individual was able to record an audio clip saying "ahhh" for 10 s.

whether smartphone type (iPhone 4s and upward) affects the accuracy attained by the methods, as well as whether biological information included in the mPower dataset (gender, other conditions) affects the accuracy, would be extremely valuable in determining how to further improve the diagnostic accuracy.

COMPARISON BETWEEN FEATURE EXTRACTION METHODS

Several feature extraction methods were implemented, with significantly lower accuracy in discriminating between Parkinson's and control subjects.

One method implemented extracted the parameters of each file, which amounted to be ∼445,000 per file. Using the wave package, the number of channels, sample width, frame rate, and number of frames were extracted. These parameters were read and transposed into a numpy array. This method attained a 55% test set accuracy, which pales in comparison to the 99.0% test accuracy achieved by the MFCC feature extraction.

Another method to extract features from the original wave file was to graph the waveform images, mapping the amplitude against time. The axes were standardized, and the images were transformed into grayscale images to add uniformity. This method of feature extraction, at best, predicted 81% of the test set accurately. Greater detail on the methods used following this feature extraction can be found below.

Ultimately, the MFCC feature extraction outperformed both the aforementioned feature extraction methods by over 15% each; hence, the use of the MFCC feature extraction in this method.

COMPARISON BETWEEN TREE-BASED, L1-NORM, AND L2-NORM

Several methods were tried with the ultimate goal of improving the SVM accuracy. Without dimensionality reduction, the SVM achieved a test set accuracy of 0.93. Tree-based feature
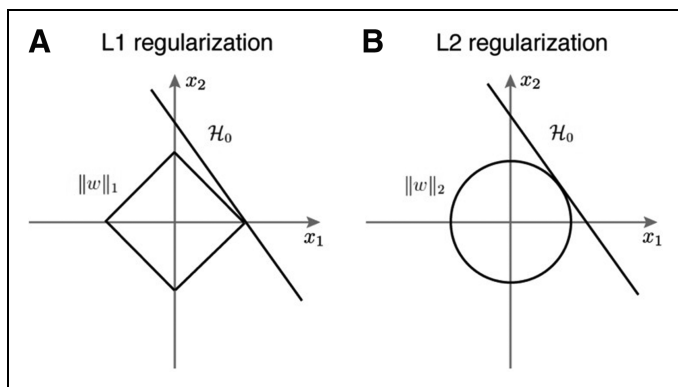
Table 1. The Effect of Frame Length and Frame Stride on Accuracy

DIMENSIONS	FRAME LENGTH	FRAME STRIDE	TEST SET ACCURACY
1,000 × 13	0.01	0.01	0.98
500 × 13	0.01	0.02	0.95
334 × 13	0.01	0.03	0.98
1,000 × 13	**0.02**	**0.01**	**0.99**
500 × 13	0.02	0.02	0.98
334 × 13	0.02	0.03	0.95
1,000 × 13	0.03	0.01	0.98
500 × 13	0.03	0.02	0.97
334 × 13	0.03	0.03	0.95

The bold values are the values that attained the highest test set accuracy and the ones we selected for our model.

selection computes importance of features using decision trees as outlined by Hall.[28] This method of feature selection actually dropped the test set accuracy to 0.86. L2-based feature selection was then implemented in place of the tree-based feature selection. The penalization with the L2-norm achieved a test accuracy of 0.96. Finally, L1-based feature was implemented, which is regularization penalized with an L1-norm. This regularization method, the penalization with the L1-norm, achieved a test set accuracy of 0.99.

Unlike L2-based feature selection, the L1-based boundaries are more rigidly linear, as seen in *Figure 3*, which suits the data

Fig. 3. Comparison between **(A)** L1 regularization and **(B)** L2 regularization.[29]

Table 2. The Effect of Feature Selection on Accuracy

METHOD	PARAMETER VALUE	NUMBER OF FEATURES	TEST SET ACCURACY
None	–	13,000	0.93
Tree based	–	1,378	0.86
L1-based	0.01	252	0.90
L1-based	0.1	578	0.98
L1-based	**1**	**995**	**0.99**
L1-based	5	1,638	0.98
L1-based	10	2,219	0.97
L2-based	0.05	5,264	0.96
L2-based	5	5,265	0.96

