ParkinSense: Contactless Parkinson's Disease diagnosis, telemonitoring, and severity analysis via Machine Learning and Multimodal Digital Biomarkers

Pulkith Paruchuri

Heritage High School

Computer Science

Texas Junior Academy of Sciences Research Paper

Acknowledgment of Major Assistance

Most of the research was conducted virtually via computational methods for development. This stage of the study was conducted from June 2020 to the present date. Authentic patient data collection was collected from an in-house online data collection site and was collected from July 2022 to September 2022. Further data was collected during this same period from authentic patients from Andhra University under the supervision of Dr.Vallurupalli. Initial training and validation, and synthetic patient generation were conducted from datasets collected from the UC Irvine Public Dataset repository.

Further data was sourced from Professor Broadley from Griffith University, from the ParkTest team at the University of Rochester, and from Dr. Gupta at Duke University. We are grateful for all of these contributors.

We would like to thank the CardinalKit Lab at the Stanford Medical School for developing the framework to ensure ParkinSense abides by federal and state data privacy laws and for helping the ParkinSense team troubleshoot compatibility issues.

We would like to thank Dr. Sunday from Yale University, as well as the Yale University Institutional Review Board for sponsoring the collection of authentic patient data for this study so that human trials could be conducted, and for advising methods of approach to enlist volunteers to participate in this study.

We would like to thank the Ghosh Lab at the Massachusetts Institute of Technology for answering queries regarding the state and nature of Parkinson's Disease Research and treatment.

We would like to thank Roshan Patel and Satyen Dhamankar from the Webb Lab at Princeton University for reading over the initial results to propose areas for further study, despite being in different fields.

We would like to thank the Awatramani Lab at Northwestern University for providing results from Parkinson's Disease diagnosis trials on mice, which provided insight into promising solutions to explore similar diagnoses in humans.

Finally, we would like to thank the Neurose Team for working to make ParkinSense a global solution that is accessible to everyone.

Table of Contents

I. Abstract

Despite being the fastest growing Neurodegenerative disease in the world, with cases expected to surge to 30 million by 2030, Parkinson's Disease (PD) currently has no definitive diagnosis. Current methods of diagnosis are inaccessible to billions due to geographic, economic, and age-related mobility barriers, disproportionally so in rural areas or regions without ready access to healthcare, such as third world countries; PD diagnosis can cost upwards of $5K, require a lengthy visit to a nearby neurologist, and can involve intensive scans. Notwithstanding, currently, almost 1 in 3 patients are misdiagnosed, and an estimated 40% of patients with PD are never successfully diagnosed with PD. Since treatment is often more effective in the earlier stages of PD, these issues are vital to remedy so that PD patients can be diagnosed as early as possible. ParkinSense, a web application was developed to combat these issues, and understand whether Machine Learning can be exploited for accurate and rapid remote PD diagnosis and severity analysis, due to success in these regards with other disorders. ParkinSense allows one to diagnose themselves for PD from anywhere in the world in under 12 minutes, requiring only a webcam, microphone, and internet connection. ParkinSense analyzes 3 cardinal symptoms of PD entirely contactlessly and automatically: Dysarthria via sustained vocal phonations, Hypomimia via Facial Action Unit analysis on prolonged facial expressions, and Bradykinesia via Eye Blink Rate from a reading task. These symptoms were chosen, because they can be analyzed remotely, and have previously been shown to be significant symptoms in early PD. As such, Machine Learning was expected to be able to diagnose PD from these modalities. ParkinSense is able to complete automated feature extraction. ParkinSense is the first attempt at diagnosing PD via rapid Eye Blink Rate Analysis, and one of the first to do so via facial expression analysis. It was developed using recent innovations in Machine Learning, in conjunction with over 16 machine learning models that utilize ensemble learning and multimodal data fusion to improve accuracy. ParkinSense boasts an overall accuracy rate of 99.72% when trained on public databases totaling 9 million data points, synthetic patients, and 106 authentic patients. Furthermore, each individual modality(symptom), was able to achieve accuracy rates above 86%. Recall, a valuable metric to maximize in medical Machine Learning solutions was 0.999. Variational Autoencoders were utilized before training, to augment the dataset, with substantial accuracy gains. Even after proper diagnosis, there is a critical need to monitor PD severity, to analyze disorder progression and the effectiveness of certain treatments. ParkinSense is able to estimate the disease severity on the two common PD severity classification scales: the Unified Parkinson's Disease Rating Scale, and the Hoehn and Yahr Scale, with a relative mean absolute error of less than 4.5%, greatly outperforming current methods. Furthermore, trends of severity can be analyzed over time via a graphical interface, and results can be submitted directly to a neurologist. Not only is ParkinSense contactless, rapid, accurate, and cheap, it is the first public method of remote PD diagnosis and severity analysis, potentially providing billions with a crucial healthcare tool, that is otherwise essentially inaccessible to them. Further research can be conducted on unutilized extracted features, such as eye saccades, which could offer additional success.

II. Introduction

Parkinson's Disease (PD) is the second most prevalent neurodegenerative disease in the world [1]. With Parkinson's Disease being the fastest growing neurodegenerative disease globally [2], the number of cases has surged to over 10 million in the last few decades, and this figure is expected to double by 2030 [3]. Parkinson's Disease results from the death of dopaminergic neurons in the substantia nigra pars compacta [4]. This neurodegeneration affects largely those over the age of 40 [5] and leads to progressive disability [6]. Due to this progressive nature of PD, early and accurate diagnosis is vital for effective treatment, to stagnate or prevent further neuronic loss [7]-[9]. This neuronic loss affects both motor and non-motor skills [10]. Some of the hallmarks of the disorder include bradykinesia (slowed movement), speech irregularities, tremors, visual impairment,

postural instability, rigid muscles, and more [11]. Traditional and industry-standard diagnostic approaches typically take the form of extensive questionnaires that involve the manual examination and measurement of certain behaviors [12]. However, these are highly subjective to the administering clinician, relying on the detection of characteristics and behavior that may be too minute to be reliably noticed by one's eyes [13]. While a few technologies, such as Magnetic Resonance Imaging (MRI) and DaTScan scans can aid in PD diagnosis, they can offer little information beyond differentiating Idiopathic Parkinsonism, or Parkinson's Disease, from atypical forms of Parkinsonism or other diseases with similar symptoms [14]. As such, they can only be used to exclude other causes, as opposed to directly issue. Not only do many neurological disorders, such as Essential Tremor [14], share symptoms with PD, but many parkinsonism symptoms arise naturally in PD-susceptible, but healthy, individuals, as a result of old age, complicating PD diagnosis [15]-[17]. Resultantly, nearly 1 in 4 patients are misdiagnosed during a PD diagnosis at least once [18], severely inhibiting the effectiveness of potential treatments, and enabling the disease time to progress [19]. Furthermore, such a diagnosis typically is conducted in a hospital or institution with trained clinicians [20]. Coupled together, these difficulties essentially render PD diagnosis inaccessible to many, specifically those mainly in developing regions, or those without access to affordable healthcare [21]. The goal of this study was to create an accurate, rapid, and accessible method of PD diagnosis, and severity analysis using machine learning for multiple symptoms. This method is to be accessible via a web platform. The study contributions are multifold:

I. Multimodal Diagnosis PD can lead to accuracy gains when compared to unimodal diagnosis tools.

II. Aggregating multiple state-of-the-art classifiers simply via late fusion is successful in improving model performance.

III. PD diagnosis to be accurately conducted entirely non-intrusively and contactlessly for cheap.

IV. The world's first PD diagnosis and severity analysis software were developed that can be deployed in an engaging and accessible manner worldwide

V. EBR is a promising modality for PD diagnosis.

VI. PD diagnosis can be done rapidly and in an automated manner with server-side feature extraction.

VII. PD severity analysis can be conducted contactlessly and accurately, using a similar feature set as the diagnosis models.

III. Related Works

With technology significantly emerging into the healthcare industry over the past few decades, a great deal of literature has been published regarding PD diagnosis with the assistance of computational resources. Multiple approaches have been tried, including things such as Wearable Technology and MRI scans [22], [23]. However, a proven and accurate method of diagnosis has been elusive. A promising portion of this healthcare-technology convergence is Machine Learning. Machine Learning has been seen to be effective in signaling PD in an individual, by analyzing data such as vocal irregularities, MRI scans, sebum RNA, and much more [24]- [26]. For the purpose of this study, a focus was placed on machine learning methods for PD diagnosis that can be conducted contactlessly (e.g. without wearable technology, or MRI scans).

