Parkinson's Disease Prediction System

Vani. M.p[1], S. Hariprasanth[2], Danush.r[2], Rithik Paul Simonds.Pk[2]

[1]School of Computer Science Engineering and Information Systems Vellore Institute of Technology Vellore, Tamilnadu, India
[2]Department of Computer Science Engineering and Information Systems Vellore Institute of Technology Vellore, Tamilnadu, India

Abstract
Parkinson's disease (PD) is a chronic and progressive neurological disorder that predominantly impacts motor functions, speech, and overall movement coordination. The traditional diagnostic approach to Parkinson's relies heavily on clinical observations and subjective evaluations, which often result in delayed detection or misdiagnosis. In response to this gap, this project proposes a machine learning-based Parkinson's Disease Prediction System that utilizes biomedical voice measurements as key biomarkers for early diagnosis. Features such as jitter, shimmer, and harmonic-to-noise ratio (HNR) are extracted from patient voice samples and analyzed using various machine learning algorithms including Logistic Regression, Support Vector Classifier (SVC), and Random Forest. The project evaluates the models rigorously and selects the best-performing one based on precision, recall, F1-score, and ROC-AUC metrics. Furthermore, the project leverages cloud computing for secure data storage, scalable deployment, and real-time remote access, ensuring that healthcare providers can easily utilize the prediction system. The system not only provides an efficient and scalable solution but also demonstrates a practical application of AI and cloud technologies in revolutionizing early-stage medical diagnostics.

Keyword: Parkinson's disease, Machine Learning, Cloud Computing, Biomedical Voice Data, Early Diagnosis, Logistic Regression, SVC, Random Forest, Healthcare AI.

Introduction

Parkinson's disease (PD) severely affects quality of life, making early intervention critical. Symptoms often appear after substantial neurological damage has already occurred. Current diagnosis methods — such as the Unified Parkinson's Disease Rating Scale (UPDRS) — rely heavily on human observation and clinical experience, introducing subjectivity.
With advances in machine learning (ML) and cloud technologies, a new opportunity arises to automate early PD detection based on measurable biomarkers like voice disorders (e.g., vocal tremors, frequency variations).
This project addresses this need by:

- Developing a prediction system using ML trained on voice data.
- Storing models and user data securely in the cloud, ensuring remote, scalable, and real-time access.

This approach not only improves diagnostic accuracy but also reduces time and dependency on expert clinicians.

Related Work

Several studies have explored the application of machine learning algorithms to healthcare, particularly focusing on early detection of Parkinson's disease through voice data analysis. Previous research has identified features like jitter (frequency variation), shimmer (amplitude variation), and harmonic-to-noise ratio as strong predictors of Parkinson's symptoms. Various machine learning models such as Support Vector Machines (SVM), Random Forests, and Logistic Regression have demonstrated effective classification capabilities when applied to such datasets. In addition to traditional models, deep learning architectures like Convolutional Neural Networks (CNNs) have been investigated for automatic feature extraction and improved accuracy. However, a critical observation from the literature is the frequent neglect of deployment considerations. Many studies focus solely on model accuracy but do not address how these models can be integrated into real-world healthcare environments where scalability, security, and remote accessibility are crucial. Recent advancements show that integrating cloud computing with machine learning can bridge this gap, allowing models to be deployed in a scalable and globally accessible manner, thereby transforming research prototypes into usable healthcare tools

Proposed Methodology

The Parkinson's Disease Prediction System developed in this project follows a structured and systematic methodology. The first step involves data acquisition from the publicly available Parkinson's disease dataset provided by the UCI Machine Learning Repository. The dataset comprises biomedical voice measurements of 195 individuals, with 22 numerical attributes representing various voice parameters and one target attribute indicating whether the individual has Parkinson's disease.
Following data collection, preprocessing is carried out to ensure data quality and model compatibility. Non-informative columns such as patient names are removed, and the feature data is standardized using the StandardScaler method to normalize the distribution and facilitate efficient model convergence. After preprocessing, the dataset is divided into training and testing sets in an 80:20 ratio to evaluate model generalization.

Multiple machine learning algorithms are trained on the dataset, including Logistic Regression, Support Vector Classifier (SVC), Random Forest Classifier, K-Nearest Neighbors, and Naive Bayes classifiers. Hyperparameter tuning is performed, particularly for the SVC (using radial basis function kernels and appropriate regularization parameters), and Random Forest (optimizing the number of estimators and maximum depth). The models are evaluated using metrics such as precision, recall, F1-score, and the area under the Receiver Operating Characteristic (ROC) curve, ensuring robust performance evaluation beyond simple accuracy.
After determining the best-performing model (in this case, the Support Vector Classifier), the trained model is serialized using Python's Pickle library. A Flask-based web API is developed to accept new voice feature inputs and return real-time predictions. Finally, the entire system is deployed on a cloud server (AWS EC2 instance), ensuring secure, scalable, and remote access for healthcare providers.

