**Advanced Parkinson’s Prediction System Using Random Forest Algorithm and Feature Engineering**

A. Alagar1, U. Balaji2, S.S. Barath3, K. Gokul Krishnan4

*1Assistant Professor, Department of Computer Science and Engineering, K.L.N. College of Engineering, Sivagangai, Tamilnadu, India.*

*2,3,4 Final Year Students, Department of Computer Science and Engineering, K.L.N. College of Engineering, Sivagangai, Tamilnadu, India.*

|  |  |
| --- | --- |
|  | ***Abstract:*** *This* *system offers an enhanced Parkinson’s Disease Prediction Model utilizing machine learning for prompt diagnosis. By utilizing health and lifestyle information, it utilizes Random Forest for categorization, assessing its precision against K-Nearest Neighbors (KNN). The model goes through data preprocessing, feature engineering, and assessment to improve prediction accuracy. An interface built on Streamlit enables users to enter patient information and obtain real-time predictions along with confidence scores. Furthermore, analyzing feature importance assists in recognizing crucial contributing elements, aiding in medical decision-making. By prioritizing real-time forecasting, model interpretability, and scalability, this system supports early identification and improved disease management. Upcoming enhancements involve speech and walking assessment, incorporation of genetic information, and advanced deep learning frameworks for improved precision and usability .*  ***Key Word****:**Machine Learning, Parkinson’s Disease Prediction, Random Forest, K-Nearest Neighbors (KNN), Data Preprocessing, Feature Engineering, Model Evaluation, Confidence Score* |

**I. INTRODUCTION**

Parkinson's disease is a progressive neurological condition that impacts movement, making early identification vital for proper management. This system utilizes machine learning methods to estimate the probability of Parkinson's disease using health and lifestyle information. Employing Random Forest for classification and assessing its performance against K-Nearest Neighbors (KNN) guarantees high accuracy and dependability. Data preprocessing and feature engineering improve model performance, while an intuitive Streamlit interface enables smooth user interaction for real-time predictions accompanied by confidence scores. Furthermore, analyzing feature importance reveals significant elements affecting the disease, supporting more effective medical decision-making. This system, created for scalability and real-time forecasting, offers an innovative and understandable method for predicting Parkinson’s disease. Upcoming enhancements seek to

combine speech and gait analysis, genetic information, and deep learning methods to improve accuracy and functionality.

**II. OBJECTIVE**

The objective of this system is to create a precise and effective model for predicting Parkinson’s disease through the use of machine learning methods. The goal is to examine health and lifestyle information, using Random Forest for classification and assessing its effectiveness against KNN. The system offers instant predictions via an easy-to-use interface, improving access for prompt diagnosis. Moreover, it emphasizes the analysis of feature significance to assist healthcare practitioners in pinpointing crucial risk elements for improved decision-making.

**III. LITERATURE SURVEY**

**1. Machine Learning in Parkinson’s Disease Diagnosis:**

Conventional diagnostic techniques for Parkinson’s disease depend on clinical evaluations, neurological assessments, and medical imaging, which tend to be lengthy and subjective. Recent research has investigated machine learning strategies to enhance the accuracy of early detection by utilizing health and lifestyle information. Methods including Support Vector Machines (SVM), Decision Trees, and Neural Networks have shown encouraging outcomes in recognizing symptoms of Parkinson’s. Nonetheless, selecting features and preprocessing data continue to be significant obstacles in optimizing these models for real-world use.

**2. Feature Importance and Its Role in Prediction Accuracy:**

Precise disease prediction relies on pinpointing the most pertinent features that aid in diagnosis. Research on feature selection methods, including Principal Component Analysis (PCA) and Recursive Feature Elimination (RFE), emphasizes their importance in enhancing model efficiency. In predicting Parkinson's, vocal patterns, hand tremors, and lifestyle elements are some of the primary indicators. Utilizing explainable AI techniques, like SHAP and LIME, improves model interpretability, offering healthcare providers essential insights into significant risk factors.