While numerous symptoms are present in PD patients, three were chosen for this study, stemming from their ability to easily and accurately be assessed noninvasively, with minimal technical requirements, ensuring accessibility while retaining reliability.

3.1 Dysarthria

Dysarthria categorizes the vocal irregularities deriving from the hypokinetic nature of parkinsonism [27]. It can cause abnormalities and monotonicity in intonation, tone, volume, fluidity, and more [28].

Several studies have attempted to identify vocal irregularities via Machine Learning for PD diagnosis, and it was found that sustained vowel phonations, of /a/, for example, have proven to be effective in differentiating PD patients from Healthy Controls (HC) [29], [30].

3.2 Hypomimia

Also referred to as 'Facial Masking', Hypomimia is the reduction in pronunciation, and rapidness of facial

TABLE I PUBLIC DATASET SOURCES & DESCRIPTIONS

Dataset	Size	Features	Participants PD	Participants HC	Mean Age	% Male	Content Utilized	H&Y Scale	UPDRS Score
A	240	46	40	40	N/A	60%	/a/ phonation n	N	N
B	1040	26	20	20	64.9±9	60%	/a/, /o/, /u/ phonations, words, short sentences	N	Y
C	168	26	28	0	62.7±11	N/A	/a/, /o/ phonations	N	N
D	5875	26	42	0	65.4±9.24	67%	/a/ phonation	N	Y
E	756	754	188	64	64.1±10.6	52%	/a/ phonation	N	N
F	1812	12	61	543	63.9±7.8	38%	AU Variances	N	N
G	61	10	20	41	70.7±9.1	41%	Reading EBR	Y	N

expressions, arising from the rigidity many PD patients encounter in their facial muscles [31].

The human face can be partitioned into groups of Action Units (AU), that represent groups of facial muscles [32]. When making certain facial expressions, such as disgust or surprise, numerous AU's are activated, and the variance of the magnitudes of this activation quantifies the amount of muscle movement, which in turn, can divulge information about the presence of PD [33]. Reference [33] discovered that Hypomimia is a promising method for PD diagnosis.

3.3 Bradykinesia

Bradykinesia describes the slowness-of-movement characteristic of many PD patients, and Eye Blink Rate (EBR) can be used as a test for Bradykinesia [14].

Although no one in literature has attempted to use Machine Learning on EBR for PD diagnosis, numerous studies have shown statistically significant differences between the EBR of PD patients, and HC, providing a promising space to exploit for PD diagnosis [34]. During a reading task, it was found that PD individuals averaged only 2.4 blinks per minute, while HC averaged 10.7 blinks per minute [35].

3.4 Multimodal Diagnosis

While many classification algorithms with Machine Learning can perform quite well on PD diagnosis, they often rely on only one symptom, as can be seen in sections 3.1-3.3. While this can be effective, the onset

and severity of many PD symptoms vary greatly as some symptoms may never manifest in a PD patient, while others typically only manifest noticeably in the middle to late stages of PD after 20 years [36], [37]. As such, it is unreliable to use only unimodal classifiers. Furthermore, past multimodal studies, for not only PD diagnosis, but the diagnosis of other neurodegenerative diseases, have shown that multimodal models can offer significant performance improvements over unimodal classifiers [38], [39].

3.5 Severity Analysis

The Hoehn and Yahr Scale (H&Y Scale) is a 5-point scale [40], and the Unified Parkinson's Disease Rating Scale (UPDRS) is a 199-point scale [12], widely used for quantifying disease severity and progression. Both of these scores are conventionally calculated through lengthy questionnaires. that rates a PD patient's level of disability and disease progression, from the evaluation of dozens of tasks by a clinician. A few have managed to estimate disease severity on these scales from symptoms with Machine Learning effectively [41].

IV. Data Sources

Data was gathered from a variety of sources. All data was gathered or collected from sources that obtained data in scientifically rigorous environments and methods. An outline of public data sources, content, size, and other information can be found in Table 1. Along with the information explicitly stated in Table 1, most datasets also contain demographic information (but no identifying information) regarding the participants, such as age, gender, and how long they've had PD for those who tested positive for PD.

4.1 Dysarthria

Datasets A-E of Table 1 describe the unimodal data collected for the Vocal Analysis Modality. These datasets contain information on various vocal features, such as fundamental frequency, variations in this frequency(absolute shimmer), and the ratio of noise to tonal components. These features are collected as the user vocalizes sustained vowel phonations, typically for 10 seconds. Certain datasets were aggregated together, the first time this has been attempted in literature for Parkinson's Disease diagnosis. The features used for most vowel phonations represent the intersection of the features in the datasets for that specific phonation.

4.2 Hypomania

Facial Expression Data was collected from the study in [33], and from the ParkTest team at the University of Rochester. This data is shown as a single row in Dataset F in Table 1, since the data in [33] was also collected from the ParkTest team, and contains information regarding the variance of the magnitude of the corresponding Action Unit involved with Smile, Surprise, and Disgust facial expressions, during activated phases.

4.3 Bradykinesia

Eye Blink Rate Data, sourced from [35] is represented in Dataset G of Table 1 and provides information about the Eye Blink Rate of participants during a reading task, a conversation task, and a television watching task. For the purpose of the study, only data from the reading task was utilized. Half Blinks, a phenomenon in many PD patients, where the eye may twitch and cause a blink that only partially obscures the eye, was also collected. Finally, it was noted whether or not the patients wore glasses during the study.

4.4 Authentic Patients

An online data-collection platform, which can be found in the Appendix, was developed to collect data from authentic patients, without requiring the patients to commute to a facility or clinic for data extraction. This also allows a mimic of how data will be collected and analyzed with the diagnosis site. After approval was gathered from the Yale Institutional Review Board for human trials, 124 Healthy Controls (including individuals who were diagnosed with other neurodegenerative diseases, but did not develop PD), and PD patients were recruited from Andhra University**,** and online via advertising. Data was collected as they completed the process as outlined in section 5.2, and their PD diagnosis was recorded, as with their age, gender, H&Y rating score, and UPDRS score as applicable. Care was taken to preserve the

anonymity and privacy of the patients. The only information that was collected that could be used to identify the patient was data collected from the videos and audio clips submitted by the patient (but no name, address, email, etc… was collected). However, after the data was featured and converted to a numerical format, all the identifying data (facial expressions, vowel phonations, etc…), was discarded. The only individual who worked with the identifying data in order for it to be featured was the researcher itself, and as a result, no other individual was in contact with the identifying information.

4.5 Severity Analysis

A few of the datasets contained information regarding the severity of PD patients, which is used for the severity profiling stage of ParkinSense. The specific Datasets which contain such information can be found in Table 1. Datasets either contained the severity on the UPDRS or the H&Y scale.

V. Methodologies

5.1 Website Overview

Two websites are involved in this study. The Data Collection Site (DCS) labels the website that was used to collect data from authentic patients and returned no information to the participant. The Diagnosis & Severity Analysis Site (DSAS) is the website developed as the product of the study, allowing individuals to diagnose themselves for PD, and if a positive diagnosis is returned, classify their severity on two rating scales. Both sites are nearly identical in the layout and process for providing information / getting diagnosed, so features that allow for the ease of use in the DCS site can be translated over to allow for similar ease of use in the DSAS site.

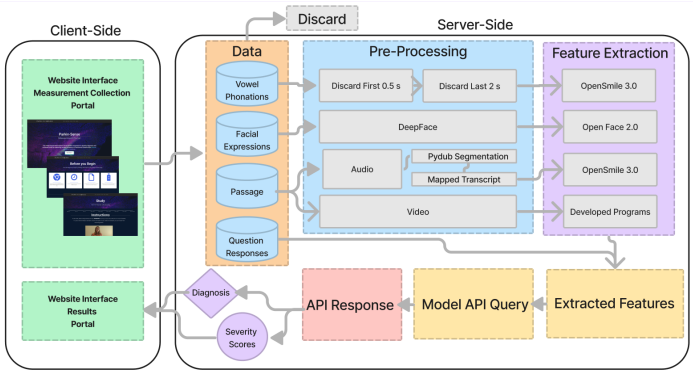

Fig. 1. Web application overview, data processing methods, and feature extraction technologies as described in sections 5.2-5.4.

