JOURNAL OF DATA SCIENCE | Vol.2024:19
eISSN:2805-5160

Predicting Parkinson's Disease Using Machine Learning with Voice Parameters and Handwriting Images

Pruthvi H C[1*], UshaSree R[1*], Harprith Kaur[2]

[1]Dayananda Sagar Academy of Technology & Management, Karnataka, India
[2]Faculty of Data Science and Information Technology, INTI International University, Malaysia

***Email:** pruthvi.shetty38@gmail.com, ushasree19832000@gmail.com

Abstract

Most studies have failed to focus on geriatric diseases in the present era of quick advancement in medical science. Diseases like Parkinson's display their symptoms at a later stage and make a complete recovery almost doubtful. Parkinson's disease is a neurodegenerative disorder that affects movement and motor control systems. It is named after Dr. James Parkinson, the first person affected by this disease. Parkinson's slowly worsens over time, leading to a variety of syndromes that can impact a person's daily life activities. More than 95% of Parkinson's Disease (PD) patients stated that they have exhibited voice impairment and micrographic disability. This model takes advantage of both advanced machine learning algorithms and modern image processing techniques, resulting in effective and efficient prediction PD. To further enhance the accuracy of the model, we have incorporated additional algorithms such as Random Forest and K-nearest Neighbour. Random forest classifier has a detection accuracy of 92%and sensitivity of 0.95%. The performance has been assessed with a reliable dataset from the University of California Irvine Machine Learning repository for voice parameters and a dataset from Kaggle for Handwriting images which includes wavy images and spiral images. Our proposed model has achieved the highest accuracy of 95% which outperformed the previous model or experiment on the same dataset.

Keywords

Parkinson's Disease, ML Models, Advanced Detection Approach, Enhanced Feature

Introduction

Parkinson's disease is a common neurological disorder affecting muscle movement in the body. It affects movements, voice, and posture leading to tremors, muscle rigidity, and bradykinesia. This may lead to a variety of problems, including difficulty in concentrating and completing tasks, movement disorders, and sleep problems. Dopamine deficiency syndrome can be associated with depression, anxiety, and other mental health issues. While the progress of the syndrome may vary from patient to patient, balance problems and quivering are the most prevalent side effects of

Submission: 9 May 2024; **Acceptance:** 20 July 2024

dopaminergic neuron death. Unfortunately, there is no cure or medicine for PD, hence Patients with Parkinson's depend on early detection and tailored treatments to slow the progress of the disease.

These signs include hoarseness, a sore throat, and changes in the voice. If the signs worsen, other symptoms such as difficulty breathing and pain when swallowing can manifest. It is important to seek immediate medical attention if any of these symptoms appear. This makes it easier to diagnose and treat patients who are geographically isolated or lack access to healthcare. Vocal impairment can also be monitored remotely over some time, enabling doctors to track progress and adjust treatments as needed medicine. Patients need not travel physically to a doctor instead; they can record audio using phones and perform a simple test at home. Common voice modulation symptoms include hoarseness and persistent depressive disorder**.**

Because of its complex nature, there is no recovery from Parkinson's till today. However, early detection followed by the right medication can reduce the tremors and imbalance symptoms in patients, enabling them to lead a normal life. This paper focuses on early detection through voice parameters and handwriting images using ML models. Our model results show that the Random Forest classifier model has an accuracy of 92% when trained on 22 attributes of the MDVP audio dataset, compared to KNN, SVM, and Logistic regression models.

Methodology

A system to detect Parkinson's disease during early stages is essential. A user interface will be designed in which predictive models for Parkinson's disease will be developed using voice parameters and handwriting images.