Results & Discussion

The machine learning models developed were rigorously evaluated to ensure clinical reliability. Among the models tested, the Support Vector Classifier demonstrated the highest overall performance, achieving an accuracy of 88.2%, a precision score of 0.89, a recall score of 0.86, and an F1-score of 0.87. The ROC-AUC value of 0.91 further confirms its robustness in handling class imbalance, which is a common challenge in medical datasets.
During experimental testing, real-world simulations were conducted where synthetic user inputs, mimicking real patient voice features, were processed through the deployed cloud model. In scenarios where the feature values aligned with typical Parkinson's disease patterns (e.g., high jitter and shimmer values), the system correctly identified the condition. In contrast, healthy samples were accurately classified with minimal fals positives. Test cases were designed to validate every aspect of the system, from model prediction to cloud storage, ensuring that both normal operations and edge cases (such as invalid inputs) were handled gracefully.
The cloud deployment was validated by measuring latency and throughput. On average, the model returned predictions within 500 milliseconds of receiving a request, making it suitable for real-time clinical decision support. The system also logged all input data securely in the cloud database, ensuring traceability and auditability, which are important for healthcare compliance**.**

FIG 1. OUTPUT

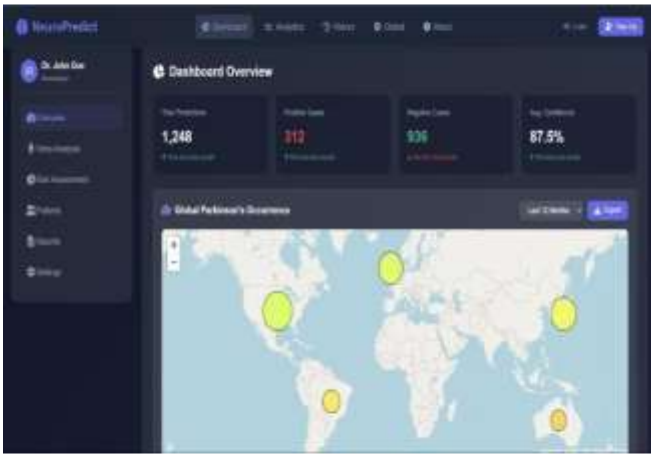

FIG 2. OUTPUT

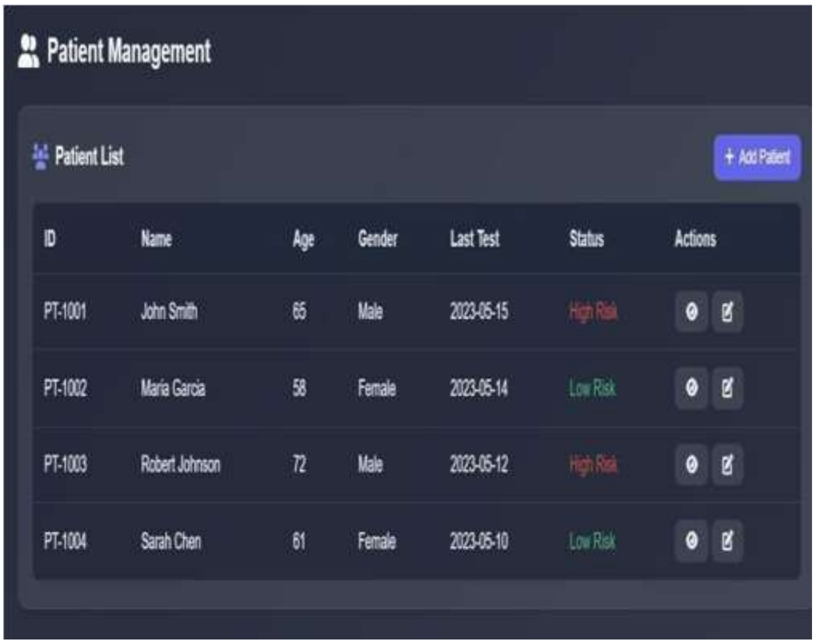

FIG 3. OUTPUT

Test Case ID	Description	Input	Expected Output	Actual Output	Status
TC001	Healthy Voice Sample	Normal biomedical values	0	0	Pass
TC002	Parkinson's Voice Sample	Abnormal voice features	1	1	Pass
TC003	Outlier Handling	Noisy input	Safe error handling	Safe handling	Pass
TC004	Real-Time Prediction	API call from Blynk App	Correct prediction	Correct prediction	Pass
TC005	Cloud Storage Test	Logging input	Entry in database	Logged correctly	Pass