5.2 Data Collection Site

A diagram of the DCS site, which is to be further described in this section, can be found in Fig. 1. The specific DCS site host link can be found in the appendix.

When users navigate to the web application and begin the diagnosis, they are first asked to be in a seated position, with their laptop or computer one foot away from them, such that their entire head and part of their torso are in the frame, not dissimilar from a passport photo. They are requested to complete the study at a time when they are well-hydrated and not tired. Finally, they should be in a quiet location, with their face being well-lit.

Before they begin the diagnosis, they are prompted with a screen that allows them to adjust their camera and microphone volume, so they are able to be in the optimal position relative to the above conditions when they begin the diagnosis.

1) *Vocal Analysis:* When users begin the diagnosis through the web application, they are first asked to vocalize sustained vowel phonations, for as long as they can, or until 10 seconds elapse, whichever is shorter. It was emphasized that they should maintain constant volume and pitch. They are requested to complete three such phonations: /a/, /u/, and /o/

2) *Facial Expressions:* Then the user is asked to make 3 pronounced facial expressions: joy, surprise, and disgust. To improve accuracy users were asked to make each face three times in succession, alternating with a neutral face.

3) *Short Passage:* In the next section of the diagnosis, users were asked to read as much as they could of a selection of 5 passages for 60 seconds. They were asked to read at their regular reading speed, volume, and voice. Text spanned 90% of the viewport width, centered on the screen, at a font size of 15 pixels. The specific reading passages were chosen for

having a diverse set of words, ranging from short to long, monosyllable to polysyllabic, containing most sounds in the English Language, and suspected phrases that might divulge important information for the machine learning models to pick up on. The specific passages can be found in the link in the Appendix. Paragraphs were ordered in order of expected relative importance, since users may not have time to read all five paragraphs in the allotted time with their regular reading speed. From this section, data for 2 of the modalities were collected. Specifically, data was gathered for vocal analysis, to be split up into small sentences, phrases, and individual words. Furthermore, video data was collected for EBR analysis.

4) *Questionnaire:* The final section of the diagnosis involved a short questionnaire, asking a few questions including gender, age, and whether the individual wore glasses during part 3 (Short Passage).

5.3 Preprocessing

Data processing and Feature Extraction was done entirely through an automated pipeline, to mimic the requirements for automated data processing when used as a web application for rapid diagnosis. Certain requirements, such as a decibel ceiling and floor were enacted to prevent data from which it was difficult to extract features accurately to transform into a form of adversarial data that could lead to an incorrect diagnosis. This step of preprocessing led to the loss of some participants from the authentic dataset, reducing the functional size of the dataset from 124 to 106.

1) Vocal Analysis: First vocal data was separated from webcam data, and webcam data was discarded. Additionally, the recording was trimmed, such that the first 0.5 seconds of the recording were discarded, since the onset of speech results in nonuniform levels of pitch and volume. Additionally, the last 2 seconds were thrown out, since, during this timeframe, the user typically runs out of breath, resulting in diminishing volume and degrading pitch that does not represent the user's true vocalization.

2) Facial Expressions: Each participant submits 3 videos, each containing 6 facial expressions (3 of the requested facial expression, and 3 neutral faces), for a total of 18 facial expressions(9 total facial expressions for smile, surprise, and disgust). For facial data, each of the 18 facial expressions was run through DeepFace [48] to ensure the correct facial expression was made. If the incorrect facial expression was detected, the recording was discarded.

3) Short Passage: For data derived from the reading passage, vocal and webcam data were split from each other. Vocal data was split up into individual words using the Pydub module [50], which was temporally paired with the words in the passage, to create a mapped transcript.

4) Data Augmentation: Because of the relatively smaller size of the EBR dataset, even relative to the unaggregated datasets of the Vocal Analysis Modality, Data Augmentation techniques were used to augment the size of the dataset. In this case, because the HC class was quite larger than the PD class, the Synthetic Minority Oversampling Technique (SMOTE) was used to generate new examples for the PD class,

5.4 Feature Extraction

1) Vocal Analysis: Features were extracted from vocal data that represents the intersection of Features in datasets A-E. These features were extracted using openSMILE 3.0 [47].

2) Facial Expressions: With a server-hosted instance of OpenFace 2.0 [49], the magnitude of the AU's associated with each of the 3 requested facial expressions were calculated when the expression was being made, and the variance of these AU magnitudes was calculated, representing a metric of the amount of facial movement. Recordings were split at the midpoint troughs of AU non-activation, indicating a neutral face.

3) Eye Blink Rate: Because of the need to measure half-blinks in the study, no suitable pre-developed was found that was able to calculate the measurements needed for this study. A solution was built instead to count these numbers of eye blinks.

First, we landmark the face to localize important regions of the face, and to derive 6 points around the perimeter of the eye. p_1 and p_4 are on the left and right corners of the eye, p_3, p_2, p_6, and p_5 are along the center of the circumference of the eye in Quadrants 1, 2, 3, and 4 respectively. Then in this landmarked region, we sample a small number of points to

determine the Eye-Aspect-Ratio (EAR), determining a ratio of the eye height to eye width as calculated by:

$$EAR = \frac{\|p_2 - p_6\| + \|p_3 - p_5\|}{2\|p_1 - p_4\|}$$

For this study, an EAR equal to or below 0.20 was counted as a full blink, and an EAR below 0.30 was determined to be a half-blink.

However, it was found that EAR was unreliable when detecting full blinks due to varying eye sizes. As a result, another metric was defined, the Eye Coverage Ratio (ECR), which calculates the amount of eye, bounded by the 6 landmarked points, that is covered by the eyelid. An ECR above 0.9 coupled with an EAR below 0.2 was counted as a full blink.

To calculate this ECR on the eye region, a few approaches were attempted. Simple Monte Carlo Estimation, where a large number of random points are sampled, and the estimation of the ECR can be calculated. However, since the ECR had to be calculated for hundreds of frames a second, this was deemed unfeasible. Then a predetermined sample of points was to be used, but due to the depth and non-spherical shape of the eye, it was difficult to determine a uniform sampling. As a result, points were generated with the constraint that they are well-spaced and roughly uniform. To accomplish this, the eye was generalized as a 2-dimensional horizontal plane. Then each candidate point, z, was normalized on this plane:

$$z \in [0,1]^M$$

Where $M = 2$. Then a maximin construction was attempted, which sequentially generates points to maximize the distance of a point to all its neighbors. At a point where k points have been generated, this can be described as satisfying the condition below to generate the $(K+1)$th point $z^{(k+1)}$

$$\text{Maximize} \min_{i=1}^{k} \left\|\mathbf{z}^{(k+1)} - \mathbf{z}^{(i)}\right\|$$

$$\text{subject to} \sum_{m=1}^{M} z_m^{(k+1)} = 1$$

$$z_m^{(k+1)} \geq 0, \quad m = 1, \ldots, M$$

However this approach was found to generate a proportionally surplus number of points on the perimeter of the eye-plane, and was thus abandoned. In its place, Reisz S-energy, a generalization of potential energy was used. In this, the potential between two points z^i and z^j is defined as

$$U\left(\mathbf{z}^{(i)}, \mathbf{z}^{(j)}\right) = \frac{1}{\|\mathbf{z}^{(i)} - \mathbf{z}^{(j)}\|^s}$$

With the overall s-energy of a region being written as:

$$U_T(\mathbf{z}) = \frac{1}{2} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \frac{1}{\left\|\mathbf{z}^{(i)} - \mathbf{z}^{(j)}\right\|^s}, \quad \mathbf{z} \in \mathbb{R}^{n \times M}$$

We then aim to minimize U_T. To reduce the computational cost, we take the logarithm of U_T, and compute the partial derivative $F_E = log(U_T)$ with respect to $z_m^{(i)}$ as shown below:

$$\frac{\partial F_E}{\partial z_m^{(i)}} = -\frac{d}{U_T} \left[\sum_{\substack{j=1, \\ j \neq i}}^{n} \frac{\left(z_m^{(i)} - z_m^{(j)}\right)}{\left\|\mathbf{z}^{(i)} - \mathbf{z}^{(j)}\right\|^{s+2}} \right]$$

3,000 iterations of Adam gradient-based optimization iterations were performed to generate a sample of well-spaced points to calculate the ECR rapidly. This approach proved to be successful both with and without glasses.