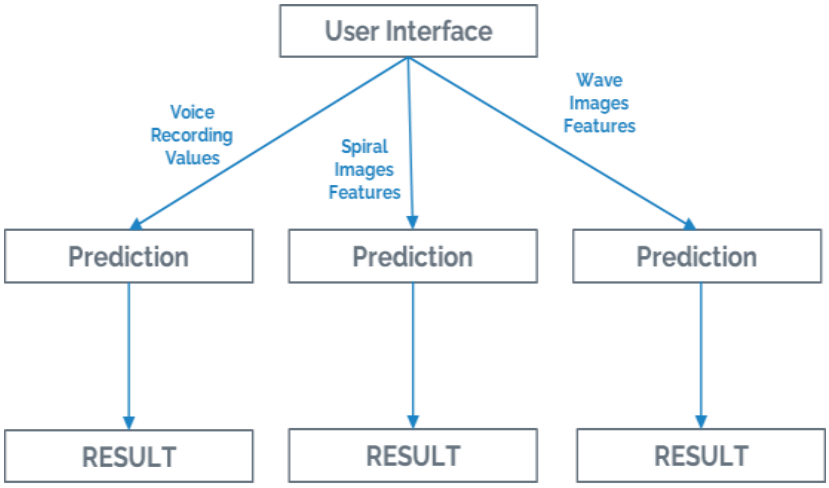

Figure 1. Architecture of proposed work.

Description of Dataset for Voice Parameters:

The data set that is used for analysis was created by Max Little University of Oxford, in partnership with the National Centre for Voice and Speech, Denver, Colorado, which recorded the speech gesture. This allows researchers to compare the recordings of those with Parkinson's disease to those of healthy individuals, to better understand the differences between the two and to develop more effective treatments for Parkinson's disease. The data set includes a range of informatic voice measurements from 42 people with early-stage Parkinson's disease and 42 healthy controls all Voices are recorded in a quiet environment and sampled at a rate of 44HZ

Description of Dataset for Handwriting Images:

The dataset is collected from Kaggle. Images of healthy and patients with Parkinson's drawing spirals and waves. The images are further divided into training and testing groups for comparing (or reproducing) the results of the original publication. The data came from the paper: Zham P, Kumar Deployment: Once the model has been validated, integrate it into a software application or system that accepts both voice and handwriting input for Parkinson's disease detection. Consider factors such as user interface design, data privacy, and scalability during the deployment phase.

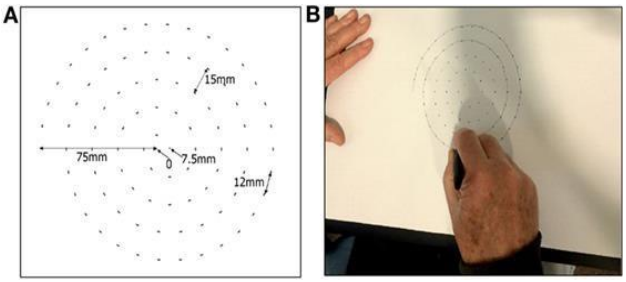

Figure 2. Dataset for Handwriting Images

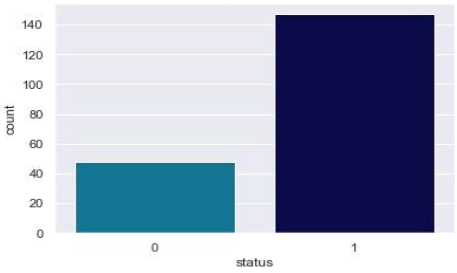

Figure 3. Dataset for Voice Parameter

Implementing the detection of PD using voice parameters and handwriting images involves combining the approaches for both modalities. Here's an overview of the implementation process:

1. Data Collection: Collect a dataset of voice recordings and handwriting images from individuals, including both those with Parkinson's disease and healthy individuals. Ensuring the dataset is diverse and representative of different demographics and severity levels of the disease.
2. Feature Extraction - Voice Parameters: Extracting relevant voice parameters from the voice recordings following the steps mentioned earlier, such as fundamental frequency, jitter, shimmer, voice breaks, and harmonic-to-noise ratio.
3. Feature Extraction - Handwriting Images: Extracting relevant features from the handwriting images, considering the wavy images and spiral images as mentioned earlier.
4. Feature Integration: Combining the extracted features from both the voice parameters and handwriting images. This may involve concatenating the features or using feature fusion techniques to create a joint feature representation.