The bold values are the values that attained the highest test set accuracy and the ones we selected for our model.
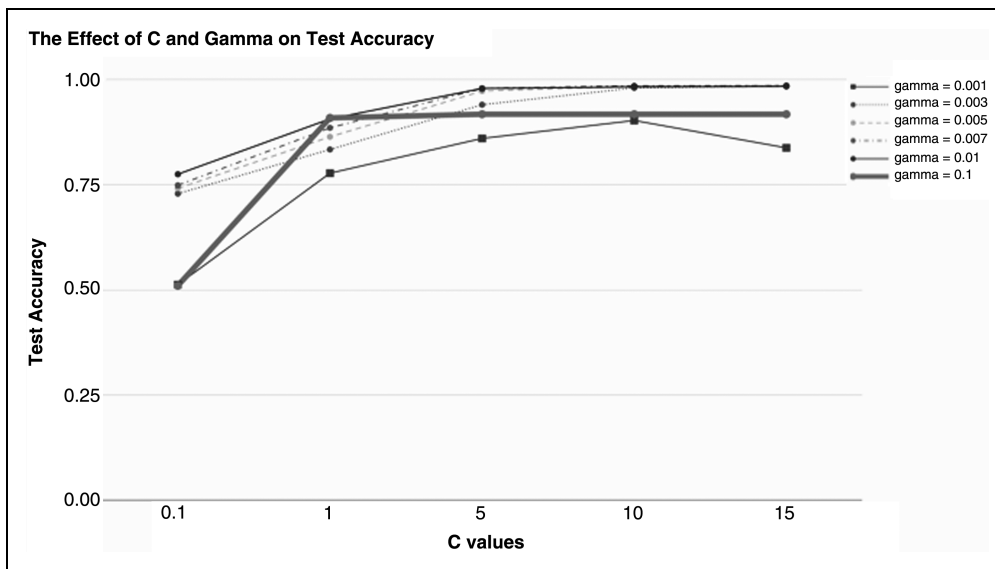
Fig. 4. The Effect of *C* and gamma parameters of the SVM on the test set accuracy. All other parameters were held constant to ensure uniformity. We selected a *C* of 15 and a gamma of 0.005 due to the improvement in test set accuracy.

more, as seen by the significant increase in accuracy. The figure illustrates the effect that the penalization method affects the features selected.

SELECTING OPTIMAL PARAMETERS

The MFCC feature extraction is the basis upon which the rest of the steps follow. Thus, it is essential to tailor the frame stride

Table 3. Test Set Accuracies for all 12 Methods Developed	
DESCRIPTION	**HIGHEST TEST SET ACCURACY**
MFCC, L1, RBF-SVM	0.99
MFCC, L2, RBF-SVM	0.96
MFCC, RBF-SVM	0.93
MFCC, tree-based, RBF-SVM	0.86
Autoencoder, linear-SVM	0.81
MFCC, linear-SVM	0.74
Autoencoder, RBF-SVM	0.73
Grayscale images, RBF-SVM	0.73
nFrames, L1, RBF-SVM	0.55
Grayscale images, linear-SVM	0.52
Grayscale images, logistic regression	0.46
Convolutional neural network	0.46

MFCC, mel-frequency cepstral coefficient; RBF, radial basis function; SVM, support vector machine.

and frame length in the MFCC extraction precisely. The frame length is the width of each frame, each frame is studied individually, and 13 cepstral coefficients are extracted per frame. The frame stride is the distance between each frame, and short strides can allow for overlaps while longer strides can reduce the dimensions. As observed in *Table 1*, the optimal frame length and frame stride are 0.02 and 0.01, respectively. Parkinson's patients have distinct vocal features, including voice arrests.[21] This combination allows for enough overlap to pick up on the distinct vocal features exhibited by Parkinson's patients.

The parameter *C* in L1-based feature selection has a significant effect on accuracy attained. As aforementioned, the parameter *C* controls sparsity: a smaller *C* results in fewer features selected. *Table 2* indicates that a value of 1 for the parameter of *C* reduces the dimension to 995 while achieving a remarkable test set accuracy of 0.99. Clearly, a higher accuracy is preferred so we selected the value of *C* as 1.

C and gamma parameters in the RBF-Kernel SVM affect the way the SVM creates a decision boundary to discriminate samples into Control and Parkinson's classes. As previously stated, gamma controls the effect each individual sample has on the decision boundary. Tuning the gamma parameter is vital to ensure that outliers don't skew the boundary. *Figure 4* illustrates the extensive effect of changing these parameters. We selected a *C* of 15 and a gamma value of 0.05, which yielded a test set accuracy of 0.99.