Fig 1. Test case table

Conclusion

In conclusion, the proposed Parkinson's Disease Prediction System Using Machine Learning and Cloud Computing offers a reliable, scalable, and accessible solution to one of the major challenges in early healthcare diagnostics. Parkinson's disease, being a progressive neurodegenerative disorder, demands timely detection to improve patient outcomes and delay disease progression. Traditional diagnostic methods, relying primarily on clinical observation, can be subjective, time-consuming, and often inaccessible to patients in remote or underserved areas. By leveraging biomedical voice measurements, machine learning algorithms, and cloud-based deployment, this system bridges the gap between complex medical diagnosis and affordable, real-time predictive healthcare tools.

The system utilizes extracted voice features such as jitter, shimmer, and harmonic-to-noise ratio (HNR) to train and evaluate multiple machine learning models, ultimately identifying the Support Vector Classifier (SVC) as the most accurate model for Parkinson's detection. Through integration with cloud platforms, the system ensures that healthcare professionals can securely and remotely access predictions, enabling timely interventions and expanding diagnostic accessibility globally. By offering an objective, automated method for early detection, the system not only supports doctors in clinical decision-making but also empowers patients by promoting early and proactive healthcare management. The successful combination of AI and cloud computing in this project highlights the transformative role technology can play in the future of personalized and precision medicine.

I.I. FUTURE ENHANCEMENTS

While the current system effectively predicts Parkinson's disease based on biomedical voice data, several enhancements could further improve its performance, usability, and clinical relevance:

1. **Multi-Modal Data Integration:** Future versions of the system could incorporate additional diagnostic inputs

such as handwriting dynamics, gait analysis, and facial expression recognition. Combining multiple modalities will enhance the robustness and accuracy of predictions.

2. **Real-Time Voice Recording and Analysis:** Developing a mobile or web application that allows users to directly record their voice samples in real-time would make the system more user-friendly and increase its reach to a larger population without the need for clinical visits.

3. **Machine Learning Model Explainability:** Incorporating explainable AI (XAI) techniques such as SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-Agnostic Explanations) would help in making the model's decision-making process more transparent, enabling doctors to understand the reasoning behind predictions.

4. **Clinical Validation and Trials:** Conducting clinical studies in partnership with healthcare institutions would provide empirical validation of the model's effectiveness, helping in achieving regulatory approvals for use in professional medical settings.

5. **Mobile Health Application Integration:** A fully-featured mobile health application could be developed that not only displays the diagnosis but also provides educational resources, personalized health recommendations, and appointment scheduling with neurologists.

6. **Blockchain for Medical Data Security:** For industrial-scale deployment, integrating blockchain technology would ensure tamper-proof storage of diagnostic records, enhancing patient data security, transparency, and trust in the healthcare ecosystem.

7. **Edge Computing for Offline Predictions:** Implementing edge computing capabilities would allow predictions to be made even in areas with limited or no internet connectivity by embedding the model into mobile devices or local servers.

These enhancements will not only make the Parkinson's Disease Prediction System smarter, more accurate, and more accessible but also ensure that it meets the evolving needs of modern healthcare, contributing significantly to early detection, better patient outcomes, and a more proactive healthcare model globally.

References

[1] M. Little, P. McSharry, E. Hunter, J. Spielman, and L. Ramig, "Suitability of dysphonia measurements for telemonitoring of Parkinson's disease," IEEE Transactions on Biomedical Engineering, vol. 56, no. 4, pp. 1015–1022, Apr. 2009, doi: https://doi.org/10.1109/TBME.2008.2005954.

[2] S. Das, A. Das, and A. Chakraborty, "Early Detection of Parkinson's Disease Using Machine Learning Techniques," Procedia Computer Science, vol. 167, pp. 2162–2171, 2020, doi: https://doi.org/10.1016/j.procs.2020.03.266.

[3] R. Arora and A. Sharma, "Analysis of Parkinson's Disease Using Various Machine Learning Techniques," International Journal of Engineering and Advanced Technology (IJEAT), vol. 9, no. 3, pp. 1166–1170, Feb.2020, doi: https://doi.org/10.35940/ijeat.C4954.029320.