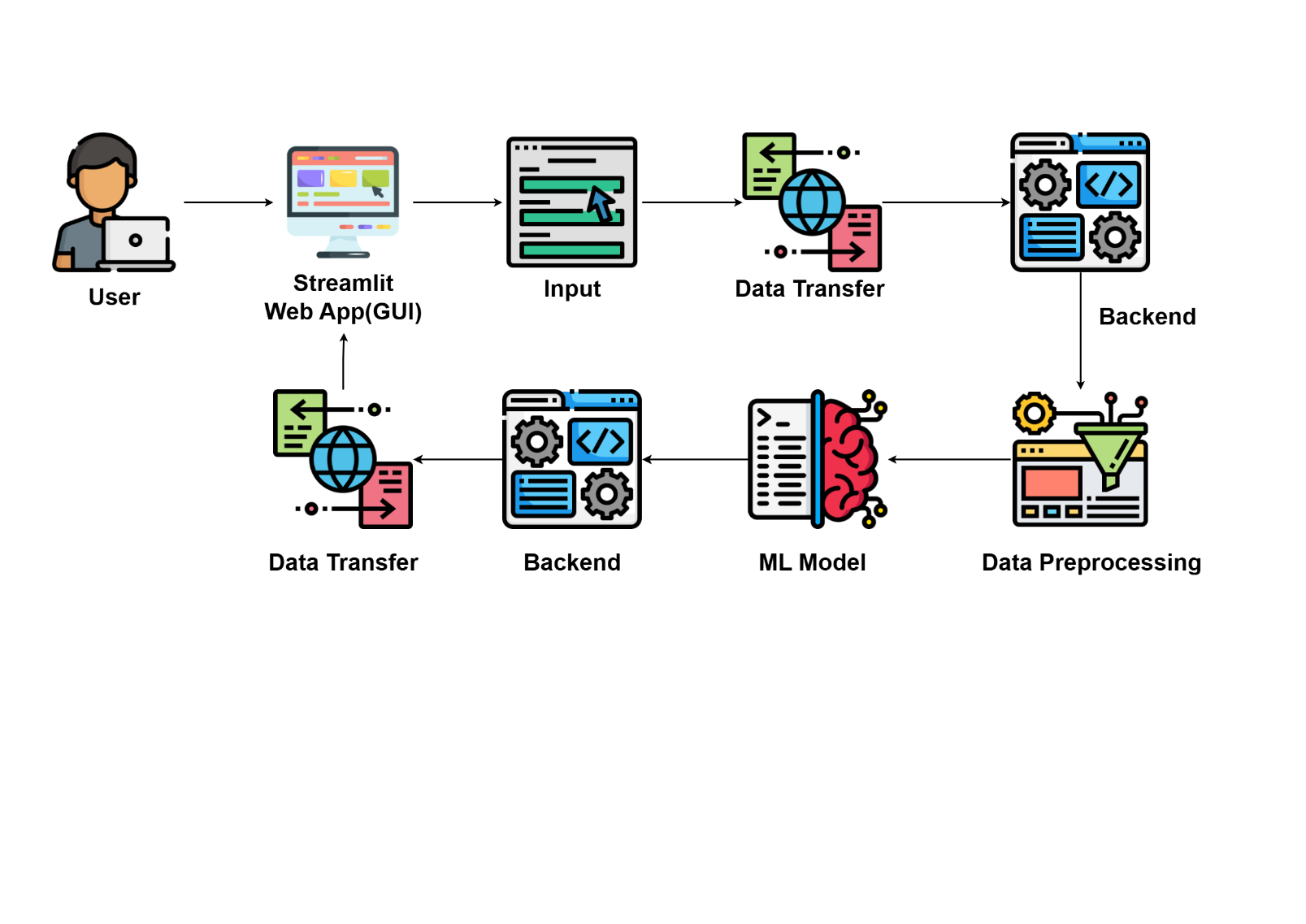
**IV. EXISTING SYSTEM**

Parkinson’s disease by examining voice metrics through a machine learning model. It presents group-wise scaling to minimize biases, enhancing accuracy by 9. 5% compared to traditional techniques. The model reaches 82% accuracy on previously unseen data and employs SHAP values for clarity. Results show that monotonic voice patterns with regular pauses are significant signs of Parkinson’s disease.