OTHER METHODS IMPLEMENTED

Several other methods were implemented, including deep learning methods, with a wide range of accuracies as listed in *Table 3*. Three feature extraction methods were tried, namely, graphing the waveform, utilizing parameters of the wave, convolutional autoencoders, and MFCC extraction. Following the different feature extraction methods, several distinct methods were used.

For the waveform images, deep learning methods were deployed. Following the creation of the standardized images, five different methods were implemented. First, a convolutional autoencoder was fine-tuned to compress the high dimensional
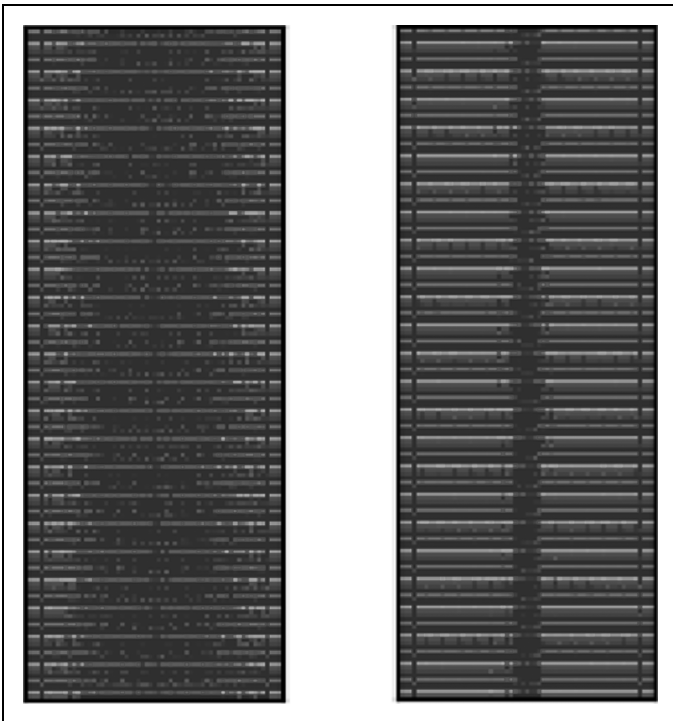
Fig. 5. Compressed images. The waveform images were run through the autoencoder framework and created these compressed images, which were then fed into the SVC classifier. One control (left) and one Parkinson's example (right) were randomly selected. SVC, support vector classification.

images with dimensions of 592 by 192 to 194 by 91 by 4. Convolutional autoencoders traditionally outperform regular autoencoders for image processing. An autoencoder seeks to recreate the input from a compact feature representation it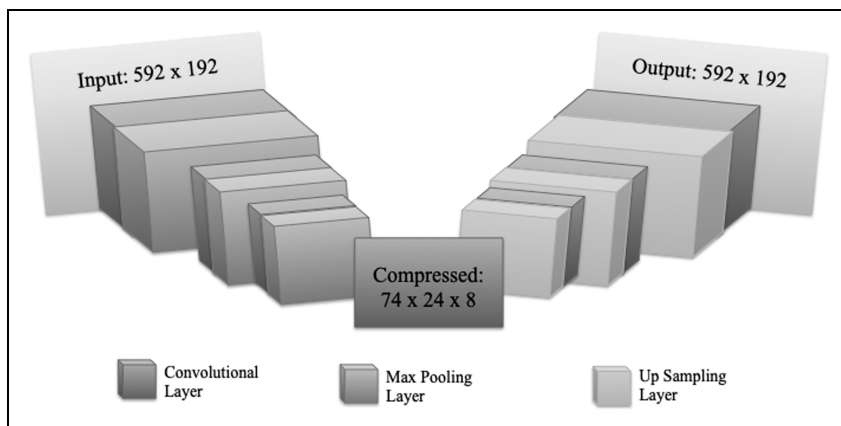 encodes. Through convolutions and pooling, the input is compacted and encoded and it is recreated by applying the reverse transformations through steps. The loss is defined as the difference between the recreated output and the initial input image. We extract the middle layer, which is significantly smaller than the original input. A randomly selected control compressed image and a randomly selected Parkinson's compressed image are seen in *Figure 5*.