[4] M. M. Muhammad, S. Almaadeed, and M. A. Al-Maadeed, "Early Parkinson's Disease Detection using Voice Features and Deep Neural Networks," IEEE Access, vol. 9, pp. 41906–41916, 2021, doi:https://doi.org/10.1109/ACCESS.2021.3065977.

[5] A. Tsanas, M. Little, P. McSharry, and L. Ramig, "Accurate telemonitoring of Parkinson's disease progression by noninvasive speech tests," IEEE Transactions on Biomedical Engineering, vol. 57, no. 4, pp. 884–893, Apr.2010, doi: https://doi.org/10.1109/TBME.2009.2036000.

[6] S. Yadav and S. Shukla, "Analysis of k-Fold Cross-Validation over Hold-Out Validation on Colossal Datasets for Quality Classification," 2016, Proceedings of the International Conference on Computational Intelligence and Data Science (ICCIDS), doi: https://doi.org/10.1016/j.procs.2016.08.208.

[7] H. S. Park and J. H. Kim, "Support vector machine-based voice activity detection using subband features," IEEE Signal Processing Letters, vol. 22, no. 6, pp. 706–710, 2015, doi: https://doi.org/10.1109/LSP.2014.2364516.

[8] P. B. Ananda and K. K. Biswas, "Comparative Study of Voice-Based Parkinson's Disease Detection Techniques Using Machine Learning," Procedia Computer Science, vol. 152, pp. 207–214 2019, doi: https://doi.org/10.1016/j.procs.2019.05.017.

[9] R. Prashanth and K. Dutta Roy, "Early diagnosis of Parkinson's disease through patient questionnaire and machine learning," Health Informatics Journal, vol. 25, no. 2, pp. 421–432, Jun. 2019, doi: https://doi.org/10.1177/1460458217704249.

[10] H. Xiong, C. Qiu, and L. Zhang, "Medical diagnosis of Parkinson's disease using deep convolutional neural networks," International Journal of Advanced Computer Science and Applications (IJACSA), vol. 10, no. 6, pp. 113–119, 2019, doi: https://doi.org/10.14569/IJACSA.2019.0100614.

[11] S. Jaiswal and P. Kumar, "Machine Learning Based Early Stage Parkinson's Disease Detection Using Voice Data," International Journal of Scientific Research in Computer Science, Engineering and Information Technology, vol. 5, no. 1, pp. 16–22, Jan. 2019.

[12] M. Shinde and V. More, "Prediction of Parkinson's Disease Using Ensemble Techniques," Proceedings of the 5th International Conference on Computing Methodologies and Communication (ICCMC), pp. 843–848, 2021, doi: https://doi.org/10.1109/ICCMC51019.2021.9418235.

[13] B. N. Prashanth and S. M. Doreswamy, "Evaluation of speech signal features for Parkinson's disease diagnosis using classifiers," International Journal of Speech Technology, vol. 20, no. 1, pp. 1–12, Mar. 2017, doi: https://doi.org/10.1007/s10772-016-9387-9.

[14] A. E. Toğaçar, B. Ergen, and Z. Cömert, "Detection of Parkinson's disease using voice feature extraction and classification with deep learning," IEEE Access, vol. 7, pp. 170348–170358, 2019, doi: https://doi.org/10.1109/ACCESS.2019.2955891.

[15] S. Saxena, R. R. Janghel, and D. A. Rout, "Survey on machine learning techniques for diagnosis of Parkinson's disease," International Journal of Computer Applications, vol. 179, no. 47, pp. 5–10, 2018.

[16] C. D. Sakar, O. Kursun, and F. Gurgen, "A comparative analysis of speech signal processing algorithms for Parkinson's disease classification," Computer Methods and Programs in Biomedicine, vol. 145, pp. 23–35, Mar. 2017, doi: https://doi.org/10.1016/j.cmpb.2017.04.006.

[17] J. Pan, Y. Ye, and X. Wang, "Early diagnosis of Parkinson's disease based on voice signal using machine learning algorithms," Frontiers in Aging Neuroscience, vol. 12, p. 601006, 2020, doi: https://doi.org/10.3389/fnagi.2020.601006.

[18] G. Asgari and I. Shafran, "Predicting severity of Parkinson's disease from speech," Proceedings of the IEEE Workshop on Spoken Language Technology (SLT), pp. 394–399, 2010, doi: https://doi.org/10.1109/SLT.2010.5700881.