5.5 Website Deployment

The diagnosis website was developed in React.js and deployed on a personal domain with a signed certificate. Data is collected client-side and processed server-side. The unprocessed video and audio data was stored server-side, then discarded after being processed to protect user privacy. The

Server-side was written in (PHP), and connected to the front-end through (AJAX) requests. The trained models were created in Python via Scikit-Learn, hosted in Amazon Web Services (AWS), and accessed by the server through an AWS Gateway API.

5.6 Model Framework

A diagram of ParkinSense's Diagnosis Model Framework can be seen in Fig. 2. ParkinSense's diagnosis framework revolves around the use of multiple classifiers to provide a binary classification of PD presence in a subject.

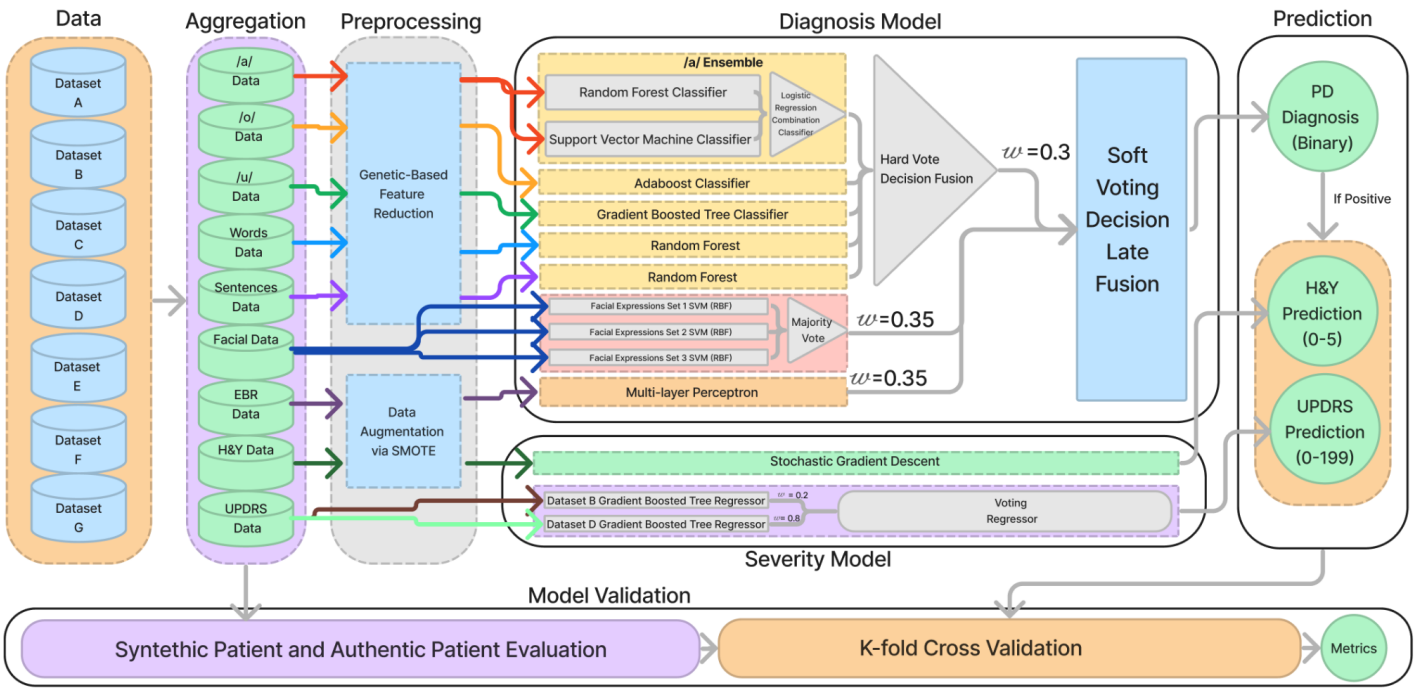

Fig. 2. Data sources, aggregation categories, preprocessing, and the model pipeline. The pipeline includes the diagnosis model, severity analysis model, and validation on synthetic patients. *w* represents the weights of certain estimators when used in voting classifiers and regressors. Severity is only evaluated following a positive diagnosis.

1) Dysarthria: Given the ample literature on Vocal-based methods of PD diagnosis, numerous classifiers and models have been well-tested and documented. In this step, the feature set was reduced to 10 using Genetic Algorithm-based feature set selection for all datasets. A Support Vector Machine (SVM) and Random Forest were used via Ensemble Learning with a Logistic Regression combination classifier to provide a single binary classification for /a/. For /o/ and /u/, an Adaboost Classifier and Gradient Boosted Classifier were used respectively, as they provided the best performance out of the tested classifiers. Finally, a random forest was used for both words and short-sentence classification. A final binary classification for Dysarthria-based methods was generated through hard majority-voting decision fusion, from the 5 individual classifiers. Weighting individual classification during the decision fusion stage did not seem to outperform equal weighting in this study.

2) Hypomimia: Three sets of 3 facial expressions were recorded for each user. A Support Vector Machine with a Radial Basis Function Kernel was used on each set and combined with a majority voting decision fusion to generate a single binary classification for this modality.

3) Bradykinesia: For EBR, a Supervised Multi-layer Perceptron was used to predict a binary classification.

4) Modality Decision Fusion: To combine the decisions from each modality, Soft Voting was used, with weights of 0.30, 0.35, and 0.35 used for Dysarthria, Hypomimia, and Bradykineia respectively. In the case of the public model, if the confidence of the algorithm was low, the application did not return a diagnosis to prevent a misdiagnosis. However, for the purpose of the study, the diagnosis was always returned irrespective of the diagnosis confidence.

5) Severity Analysis: In the case of a positive diagnosis, we predicted two scores: H&Y and UPDRS. H&Y Score was predicted via Stochastic Gradient Descent Regression using EBR data. UPDRS scores were predicted from datasets B and D individually, using Gradient Boosted Tree Regressors, and used a voting regressor, with 80% weight assigned to dataset B, due to its larger size.

6) Evaluation: Due to the difficulty in finding data, especially on all modalities simultaneously for overall model evaluation, high-fidelity synthetic patients were generated as described in section 4.7, with specific modality instances being paired by factors such as gender, age, and disease severity. Synthetic patients can provide near identical results when compared to live patients [52]. This data was combined with authentic data. K-fold cross-validation was used for all model evaluations, coupled with metrics discussed in section 6.

5.7 Synthetic Patients

To test the effectiveness of the multimodal model, over the individual unimodal classifiers, data had to represent individuals with data for all modalities. However, no such data exists due to the novelty of this study. As a result, either new data had to be collected, or synthetic data had to be generated. In this case, both of these were entertained.

Synthetic data was generated by using variational autoencoders. The variational autoencoders took the original dataset for each modality, mapped it to a lower latent space, and remapped it back to the higher latent space. As a result of this mapping, patients which were similar, but slightly different from the original modality dataset were generated. From this, full patients were generated across modalities, by matching data instances from each modality that were most similar in age and gender, and in the case of PD patients, severity, and years since diagnosis.

Gender and PD diagnosis were strict matching criteria, meaning that a data instance from one modality had to equal the gender and PD diagnosis of the paired data instance from another modality. The other metrics, such as severity, gender, and diagnosis, were loose criteria and were matched as described below.

As the matching criterion, a data instance from one modality was different from another modality by the propensity score. For each modality instance, its loose demographic information was normalized between 0 and 1, and the propensity score between x_i and x_j is defined as:

$$\delta(x_i, x_J) = \sqrt{(x_i - x_j)' S^{-1} (x_i - x_j)}$$

With x being the $z * 1$ vector containing the value of the z included covariates for a specific modality, S is the identity matrix to disregard scaling, and S' is the generalized inverse of S.

From this, an optimal matching was generated that ideally generated high-fidelity synthetic patients.

Furthermore, authentic patients were utilized to collect ground truth data that was not generated. These authentic patients also provide a testament to the effectiveness and resemblance of synthetic patients.