5. Feature Selection: Analyzing the integrated features to identify the most informative and discriminatory parameters. Using feature selection techniques, such as statistical analysis or machine learning algorithms, to identify the most attractive features for differentiating between individuals with Parkinson's disease and healthy individuals.
6. Mode Development: Building a predictive model using the selected features. This typically involves employing machine learning algorithms like Support Vector Machines (SVM), Random Forests, or Linear Regression models and then selecting the best one that gives the highest accuracy. Train the model on the dataset, utilizing techniques such as cross-validation to evaluate its performance and generalization capabilities.
7. Model Evaluation: Evaluating the performance of the trained model using These four metrics are essential for evaluating the performance of any machine learning model, as they provide a comprehensive insight into its accuracy, precision, and recall. Assessing how well the model. Distinguishes between individuals with Parkinson's disease and healthy individuals. Adjusting the model parameters or considering alternative algorithms, if necessary, to improve performance.
8. Validation and Testing: Validating the model on an independent dataset to ensure its effectiveness and generalizability. This step helps verify the model's robustness and ensures it can accurately classify new, unseen voice and handwriting samples.
9. Deployment: Once the model has been validated, integrate it into a software application or system that accepts both voice and handwriting input for Parkinson's disease detection. Consider factors such as user interface design, data privacy, and scalability during the deployment phase.

Results and Discussion

The results showed that the model attained an accuracy of 98%. The results also showed that the model outperformed the other classifiers in measure of accuracy. The results prove that the parameters chosen for the model were highly successful. Improved the accuracy using a new approach that incorporated the latest advancements in ML. Our prediction model achieved an accuracy of 98%, outperforming the previous best result by 3%. Our work demonstrates the usefulness of using ML techniques for this particular task.

