ParkinSense: A novel approach to remote Idiopathic Parkinson's Disease Diagnosis, Severity Profiling, and Telemonitoring via Ensemble Learning and Multimodal Data Fusion on Webcam-Derived Digital Biomarkers

For double blind review purpose

Please do not write your mames and affiliations. During the initial submission.

***Abstract*—Despite being the fastest growing neurodegenerative disease in the world, with over 10 million patients worldwide, there is no definitive diagnosis method for Parkinson's Disease currently. Current diagnosis technologies misdiagnose one in three patients, and rely on inaccurate technologies or are subjective to the administrating clinicians. Furthermore, they are inaccessible to billions due to immobility, geographic barriers, or associated costs. With early and accurate diagnosis being vital to effective treatment, a tremendous issue emerges. ParkinSense addresses these issues, serving as a web application that contactlessly analyzes three cardinal symptoms of Parkinson's Disease over a standard webcam and microphone: Hypomimia, Dysarthria, and Bradykinesia. With Ensemble Learning coupled with various classifiers, such as Random Forests and Support Vector Machines, Parkinsense was able to achieve accuracy rates of 99.72% when tested on synthetic patients. By examining, and combining multiple modalities, Parkinson's boasts accuracy rates higher than many advanced unimodal technologies, as patients often do not display all the symptoms of Parkinson' Disease substantially, leading to misdiagnosis when relying on only one symptom for diagnosis. ParkinSense strongly suggests that entirely contactless diagnosis of PD through digital biomarkers, can be effective, and web applications can be utilized for rapid, automated, and remote diagnosis, which can be crucial for those affected by barriers that prevent access to diagnosis.**

Keywords—Diagnosis; Machine learning; Multimodal learning; Parkinson's disease; Severity;

I. 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 MRI scans and DaTScan 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 for PD 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.

IV. PD diagnosis software 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.

II. Related Work

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 Magnetic Resonance Imaging (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.

A. *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].

A number of 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 HC [29], [30].

B. *Hypomimia*

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

C. *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].

D. *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 parts A-C of section 2. 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 [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].

TABLE I. DATASET INFORMATION

Dataset	Size	Features	Participants PD	Participants HC	Mean Age	% Male	Content Utilized	H&Y Score	UPDRS Score
A	240	46	40	40	N/A	60%	/a/ phonation	N	N
B	1040	26	20	20	64.9±9	60%	/a/ phonation, /o/ phonation, /u/ phonation, words, short sentences	N	Y
C	168	26	28	0	62.7±11	N/A	/a/ phonation, /o/ phonation	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

E. Severity profiling

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]

III. DATA AND METHOLDOLOGIES

Since we theorized that *Dysarthria, Hypomimia, and Bradykinesia* features could reliably be collected and analyzed via a web platform, these symptoms of PD were focused on. ParkinSense, a web application, was developed that enabled individuals to complete a 6-minute study. After the study was complete, data processing, feature extraction, model diagnosis, and severity estimation took under 5 minutes.

A. 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.

Extracted Features from PD and HC vowel phonations and voice recordings were gathered from numerous datasets in UC Irvine Database [42]-[45]. Datasets A-E of Table 1 outline the size, demographics, and information within each dataset used for this modality. The vowel phonations of /a/ from Datasets A, B, C and E were aggregated into a single large dataset, with the features represented being the intersection of the features in these datasets. A similar procedure was done for the vowel phonation of /o/ from datasets B and C. Additionally, UPDRS scores were reported in Datasets B and D.

Facial Expression Data from the study in [33], can be seen in Dataset F of Table 1, and contains information regarding the variance of the magnitude of AUs involved with Smile, Surprise, and Disgust facial expressions, when the AU was activated.

EBR data, sourced from [35], is represented in Dataset G of Table 1, and provides information about the EBR of PD patients and HC during a reading exercise, and also contains information about the H&Y Scale of diagnosed individuals, and whether or not they wore glasses during the exercise. The EBR dataset was augmented via Test-Time [46].