Finally, authentic patients were used to testing the effectiveness of an active machine learning paradigm. In one trial, after every 5 iterations of the diagnosis on the authentic patients, the individual dataset modalities will be augmented with the true results of the authentic patients that have been diagnosed so far, and the model will be retrained.

This is how the researchers plan the final website product to work as well. After a successful diagnosis, the participant can choose to get a diagnosis from their neurologist and can submit this official diagnosis to ParkinSense. After a specific number of these official diagnoses have been submitted, the model will be retrained with the new data, allowing for an increase in accuracy over time. However, the model will only be retrained on data where ParkinSense and the official diagnosis do not differ, as in cases where they do differ, it cannot be deduced whether the misdiagnosis was from the official diagnosis or from ParkinSense.

VI. Results & Discussion

In most cases, synthetic data, and authentic data are evaluated to quite equal metrics, especially considering the difference in the number of instances between the two. As a result, the metrics reported in sections 6.1 and 6.2 represent the aggregation of both the synthetic and authentic data to present a more accurate representation of the model.

6.1 Model Diagnosis

Accuracy (*a),* precision (*p*), recall(*r*), F-1 score (*f*), specificity (*s*), and negative predictive value (*n*) scores were used as the metrics for this task. The formulas for these metrics are defined below:

$$a = \frac{TN + TN}{TP + TN + FP + FN}$$
$$p = \frac{TP}{TP + FP}$$
$$r = \frac{TP}{TP + FN}$$
$$s = \frac{TP}{TP + FN}$$
$$S = \frac{TN}{TN + FP}$$
$$f = \frac{2 * TP}{2 * TP + FP + FN}$$
$$n = \frac{TN}{TN + FN}$$

Where (TP) defines the number of True Positives, (FN) defines the number of False Negatives, (TN) defines the number of True Negatives, and (FP) defines the number of False Positives.

TABLE II CLASSIFICATION EVALUATION METRICS

Model	*a*	*p*	*r*	*s*	*f*	*n*
Vocal Modality	0.963	0.947	0.985	0.939	0.966	0.982
Facial Modality	0.958	0.909	0.976	0.941	.941	0.987
EBR Modality	0.860	0.714	0.938	0.824	0.811	0.966
Overall Model	0.997	0.994	0.999	0.996	0.997	0.999

The results for each modality, specific internal classifiers, and the overall model, can be seen in Table 2. In large, all modalities had accuracy rates above 86%, with vocal analysis being the most accurate, and EBR being the least accurate.

The high value of *r* is important, as it is vital to reduce false negatives as much as possible in the medical diagnosis industry. Additionally, the high value of *r* is a testament to the accuracy of ParkinSense, since the HC class was quite larger than the PD class, yet ParkinSense was able to achieve both a high *a* and a high *r* simultaneously.

SMOTE data augmentation led to an increase in accuracy from 0.73 to 0.86, in agreement with previous uses of SMOTE augmentation.

Additionally, features were ranked according to their predictive value (i.e. a feature with a higher predictive value is more vital for an accurate diagnosis), although these figures will be omitted for brevity.

6.2 Severity Analysis & Telemonitoring

Mean Absolute Error (MAE), which can be calculated as

$$MAE = \frac{1}{N} * \sum_{i=i}^{N} |x_i - x|$$

where N represents the number of validation instances, x represents the actual severity value, and x_i represents the estimated severity value, for a specific i.

TABLE III SEVERITY EVALUATION METRICS

Model	MAE
Overall H&Y Prediction	0.43
Dataset B UPDRS Prediction	18.6
Dataset D UPDRS Prediction	12.2
Overall UPDRS Prediction	9.7

The results from the severity regression models can be found in Table 3.

Overall, ParkinSense's severity analysis model can be used to quite accurately predict both UPDRS and H&Y rating scale scores with a quite significant accuracy. This allows users to monitor disease progression and medication effectiveness over time. Although an analysis of the effectiveness of this telemonitoring feature could not be evaluated as no longitudinal observation studies were conducted for a few years, it can be reasonably inferred they would be quite practical. Furthermore, users can submit their past and current severity scales to a neurologist for expert analysis or for benchmark comparison.

6.3 Comparative Analysis

ParkinSense's individual classifiers outperform the vast majority of unimodal diagnosis models, in literature [48], especially the vocal modality, likely owing to the dataset aggregation, synthetic patient generation, and use with authentic patients. ParkinSense as a whole outperforms all unimodal models when tested on datasets larger than 25, all multimodal models, and even models that use invasive or nonremote data, such as MRI scans and extracted DNA sequences [53]. Additionally, ParkinSense outperforms the vast majority of models for severity analysis on both the UPDRS and H&Y severity prediction, although this phenomenon is much more apparent with the UPDRS. Additionally, if we limit analysis to models that only utilize data that can be collected remotely and noninvasively, ParkinSense provides massive improvements over current prediction models for UPDRS estimation.

6.4 Website Analysis

All methods of data collection, feature extraction, and model querying via the web application were validated to be effective and successful when done in an automated manner. Data collection depended on the user but usually took under 6 minutes. Pre-processing and Feature extraction usually took 4 minutes, and prediction via the model usually took 2 minutes. As such the goal of a rapid diagnosis with ParkinSense was achieved, with the entire diagnosis taking under 12 minutes, in sharp contrast to current industry methods that can take days to weeks. Therefore, ParkinSense suggests that an entirely contactless and accurate diagnosis via web-and-microphone-derived biomarkers is quite effective.

Additionally, users could track and compare their severity over time on both scales, via a Graphical User Interface that involves graphs and plots.

Furthermore, since the target demographic of the website is quite different in age than those usually quite experienced in technology, especially computer and web applications, it is important to ensure that the target demographic is able to easily and accurately use the website. As part of the end-of-study survey, the web application was rated *4.8 out of 5* in terms of ease of use. Although this rating was from the data-collection site, the actual diagnosis site is quite similar in layout and process, so the ease of use likely carries over.

We found no notable performance improvements when using an Actv Larrning Paradigm, although this was likely due to a small authentic sample size.

VII. Conclusion

7.1 Error Analysis:

The largest source of error in correctly diagnosing a patient for PD is the EBR classifier. It had the least number of data points out of all the classifiers, and data had to be augmented for its use in the overall model. While this proved to be effective, it may have introduced certain biases or errors that may only be observable in large validation trials.

Additionally, for dataset aggregation in part 4.1, certain features likely were not collected in similar manners for each dataset, which may have introduced some noise that made the model underperform. Furthermore, during this aggregation, some features were discarded in favor of obtaining a larger dataset that was compatible with more child datasets. As a result, this may have discarded some features that had high predictive power in differentiating PD patients from HC.

Finally, some errors in the model may be a result of some authentic patients, testing negative for PD, but also have some other disorder, perhaps a neurodegenerative one, that could share similar symptoms to PD or characteristics prevalent in HC. Because of the high level of r, the model tends to be biased towards a positive PD diagnosis, which may result in an error, when the individual does not have PD and instead has a disorder that imitates a few symptoms of PD. This may even arise from the individual modality datasets as well. However, rather than an 'error' of ParkinSense, it would be more favorable to rather label it perhaps as a 'feature', since during an actual diagnosis, it is very difficult to impossible to ensure that the participant does not have any other neurodegenerative disorder that could hinder diagnosis or lead to a misdiagnosis.

7.2 Future Work and Applications

One quite promising approach involves measuring saccades, or rapid eye movements[55]. Individuals with PD often have oculomotor abnormalities in regard to saccades that can be measured and analyzed [55]. While this approach may be difficult to implement with the typically low-quality webcams found on most computers (relative to gaze-tracking headsets and technologies), [55], found saccade analysis to be an effective method of PD diagnosis when the video was recorded on a mobile phone. Saccade analysis may hold the key to unlocking another modality that can be tapped into for remote PD analysis.

Indeed, in this study, saccades were analyzed as a possible addition to the multimodal application. However, a suitable approach was not developed for use in ParkinSense. In the regard that future researchers may be able to build upon this line of methodology, the approach used by the researchers in this study will be described below.