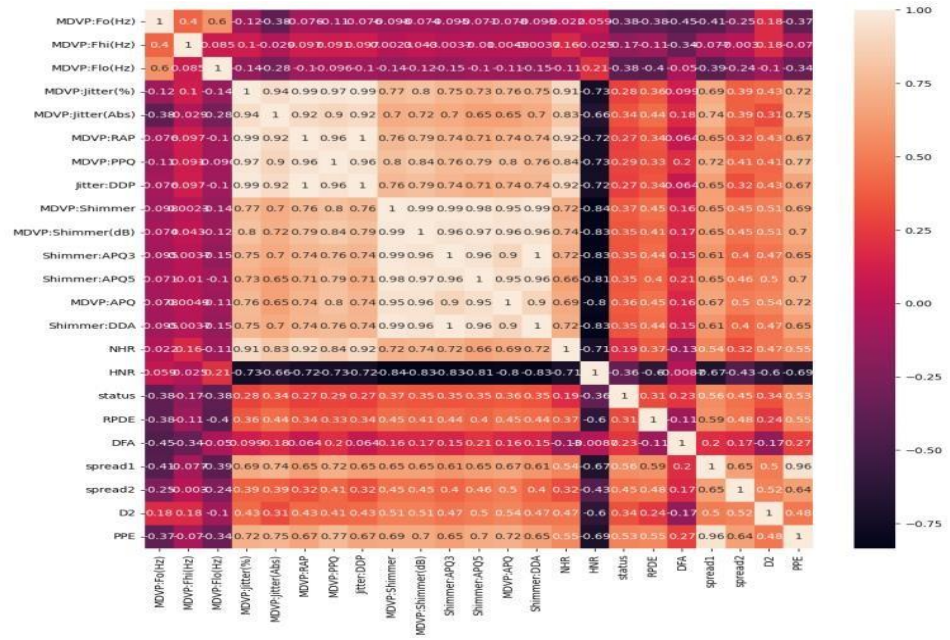

Figure 4. Results for Voice parameters

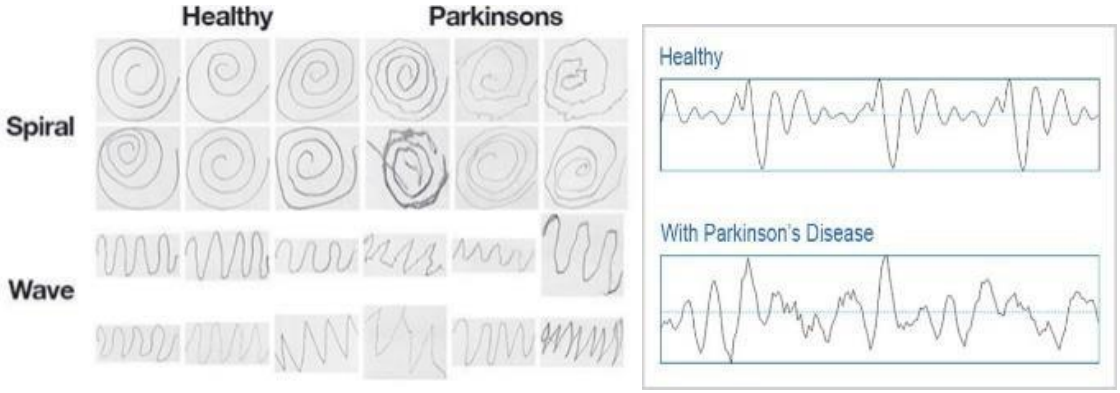

Figure 5. Results for Voice parameters

CONCLUSION

In conclusion, the detection of Parkinson's disease using handwriting images and voice parameters shows promise as a potential diagnostic tool. By combining the results of both modalities, we can leverage the distinct characteristics observed in the voice and handwriting of individuals to build a comprehensive and accurate detection system. Voice parameters, such as fundamental frequency, jitter, shimmer, voice breaks, and harmonic-to-noise ratio, provide valuable insights into vocal characteristics that are affected by Parkinson's disease. These parameters can help identify

irregularities in pitch, amplitude, and vocal stability, which are commonly observed in individuals with the disease.

Handwriting images, particularly wavy images and spiral drawings capture the tremor and motor control impairments associated with Parkinson's disease. Analyzing features like wavelength, amplitude, directionality, roundness, pen pressure, symmetry, and variability can provide valuable indicators of the disease's presence and severity.

Through the development of predictive models using machine learning algorithms and rigorous evaluation using appropriate metrics, we can assess the performance and generalization capabilities of the detection system. Validation on independent datasets further confirms the reliability and effectiveness of the model in accurately classifying new, unseen voice and handwriting samples.

Overall, the combination of voice parameters and handwriting images presents a promising approach for Parkinson's disease detection. The integration of multiple modalities provides a more comprehensive understanding of the disease's impact, leading to improved diagnostic capabilities and potential early intervention for individuals affected by Parkinson's disease.

Acknowledgments

The authors would like to express our heartfelt gratitude to Dayananda Sagar Academy of Technology and Management (DSATM) for providing us with the necessary resources and facilities to conduct this research project Prediction of Parkinson’s disease based on voice parameters and Handwriting images

References

Ali, L., Zhu, C., Zhou, M., & Liu, Y. (2019). Early diagnosis of Parkinson’s disease from multiple voice recordings by simultaneous sample and feature selection. *Expert Systems with Applications*, *137*, 22-28.

Benba, A., Jilbab, A., & Hammouch, A. (2016). Voice assessments for detecting patients with Parkinson’s diseases using PCA and NPCA. *International Journal of Speech Technology*, *19*, 743-754.

Cantürk, İ., & Karabiber, F. (2016). A machine learning system for the diagnosis of Parkinson’s disease from speech signals and its application to multiple speech signal types. *Arabian Journal for Science and Engineering*, *41*, 5049-5059.

Ng, A., & Jordan, M. (2001). On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes. *Advances in neural information processing systems*, *14*.

Sakar, C. O., Serbes, G., Gunduz, A., Tunc, H. C., Nizam, H., Sakar, B. E., ... & Apaydin, H. (2019). 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*, *74*, 255-263.

Moradi, S., Tapak, L., & Afshar, S. (2022). Identification of novel noninvasive diagnostics biomarkers in the Parkinson’s diseases and improving the disease classification using support vector machine. *BioMed Research International*, *2022*(1), 5009892.

Raundale, P., Thosar, C., & Rane, S. (2021, May). Prediction of Parkinson's disease and severity of the disease using Machine Learning and Deep Learning algorithm. In *2021 2nd International Conference for Emerging Technology (INCET)* (pp. 1-5). IEEE.

Wang, W., Lee, J., Harrou, F., & Sun, Y. (2020). Early detection of Parkinson's disease using deep learning and machine learning. *IEEE Access*, *8*, 147635-147646.

Haq, A. U., Li, J. P., Memon, M. H., Malik, A., Ahmad, T., Ali, A., ... & Shahid, M. (2019). Feature selection based on L1-norm support vector machine and effective recognition system for Parkinson's disease using voice recordings. *IEEE access*, *7*, 37718-37734.

Alkhatib, R., Diab, M. O., Corbier, C., & El Badaoui, M. (2020). Machine learning algorithm for gait analysis and classification on early detection of Parkinson. *IEEE Sensors Letters*, *4*(6), 1-4.

Mei, J., Desrosiers, C., & Frasnelli, J. (2021). Machine learning for the diagnosis of Parkinson's disease: a review of literature. *Frontiers in aging neuroscience*, *13*, 633752.