Keras framework using the TensorFlow backend was used, and a detailed depiction of the steps deployed by the autoencoder is shown in *Figure 6*. The convolutional autoencoder had a 3×3 kernel with 1×1 strides to allow for overlap. To compact the image, the convolutional autoencoder used Max Pooling with a 2×2 kernel. *Figure 6* illustrates the entire method in further detail.

Different depths of convolutional autoencoders were tried for reducing the loss, which was minimized to 0.14 after 35 epochs when seven total convolutional layers were used.

Following the convolutional autoencoder, several different classification methods were implemented to classify the compressed feature representation as Parkinson's or control. A SVC method using a linear kernel attained a test set accuracy of 0.81, while a SVC with an RBF-Kernel accurately predicted 73% of the test set. While these are impressive results, they pale in comparison to the 99% achieved by our method explained in the Framework Section, which was minimized to 0.14 after 35 epochs when seven total convolutional layers were used.

Due to the high loss from the autoencoder, we also ran several methods on the grayscale standardized images without compressing the images. A logistic regression model was trained, which achieved a dismal test set accuracy of 0.46. A linear SVC accurately predicted only 52% of the test set. We also tried a convolutional neural network, which uses convolutions to analyze an image in larger frames of reference instead of a single pixel-by-pixel analysis. With 32 filters, a 3 by 3 convolution, and a 2 by 2 pooling size, the accuracy on the training set is 0.455.

Another method of feature extraction implemented was to extract the wave parameters, as previously mentioned. A Linear SVM and a RBF Kernel SVM were both independently run on these data and at best attained 55% training set accuracy. The computation time was also over 30 times greater than the method we proposed in the Framework Section. Accounting for the higher computational cost and the lower accuracy, we selected the MFCC model.

Fig. 6. The architecture of the Convolutional Autoencoder. It contains seven convolutional layers. The input is of dimensions 592×192 and through the convolutional and max pooling layers is transformed into a $74 \times 24 \times 8$ array. From this $74 \times 24 \times 8$ array, the input is recreated through convolutions and up sampling.

Conclusion

In this article, we proposed a method to predict Parkinson's disease from a 10-s audio clip of an individual saying "aaah" into a smartphone. Our method entails MFCC feature extraction, L1-based feature selection, and an RBF-kernel SVC and can run in real time on a smartphone. Experiments with 1,000 voice samples from the mPower dataset illustrate that our method achieves a 99% diagnostic accuracy. We also developed 11 other methods, including several deep learning methods and compared our method with them. Our method significantly outperformed all other state-of-the-art methods and is significantly faster than clinical diagnoses.

Acknowledgments

Disclosure Statement

No competing financial interests exist.

REFERENCES

1. Parkinson's Foundation. Statistics. [online]. **2018**. Available at: http://parkinson.org/Understanding-Parkinsons/Causes-and-Statistics/Statistics (last accessed August 3, 2018).

2. Schrag A, Ben-Shlomo Y, Quinn N. How valid is the clinical diagnosis of Parkinsons disease in the community? *J NeuroL Neurosurg Psychiatry* **2002;**73:529–534.

3. Holmes RJ, Oates JM, Phyland DJ, Hughes AJ. Voice characteristics in the progression of Parkinson's disease. *Int J Lang Commun Disord* **2000;**35:407–418.

4. Dodel R, Reese J, Balzer M, Oertel W. The economic burden of Parkinson's disease. *Eur Neurolog Rev* **2008;**3:11–14.

5. AAH works to get vocal information; Titze IR. *Principles of voice production*, 2nd ed. Iowa City: National Center for Voice and Speech, **2000**.

6. Asgari M, Shafran I. Predicting severity of Parkinson's disease from speech. *Conf Proc IEEE Eng Med Biol* **2010;**2010:5201–5204.

7. Galaz Z, et al. Degree of Parkinson's disease severity estimation based on speech signal processing. In: *2016 39th International Conference on Telecommunications and Signal Processing (TSP)*, Vienna, **2016:**503–506.

8. Zhan A, Mohan S, Tarolli C, et al. Using smartphones and machine learning to quantify Parkinson disease severity: The mobile Parkinson disease score. *JAMA Neurol* **2018;**75:876–880.

9. Tsanas A, Little MA, McSharry PE, Ramig LO. Accurate telemonitoring of Parkinson's disease progression by noninvasive speech tests. *IEEE Trans Biomed Eng* **2010;**57:884–893.