Since saccades represent rapid eye movements, a test of some sort must be developed. For this study, a black ball moves across a white screen, alternating directions when it nears the edge of the screen. While little research has been conducted on this topic, a recent study proposes that single-eye horizontal saccades may be of significant interest. In this case, when the ball changes direction, one eye deviates from the previous gaze of both eyes and follows the direction of the ball. On the other hand, the other eye will sometimes maintain the previous direction, slowing down in the direction that opposes the movement of the ball. This will result in tremors in a gaze that indicate poor motor control. Because of this, horizontal single-eye saccades were the main goal of this modality of the study.

The difficulty in this approach is accurate gaze estimation. First, the user's face is landmarked and the pupil's extracted. Then the user is requested to look at all four corners of the screen in succession to calibrate the algorithm. From here the coordinate for the y direction can be easily calculated by calculating the Euler angles. For a specific gaze, z, and a set of calibration points $\{TL, TR, BL, BR\}$ indicating the coordinates of the pupil at the top left, top right, bottom left, and bottom right of the active area during calibration, a specific gaze estimation for y can be calculated as:

$$gaze_y = (z_y - TL_y)/(BL_y - TL_y) * SH + ST$$

Where SH represents the height of the active area which can be easily measured via the website and ST measures the distance from the top of the active area to the top of the screen. A similar method can be conducted for the x axis.

However, this method suffers from inaccuracy when the user's face is not aligned perpendicular to the plane of the camera, and due to the low-quality resolution of most commercial laptop webcams.

Furthermore, the smooth pursuit can also be estimated from an application like ParkinSense. Smooth Pursuit aptly describes the movement of the eyes in a smooth motion, in contrast to saccades.

Both of these approaches can be added to the current diagnosis application, without significantly altering the number of steps for the diagnosis itself. As the participant reads the passage, the movement from left to right across the line the movement of the eye can be generalized to smooth pursuit. On the other hand, as the participant jumps from one line to the next, and moves the eye quickly to the start of the next line, the behavior can be compared to that of a horizontal saccade, albeit with a slight vertical movement.

Furthermore, certain features, such as speech rate, blink duration, blink latency, reading rate, and head tremors that could be calculated from data collected in an environment similar to ParkinSense, also provide multiple other opportunities, that have had success in PD diagnosis in the past in a labs setting, and could be combined with ParkinSense for greater accuracy rates.

Additionally, an active learning paradigm can be used to increase dataset size, and hopefully accuracy, by resubmitting consenting users' data into the dataset.

The researchers with this study, aim to test this software, combined with other modalities, for longitudinal clinical tests on live patients in the near future to ensure the telemonitoring feature is functioning as expected. While it is expected that synthetic data is quite similar to live patient data, it is imperative to verify such metrics, especially in a field as unforgiving as medical diagnosis. If such trials are a

success, ParkinSense would be open to the public for beta access. In the meantime, we will release ParkinSense without the severity analysis feature, pending approval from federal and state regulatory agencies.

Finally, other approaches to PD diagnosis, or other neurodegenerative diseases, and diseases with cardinal behavioral symptoms can benefit from a similar contactless diagnosis framework. By analyzing symptoms rapidly, accurately, and entirely contactlessly, many diagnosis tools can broad-base their technology to millions more individuals across the world, who do not readily have access to institutional or medical-facility instruments and technology.

7.3 Final Remarks

ParkinSense makes encouraging progress in the development of PD contactless diagnosis and severity analysis software via machine learning technologies and multiple modalities. ParkinSense automates data collection and feature extraction on the web application, and regarding the unimodal models, performed similarly to many top models. Furthermore, we introduce the first model that involves PD diagnosis classification through EBR. Additionally, we developed UPDMS and H&Y Scale predictors that perform quite well. ParkinSense integrates all of this together, to develop a PD diagnosis and severity analysis application, that outperforms most unimodal and multimodal diagnosis tools in literature for PD diagnosis and severity and progression metrics. All of this is accessible through a web application that is accurate and rapid. ParkinSense is the first of its type to offer such rapidness, and accuracy, for free, and entirely remotely. ParkinSense is also the first application to offer severity tracking and submission, which may prove to be vital when prescribing and using medications.ParkinSense juxtaposes current methods, which are quite invasive, requiring scans, injections, and hours of clinical diagnostic tests, and are also quite expensive, costing thousands of dollars. This makes it out of reach for billions of people, an issue ParkinSense aims to resolve. ParkinSense suggests that contactless PD diagnosis and severity tracking can be conducted accurately and rapidly, which may prove invaluable for those without access to PD diagnosis due to barriers, immobility, or associated costs. The significant accuracy of ParkinSense demonstrates the effectiveness of multimodal machine learning algorithms, data augmentation, and synthetic patient generation, and may allow millions worldwide to be diagnosed correctly, ensuring effective medication can be distributed immediately, preventing PD progression, as well as improving the quality of life for PD patients.

VIII. References

[1] L. Shen, “Gut, oral and nasal microbiota and Parkinson's disease,” Microbial Cell Factories, vol. 19, no. 1, 2020.
[2] L. D. Zerden, T. Guan, J. Shurer, L. Kreitzer, and E. Book, “Social work, Parkinson's disease care, and covid-19,” Social Work in Health Care, vol. 61, no. 3, pp. 139–157, 2022.
[3] M. Öberg, I. Fabrik, D. Fabrikova, N. Zehetner, and A. Härtlova, “The role of innate immunity and inflammation in Parkinson's disease,” Scandinavian Journal of Immunology, vol. 93, no. 5, 2021.
[4] W. Dauer and S. Przedborski, “Parkinson's disease,” Neuron, vol. 39, no. 6, pp. 889–909, 2003.
[5] A. H. Rajput, “Frequency and cause of Parkinson's disease,” Canadian Journal of Neurological Sciences / Journal Canadien des Sciences Neurologiques, vol. 19, no. S1, pp. 103–107, 1992.
[6] M. Lew, “Overview of Parkinson's disease,” Pharmacotherapy, vol. 27, no. 12 Part 2, 2007.
[7] W. Poewe, K. Seppi, C. M. Tanner, G. M. Halliday, P. Brundin, J. Volkmann, A.-E. Schrag, and A. E. Lang, “Parkinson disease,” Nature Reviews Disease Primers, vol. 3, no. 1, 2017.
[8] M. Stacy, P. Hickey, and M. Stacy, “Available and emerging treatments for Parkinson's disease: A Review,” Drug Design, Development and Therapy, vol. 2011, pp. 241–254, May 2011.
[9] F. L. Pagan, “Improving outcomes through early diagnosis of Parkinson's disease,” The American Journal of Managed Care, vol. 18. pp. 176–182, 2012.
[10] M. Olson, T. E. Lockhart, and A. Lieberman, “Motor learning deficits in Parkinson's disease (PD) and their effect on training response in gait and balance: A narrative review,” Frontiers in Neurology, vol. 10, 2019.
[11] J. C. Greenland and R. A. Barker, “The differential diagnosis of Parkinson's disease,” Parkinson’s Disease: Pathogenesis and Clinical Aspects, pp. 109–128, 2018.
[12] “The Unified Parkinson's Disease Rating Scale (UPDRS): Status and recommendations,” Movement Disorders, vol. 18, no. 7, pp. 738–750, 2003.
[13] J. Jankovic and A. Tarakad, “Faculty opinions recommendation of measuring Parkinson's disease over time: The real-world within-subject reliability of the MDS-UPDRS.,” Faculty Opinions – Post-Publication Peer Review of the Biomedical Literature, 2020.
[14] E. Tolosa, G. Wenning, and W. Poewe, “The diagnosis of Parkinson's disease,” The Lancet Neurology, vol. 5, no. 1, pp. 75–86, 2006.
[15] W. H. Chen, T. J. Chiang, M. C. Hsu, and J. S. Liu, “The validity of eye blink rate in Chinese adults for the diagnosis of Parkinson's disease,” Clinical Neurology and Neurosurgery, vol. 105, no. 2, pp. 90–92, Apr. 2003.
[16] N. P. Quinn, “Parkinson's disease: clinical features,” Baillieres Clin Neurol., vol. 6, no. 1, pp. 1–13, 1997.