B. Website Diagnosis Design

A diagram of the website, which is to be described in parts B-D of section 3 can be found in Fig. 1. A prototype example of this website 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

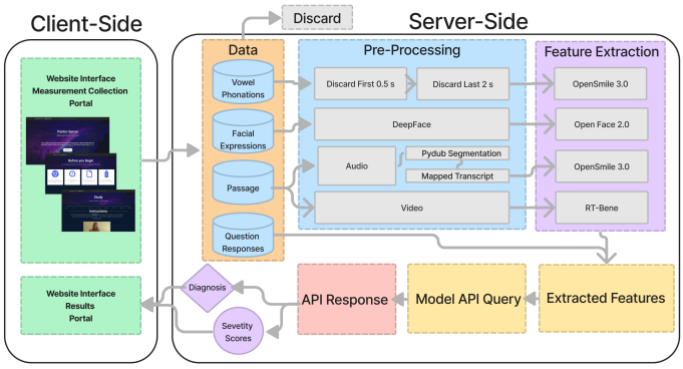

Fig. 1. Web Application Overview, Data Processing Methods, and Feature Extraction technologies.

their entire head and part of their torso are in the frame, not dissimilar from a passport photo. They are requested to be 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 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) Questionaire: 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).

C. Data Pre-processing & Feature Extraction

Data processing and Feature Extraction was done entirely through an automated workflow, to mimic the requirements for automated data processing when used as a web application for rapid diagnosis.

1) Vocal Data: 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. 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 Data: 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. Then 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) Reading Passage Data: 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. RT-Bene [51] was used to calculate the EBR and is effective both with and without lenses.

D. 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, were hosted in Amazon

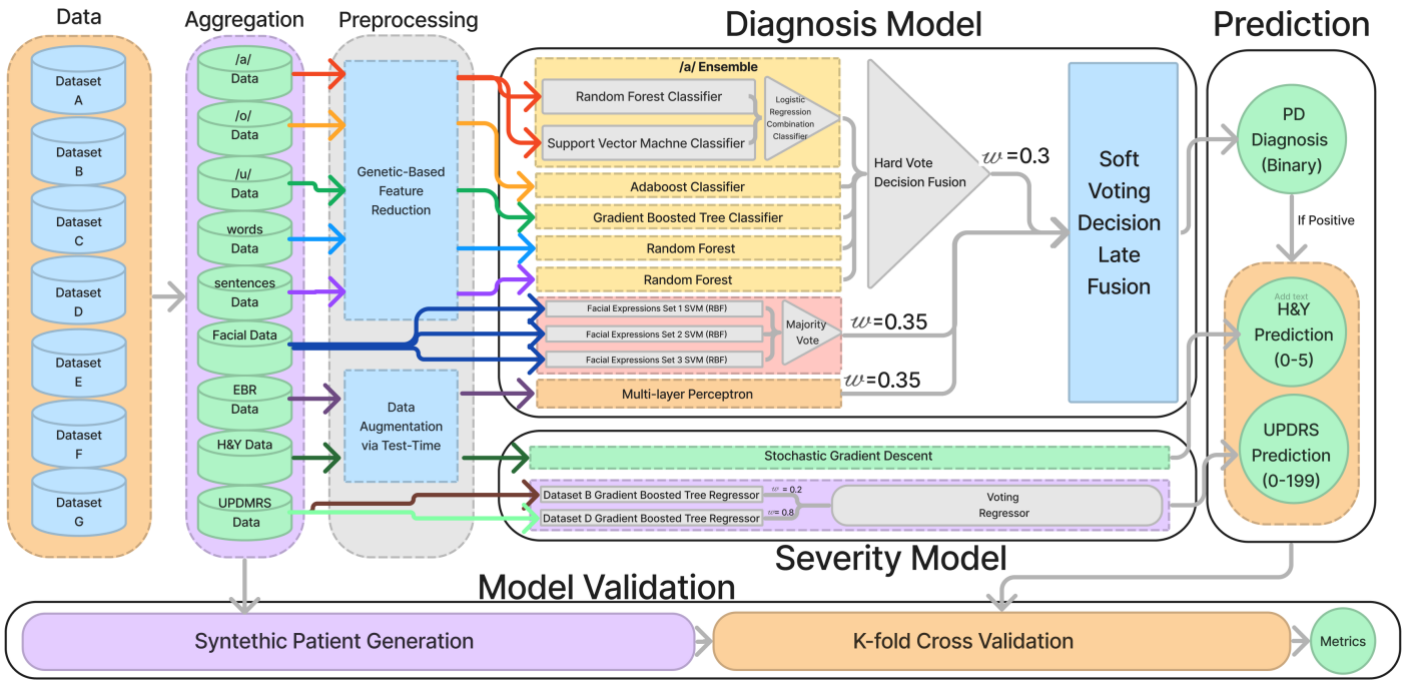

Fig. 2. Data sources and aggregation categories map. Preprocessing methods, diagnosis and severity models classifiers and flow, and output predictions.