10. Arora S, Venkataraman V, Donohue S, Biglan KM, Dorsey ER, Little MA. High accuracy discrimination of Parkinson's disease participants from healthy controls using smartphones. In: *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Florence, **2014:**3641–3644.

11. Arora S, et al. Detecting and monitoring the symptoms of Parkinson's disease using smartphones: A pilot study. *Parkinsonism Relat Disord* **2015;**21:650–653.

12. Albert MV, Toledo S, Shapiro M, Kording K. Using mobile phones for activity recognition in Parkinson's patients. *Front Neurol* **2012;**3:158.

13. Challa KNR, Pagolu VS, Panda G, Majhi B. An improved approach for prediction of Parkinson's disease using machine learning techniques. In: *2016 International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES)*. **2016**.

14. Kostikis N, Hristu-Varsakelis D, Arnaoutoglou M, Kotsavasiloglou C. A smartphone-based tool for assessing Parkinsonian hand tremor. *IEEE J Biomed Health Inform* **2015;**19:1835–1842.

15. Khan T, Nyholm D, Westin J, Dougherty M. A computer vision framework for finger-tapping evaluation in Parkinson's disease. *Artif Intell Med* **2014;**60:27–40.

16. Tolosa E, Wenning G, Poewe W. The diagnosis of Parkinson's disease. *Lancet Neurol* **2006;**5:75–86.

17. Jain S, Lo SE, Louis ED. Common misdiagnosis of a common neurological disorder: How are we misdiagnosing essential tremor? *Arch Neurol* **2006;**63:1100–1104.

18. Meara J, Bhowmick BK, Hobson P. Accuracy of diagnosis in patients with presumed Parkinson's disease. *Age Ageing* **1999;**28:99–102.

19. Tsanas A, Little MA, Mcsharry PE, Spielman J, Ramig LO. Novel speech signal processing algorithms for high-accuracy classification of Parkinson's disease. *IEEE Trans Biomed Eng* **2012;**59:1264–1271.

20. Hartelius L, Svensson P. Speech and swallowing symptoms associated with Parkinson's disease and multiple sclerosis: A survey. *Folia Phoniatr Logop* **1994;**46:9–17.

21. Gamboa J, Jiménez-Jiménez FJ, Nieto A, Montojo J, Ortí-Pareja M, Molina JA, García-Albea E, Cobeta I. Acoustic voice analysis in patients with Parkinson's disease treated with dopaminergic drugs. *J Voice* **1997;**11:314–320.

22. Ramig LA, Titze IR, Scherer RC, Ringel SP. Acoustic analysis of voices of patients with neurologic disease: Rationale and preliminary data. *Ann Otol Rhinol Laryngol* **1988;**97:164–172.

23. "Mel Frequency Cepstral Coefficient (MFCC) tutorial." *Practical Cryptography*. [Online]. Available at http://practicalcryptography.com/miscellaneous/machine-learning/guide-mel-frequency-cepstral-coefficients-mfccs/#eqn1 (last accessed August 3, 2018).

24. "1.4. Support Vector Machines." *Scikit Learn*. [Online]. Available at: http://scikit-learn.org/stable/modules/svm.html (last accessed August 3, 2018).

25. mPower: *Mobile Parkinson disease study*. [Online]. Available at http://parkinsonmpower.org/ (last Accessed August 3, 2018).

26. Reeve A, Simcox E, Turnbull D. Ageing and Parkinson's disease: Why is advancing age the biggest risk factor? *Ageing Res Rev* **2014;**14:19–30.

27. "Parkinson's Disease: Diagnosis and Treatment." *Family Health UT Health San Antonio*. [Online]. Available at http://familymed.uthscsa.edu/geriatrics/readingresources/virtual_library/Parkinsons/Parkinsons06.pdf (last accessed August 2, 2018).

28. Hall MA. Correlation-based feature selection of discrete and numeric class machine learning. (Working paper 00/08). Hamilton, New Zealand: University of Waikato, Department of Computer Science, **2000**.

29. Shi JY, Wielaard J, Smith RT, Sajda P. Perceptual decision making "through the eyes" of a large-scale neural model of V1. *Front Psychol* **2013;**4:161.

Address correspondence to:
Wenyao Xu, PhD
Department of Computer Science and Engineering
State University of New York at Buffalo
338 Davis Hall
Buffalo, NY 14260-1660

E-mail: wenyaoxu@buffalo.edu