[17] N. Fothergill-Misbah, R. Walker, J. Kwasa, J. Hooker, and K. Hampshire, "'Old people problems', uncertainty and legitimacy: Challenges with diagnosing Parkinson's disease in Kenya," Social Science & Medicine, vol. 282, p. 114148, Aug. 2021.
[18] R. Pahwa and K. E. Lyons, "Early Diagnosis of Parkinson's Disease: Recommendations From Diagnostic Clinical Guidelines," Implications of Early Treatment for Parkinson's Disease, vol. 16, no. 4, Mar. 2010.
[19] M. Tinelli, P. Kanavos, and F. Grimaccia, London School of Economics and Political Science, London, rep., Mar. 2016.
[20] Parkinson's disease: National clinical guideline for diagnosis and management in primary and secondary care. London: Royal College of Physicians, 2006.
[21] S.-Y. Lim, A. H. Tan, A. Ahmad-Annuar, C. Klein, L. C. Tan, R. L. Rosales, R. Bhidayasiri, Y.-R. Wu, H.-F. Shang, A. H. Evans, P. K. Pal, N. Hattori, C. T. Tan, B. Jeon, E.-K. Tan, and A. E. Lang, "Parkinson's disease in the Western Pacific region," The Lancet Neurology, vol. 18, no. 9, pp. 865–879, 2019.
[22] A. Channa, N. Popescu, and V. Ciobanu, "Wearable solutions for patients with Parkinson's disease and Neurocognitive Disorder: A systematic review," Sensors, vol. 20, no. 9, p. 2713, 2020.
[23] B. Heim, F. Krismer, R. De Marzi, and K. Seppi, "Magnetic resonance imaging for the diagnosis of Parkinson's disease," Journal of Neural Transmission, vol. 124, no. 8, pp. 915–964, 2017.
[24] F.Amato, L.Borzì, G.Olmo,andJ.R.Orozco-Arroyave, "An algorithm for Parkinson's disease speech classification based on isolated words analysis," Health Information Science and Systems, vol. 9, no. 1, 2021.
[25] S. Chakraborty, S. Aich, and H.-C. Kim, "Detection of Parkinson's disease from 3T T1 weighted MRI scans using 3D convolutional neural network," Diagnostics, vol. 10, no. 6, p. 402, 2020.
[26] Y. Uehara, S.-I. Ueno, H. Amano-Takeshige, S. Suzuki, Y. Imamichi, M. Fujimaki, N. Ota, T. Murase, T. Inoue, S. Saiki, and N. Hattori, "Non-invasive diagnostic tool for Parkinson's disease by sebum RNA profile with machine learning," Scientific Reports, vol. 11, no. 1, 2021.
[27] K. Tjaden, "Speech and swallowing in Parkinson's disease," Topics in Geriatric Rehabilitation, vol. 24, no. 2, pp. 115–126, 2008.
[28] S. Pinto, R. Cardoso, J. Sadat, I. Guimarães, C. Mercier, H. Santos, C. Atkinson-Clement, J. Carvalho, P. Welby, P. Oliveira, M. D'Imperio, S. Frota, A. Letanneux, M. Vigario, M. Cruz, I. P. Martins, F. Viallet, and J. J. Ferreira, "Dysarthria in individuals with Parkinson's disease: A protocol for a binational, cross-sectional, case-controlled study in French and European Portuguese (fralusopark)," BMJ Open, vol. 6, no. 11, 2016.
[29] H. Hazan, D. Hilu, L. Manevitz, L. O. Ramig, and S. Sapir, "Early diagnosis of Parkinson's disease via machine learning on speech data," 2012 IEEE 27th Convention of Electrical and Electronics Engineers in Israel, 2012.z
[30] J. S. Almeida, P. P. Rebouças Filho, T. Carneiro, W. Wei, R. Damaševičius, R. Maskeliūnas, and V. H. de Albuquerque, "Detecting Parkinson's disease with sustained phonation and speech signals using Machine Learning Techniques," Pattern Recognition Letters, vol. 125, pp. 55–62, 2019.
[31] W. E. Rinn, "The neuropsychology of Facial Expression: A review of the neurological and psychological mechanisms for producing facial expressions.," Psychological Bulletin, vol. 95, no. 1, pp. 52–77, 1984.
[32] P. Ekman and W. V. Friesen, "Facial action coding system," PsycTESTS Dataset, 1978.
[33] M. R. Ali, T. Myers, E. Wagner, H. Ratnu, E. R. Dorsey, and E. Hoque, "Facial expressions can detect Parkinson's disease: Preliminary evidence from videos collected online," npj Digital Medicine, vol. 4, no. 1, 2021.
[34] C. N. Karson, P. A. Lewitt, D. B. Calne, and R. J. Wyatt, "Blink rates in parkinsonism," Annals of Neurology, vol. 12, no. 6, pp. 580–583, 1982.
[35] E. Fitzpatrick, N. Hohl, P. Silburn, C. O'Gorman, and S. A. Broadley, "Case-control study of blink rate in Parkinson's disease under different conditions," Journal of Neurology, vol. 259, no. 4, pp. 739–744, 2011.
[36] Syed, Nadir A; Ali, Farwa; Sher, Khalid; Ikram, Amer; Soomro, Bashir; Shahbaz, Naila; Sheeran, Mughis; Jamil, Sarwar; Numan, Ahsan; Lakhair, Manzoor; Awan, Irshad; Baruch, Saleem; and Basheer, Haroon (2015) "National guidelines for diagnosis and management of Parkinson's disease in Pakistan," Pakistan Journal of Neurological Sciences (PJNS): Vol. 10: Iss. 1, Article 14.
[37] A. M. Sidahmed and H. A. Ali, "Frequency and associated factors of autonomic dysfunction in patients with Parkinson's disease in Khartoum State," Advances in Parkinson's Disease, vol. 08, no. 04, pp. 63–74, 2019.
[38] J. C. Vasquez-Correa, T. Arias-Vergara, J. R. Orozco-Arroyave, B. Escoffier, J. Klucken, and E. Noth, "Multimodal assessment of Parkinson's disease: A deep learning approach," IEEE Journal of Biomedical and Health Informatics, vol. 23, no. 4, pp. 1618–1630, 2019.
[39] G. Battineni, M. A. Hossain, N. Chintalapudi, E. Traini, V. R. Dhulipalla, M. Ramasamy, and F. Amenta, "Improved Alzheimer's disease detection by MRI using multimodal machine learning algorithms," Diagnostics, vol. 11, no. 11, p. 2103, 2021.
[40] R. Bhidayasiri and D. Tarsy, "Parkinson's disease: Hoehn and Yahr Scale," Current Clinical Neurology, pp. 4–5, 2012.
[41] C. Kotsavasiloglou, N. Kostikis, D. Hristu-Varsakelis, and M. Arnaoutoglou, "Machine learning-based classification of simple drawing movements in Parkinson's disease," Biomedical Signal Processing and Control, vol. 31, pp. 174–180, 2017.
[42] Naranjo, L., PÃ©rez, C.J., Campos-Roca, Y., MartÃn, J.: Addressing voice recording replications for Parkinson's disease detection. Expert Systems With Applications 46, 286-292 (2016)
[43] Erdogdu Sakar, B., Isenkul, M., Sakar, C.O., Sertbas, A., Gurgen, F., Dell, S., Apaydin, H., Kursun, O., 'Collection and Analysis of a Parkinson Speech Dataset with Multiple Types of Sound Recordings, IEEE Journal of Biomedical and Health Informatics, vol. 17(4), pp. 828- 834, 2013.
[44] Athanasios Tsanas, Max A. Little, Patrick E. McSharry, Lorraine O. Raming (2009), 'Accurate telemonitoring of Parkinson's disease progression by non-invasive speech tests', IEEE Transactions on Biomedical Engineering.
[45] Sakar, C.O., Serbes, G., Gunduz, A., Tunc, H.C., Nizam, H., Sakar, B.E., Tutuncu, M., Aydin, T., Isenkul, M.E. and Apaydin, H., 2018. A comparative analysis of speech signal processing algorithms for Parkinson's disease classification and the use of the tunable Q-factor wavelet transform. Applied Soft Computing,
[46] A. Torfi and E. A. Fox, "Corgan: Correlation-capturing convolutional generative adversarial networks for generating Synthetic Healthcare Records," arXiv.org, 04-Mar-2020. [Online]. Available: https://arxiv.org/abs/2001.09346.
[47] F. Eyben, M. Wöllmer, and B. Schuller, "Opensmile," Proceedings of the international conference on Multimedia - MM '10, 2010.
[48] S. I. Serengil and A. Ozpinar, "Hyperextended Lightface: A facial attribute analysis framework," 2021 International Conference on Engineering and Emerging Technologies (ICEET), 2021.
[49] T. Baltrusaitis, A. Zadeh, Y. C. Lim, and L.-P. Morency, "OpenFace 2.0: Facial Behavior Analysis Toolkit," 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018), 2018.
[50] J. Robert, PyDub. PyDub, 2014.
[51] K. Cortacero, T. Fischer, and Y. Demiris, "RT-BENE: A dataset and baselines for real-time Blink estimation in Natural Environments," 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), 2019.
[52] A. Goncalves, P. Ray, B. Soper, J. Stevens, L. Coyle, and A. P. Sales, "Generation and evaluation of Synthetic Patient Data," BMC Medical Research Methodology, vol. 20, no. 1, 2020.
[53] J. Mei, C. Desrosiers, and J. Frasnelli, "Machine learning for the diagnosis of Parkinson's disease: A review of literature," Frontiers in Aging Neuroscience, vol. 13, 2021.
[54] M.Nilashi,R.A.Abumalloh,B.Minaei-Bidgoli, S.Samad, M.Yousuf Ismail, A. Alhargan, and W. Abdu Zogaan, "Predicting Parkinson's disease progression: Evaluation of ensemble methods in machine learning," Journal of Healthcare Engineering, vol. 2022, pp. 1–17, 2022.
[55] Z. Chang, Z. Chen, C. D. Stephen, J. D. Schmahmann, H.-T. Wu, G. Sapiro, and A. S. Gupta, "Accurate detection of cerebellar smooth pursuit eye movement abnormalities via mobile phone video and Machine Learning," Scientific Reports, vol. 10, no. 1, 2020.