Web Services (AWS), and accessed by the server through an AWS Gateway API.

E. Model Framework

A diagram of the 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.

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 out-perform 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 30, 35, and 35 used for Dysarthria, Hypomimia, and Bradykineia respectively.

5) Severity Profiling: We predicted two scores: H&Y and UPDRS. HY 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 the dataset B, due to the magnitude of data in that set.

F. Evaluation

Due to the difficulty in finding data, especially on all modalities simultaneously for overall model evaluation, high-fidelity synthetic patients were generated, with specific modality instances being paired by gender, age, and disease severity. Synthetic patients can provide near identical results when compared to live patients [52]. K-fold cross-validation was used for all model evaluations, coupled with metrics discussed in section 4.

IV. Results & Discussion

A. 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.

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	0.949	0.941	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 85%, 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.

B. Severity Profiling & Telemonitoring

TABLE III. Severity Regression Evaluation Metrics

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

Mean Absolute Error (MAE) was used as the metric for both H&Y Scale and UPDRS Scale regression predictions. The MAE for each regressor can be found in Table 3.

Overall, ParkinSense's severity profiling model can be used relatively accurately to predict UPDRS Scale scores, as well as H&Y to a lesser extent, allowing users to monitor disease progression and medication effectiveness over time.

C. Comparative Analysis

ParkinSense outperforms many unimodal diagnosis models in literature [48]. ParkinSense's individual modalillity classifiers perform similarly to top models [53] but ParkinSense is the first to analyze EBR for PD diagnosis via Machine Learning**.** Furthermore, ParkinSense performs better than or equal to most multimodal PD diagnosis models, even those that analyze genes, MRI scans, or other invasive or noncontactless forms of data [53]. Additionally, ParkinSense is able to have comparable metrics for severity analysis as some of the top models for UPDRS prediction via Machine Learning [54].

D. Website Analysis

All methods of data collection, feature extraction, and model querying via the web application were validated to be effective and successful. Data collection depended on the user but usually took around 6 minutes. Pre-processing and Feature Extraction usually took around 3 minutes, and prediction via the model took around 2 minutes. As such a rapid diagnosis was achieved, as the entire process took around 11 minutes. Therefore, ParkinSense suggests that a contactless diagnosis via webcam-and-microphone-derived biomarkers is quite effective.

V. Conclusion

A. Error Analysis

The largest source of error 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 to create synthetic patients.

Additionally, for dataset aggregation in part A of section 3, certain features likely were not collected in similar manners for each dataset, which may have introduced some noise that made the model underperform.

B. Future Work & Applications

Since a large goal of this study was to diagnose PD in patients through behavioral modalities that can be collected from a patient's head, much of the proposed future work follows in that vein.

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.

Furthermore, certain features, such as speech rate, blink duration, blink latency, reading rate, 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 clinical tests on live patients in the near future. 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.

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.

C. Conclusion

ParkinSense makes encouraging process for the development of PD contactless diagnosis software via machine learning technologies and multiple modalities. ParkinSense automates data collection and feature extraction on the web application, and regarding the unimodality models, have performed similarly to many top models. Furthermore,we introduce the first models that involve PD diagnosis classification through EBR, /o/ and /u/phonations, words, and short sentences. 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-proofing tool, that outperforms all unimodal and multimodal diagnosis tools in literature for PD diagnosis, and delivers competitively accurate levels of PD severity and progression metrics. All of this is accessible through a web application that is accurate and rapid. ParkinSense suggests that contactless PD diagnosis and severity tracking can be conducted accurately, which may prove invaluable for those in which PD diagnosis is inaccessible due to barriers, immobility, or associated costs. The unparalleled accuracy of ParkinSense demonstrates the effectiveness of multimodal machine learning algorithms 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 of PD patients.

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. Gimaccia, 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, and J. 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 withz 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; Sheerani, Mughis; Jamil, Sarwar; Numan, Ahsan; Lakhair, Manzoor; Awan, Irshad; Barech, 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. Eskofier, 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., Delil, 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] M. Kimura, "Understanding test-time augmentation," *Neural Information Processing*, pp. 558–569, 2021.

[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. Yousoof 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.

APPENDIX

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