IX. Appendix

Diagnosis steps and information can be found at https://measure.parkin-sense.pulkith.com

X. Statement On Outside Assistance

Researchers Name: Pulkith Paruchuri
Title of Paper: ParkinSense: Contactless Parkinson's Disease diagnosis, telemonitoring, and severity analysis via Machine Learning and Multimodal Digital Biomarkers
What steps led you to formulate your hypothesis? (Where did you get the idea for your research?) Please be specific. After a related couple was diagnosed with Parkinson's Disease within a few years of each other (an event that is less likely than flipping a coin and getting heads 10 times in a row), I began looking at Parkinson's Disease, since I had a fair bit of free time due to the Covid-19 Pandemic. I was quite interested, since the husband was diagnosed quite a bit earlier, and had much fewer symptoms than the wife, which led to me looking into the diagnosis. From there, I learned about Machine Learning for diagnosis and realized that although dozens of papers had been published on the matter, there is still not a single public method of diagnosis that is accessible. This allowed the current issues with diagnosis to affect millions. As a result, my hypothesis derives from these papers and the belief that many of the results from these papers, as well as many of my own novel beliefs, can be combined into an application that can indeed be released to the public. I was further interested and motivated in the work along the way when a close friend's father was diagnosed with Parkinson's quite late.
Where did you conduct the major part of your work? (i.e. home, school, or another institutional setting, university lab, medical center, etc.) The majority of my work was conducted from a home setting. For most of the research only development was required, either via web development, or machine learning model development, all of which can be done remotely without a lab.
If you worked in an institutional setting, did you work on your project as part of a team or group? If so, how large was the team and who was on the team (students, adult researchers, etc.)? Describe your role on the team. I was not part of an Instuitional Setting. However, I do lead a nonprofit organization of around 30 people (https://neurose.pulkith.com), that is working to make ParkinSense a public solution. All of the research in this paper was conducted entirely by me. The Neurose team is only helping to advertise the solution, develop a mobile app out of the solution (which is not discussed in this paper), and explore solutions to treating Parkinson's Disease (which is also not discussed in this paper). All of the scientific aspects of ParkinSense were developed solely by me, while the Neurose team is working to make the application in this paper a full reality so that it can work to help millions globally. Furthermore, most of the research involved in this paper was conducted before Neurose was even formed.

Describe what parts of the research you did on your own and what parts you received help (i.e. literature search, hypothesis, experimental design, use of special equipment, gathering data, evaluation of data, statistical analysis, conclusions, and preparation of written report (abstract and/or paper).

Unless otherwise noted below, all of the work was completed solely by me.

Some of the data were sourced from public methods, mostly the UC Irvine Public Database.

Additionally, authentic data collected for this study was collected through a website I developed, although as per FDA requirements, an IRB approval was required. Approval was granted through Yale University. Further authentic data was collected from Andhra University.

Further Eye Blink Rate data was sourced from Professor Broadley from Griffith University.

Further Facial Expression data was collected from the ParkTest team at the University of Rochester.

Eye Saccade data, which was not used in this report, was collected from Dr. Gupta at Duke University.

The CardinalKit Lab at the Stanford Medical School developed the framework to ensure ParkinSense abides by federal and state data privacy laws.

The Ghosh Lab at Northwestern University answered queries regarding the state and nature of Parkinson's Disease Research and treatment when I formed my hypothesis.

Roshan Patel and Satyen Dhamankar from the Webb Lab at Princeton University for reading over the initial results to propose areas for further studies, such as analyzing which feature provided the highest predictive power.

The Awatramani Lab at Northwestern University provided results from Parkinson's Disease diagnosis trials on mice, which provided insight into promising solutions to explore similar diagnoses in humans. This was largely in the form of bradykinesia, which was the inspiration for Eye Blink Rate in this study (in the mice, bradykinesia was shown in the slowness of walking).

Finally, the Neurose Team, which I lead, is helping make ParkinSense a global solution that is accessible to everyone. More about this in the next question.

Is your research current or a continuation of previous research? If a continuation, please describe the current work and advancement(s) of this research in comparison to the prior work and results.

I would like to note, that two papers based on the research in this paper are currently being published or under review (with myself as the sole author on both). One paper will be presented at IEEE ICIIBMS in November, and the other is under review by Nature Neuroscience.

All of the research in this paper is my own and has been ongoing since 2020. It is not a direct continuation of another research project. While a few papers have been able to semi-accurately create scripts that can diagnose individuals with Parkinson's Disease, either from Vocal Symptoms, or Facial Symptoms, not much work in this vein has bee done. Additionally, the vast, vast, majority of papers that diagnose Parkinson's use invasive data, such as MRIs, gene sequencing, blood tests, and vocal data collected in a studio with a professional microphone.

- This is the first time Vocal Symptoms and Facial Symptoms are combined for Parkinson's Disease diagnosis.
- This is also the first time Eye Blink Rate is used as a symptom of Parkinson's Disease diagnosis.
- **This is also the first time a fully automated application is made from diagnosis and is available to the public. There exists no other application that can extract features**

from a recording and make a diagnosis and severity classification. ParkinSense is the first time anyone implement a solution for Parkinson's Disease diagnosis into something that can be used by anyone.

- This is the first time that severity classification and diagnosis is done with the same multimodal model.
- This is the first time a website application is developed for Parkinson's Disease diagnosis
- This is one of the first times that a multimodal application is used for Parkinson's Diagnosis, and is the first time a multimodal application is used on digital biomarkers for Parkinson's Disease diagnosis.
- This is the first time diagnosis and severity classification are combined into a single application for Parkinson's Diagnosis.
- This is the first time synthetic data is used in this manner for Parkinson's's Patients to increase accuracy.
- This is the first time that data augmentation techniques are used for Parkinson's Patients to increase accuracy,
- Furthermore, authentic data was collected, and the featured data will be published as part of a paper to help the community.

Over-current industry methods of diagnosis, ParkinSense is:

- Cheaper: Current methods cost $5,000+, while ParkinSense costs between $0-$17
- Accurate: Current Methods misdiagnose 30-40% of people, while ParkinSense misdiagnoses 0.28% of people.
- Rapid: Current Methods of diagnosis take days to weeks, while ParkinSense takes 12 minutes.
- Remote: Current methods of diagnosis involve either a neurologist or invasive scans and injections, while ParkinSense can be used completely from the comfort of one's home using a laptop or phone.