**V. PROPOSED SYSTEM**

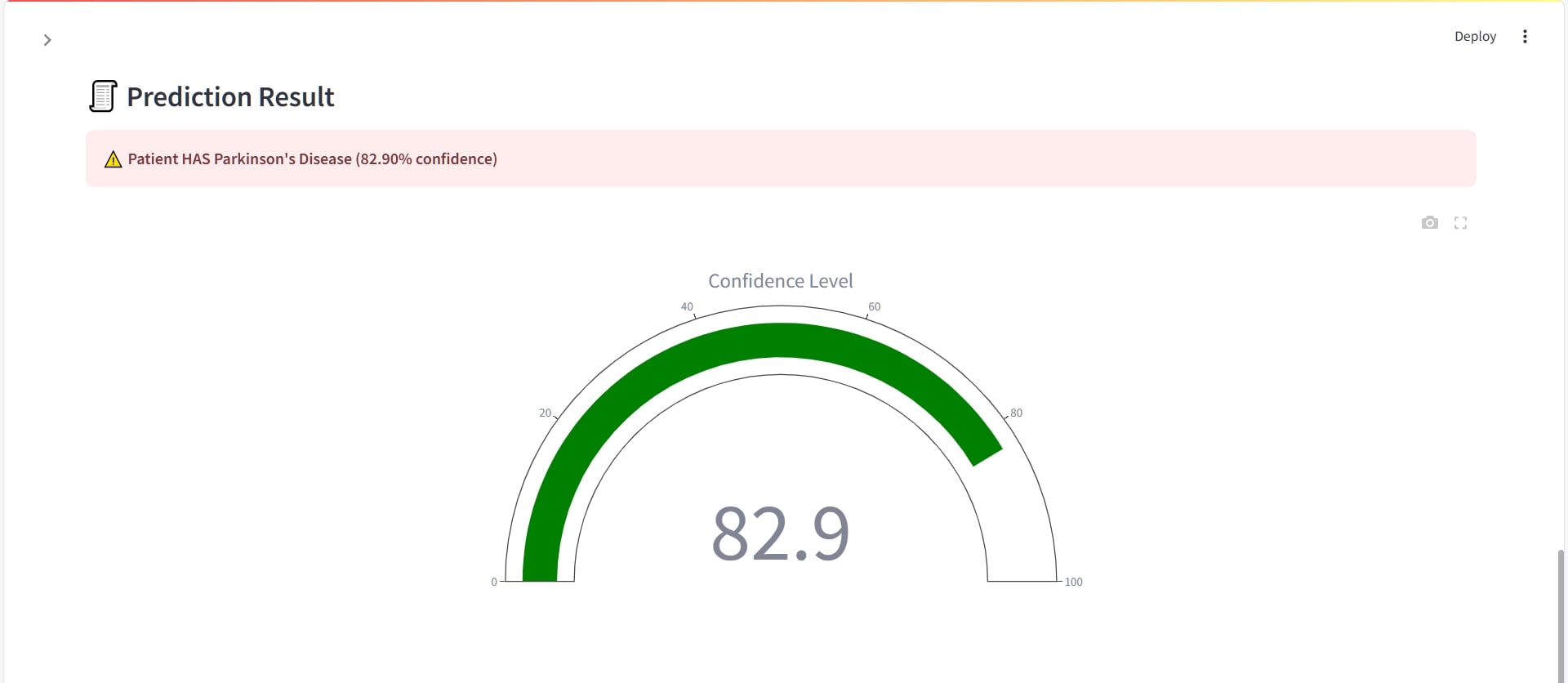
Utilizes a Random Forest model, reaching an accuracy of about 93%, which exceeds that of K-Nearest Neighbors (KNN) in classification efficacy. An online interface and command-line scripts facilitate smooth interaction, empowering users to train models and create predictions. Confidence scores paired with predictions improve result clarity, assisting healthcare professionals in their decision-making processes. Future enhancements might integrate deep learning and real-time monitoring for improved efficiency and flexibility.

**VI. ARCHITECTURE DIAGRAM**

****

**Fig 6.1 Architecture Diagram**

**VII. SYSTEM OVERVIEW**



**Fig 7.1 Prediction Result**

**1. Data Collection and Preprocessing:**

The system collects information on patient health and lifestyle, guaranteeing data consistency via preprocessing methods. Missing data, anomalies, and unnecessary features are managed to enhance model dependability. Feature engineering is utilized to derive significant patterns from the dataset. The refined data is subsequently employed for training and assessment.

**2. Machine Learning Model Training:**

The system develops both Random Forest and K-Nearest Neighbors (KNN) models to predict Parkinson’s disease. Random Forest reaches an accuracy of 93%, surpassing KNN in classification effectiveness. The training process includes hyperparameter tuning to enhance model performance. The ultimately trained models are stored for immediate predictions.

**3. Prediction and Result Interpretation:**

Users enter patient information through a command-line interface or a web application built with Streamlit. The trained model analyzes the input and returns a prediction alongside a confidence score. Analyzing feature importance assists in clarifying which elements are most influential in the prediction. Findings are displayed clearly to facilitate interpretation by healthcare professionals.

**4. User Interface and Deployment:**

A web-based interface has been created with Streamlit, rendering the system reachable and user-friendly. The platform guarantees immediate interaction, enabling users to obtain instant disease forecasts. The system is built for scalability and can be implemented on cloud platforms for wider accessibility. Upcoming improvements might feature deep learning integration and real-time monitoring functionalities.

**VIII. CONCLUSION**

The Parkinson’s Disease Prediction System utilizes machine learning to improve early detection and healthcare results. With an accuracy of 93%, the Random Forest model surpasses KNN, offering dependable predictions accompanied by confidence scores. The intuitive Streamlit interface guarantees accessibility for researchers and healthcare practitioners alike. Upcoming improvements, such as the integration of deep learning and real-time monitoring, will enhance diagnostic accuracy and user experience.

**References**

1. N. Momeni, S. Whitling, and A. Jakobsson, "Interpretable Parkinson’s Disease Detection Using Group-Wise Scaling", IEEE Access, vol. 13, pp. 29147-29161, 2025.
2. N. Momeni, S. Whitling, and A. Jakobsson, "Detecting Parkinson’s disease using voice recordings from mobile devices", Proc. 32nd Eur. Signal Process. Conf. (EUSIPCO), pp. 1516-1520, Aug. 2024.
3. C. Botelho, A. Abad, T. Schultz, and I. Trancoso, "Speech as a Biomarker for Disease Detection", IEEE Access, vol. 12, pp. 184487-184508, 2024.
4. C. R. Dhivyaa, K. Nithya, and S. Anbukkarasi, "Enhancing Parkinson’s Disease Detection and Diagnosis: A Survey of Integrative Approaches Across Diverse Modalities", IEEE Access, vol. 12, pp. 158999-159024, 2024.
5. A. Siderowf et al., "Assessment of heterogeneity among participants in the Parkinson’s progression markers initiative cohort using α-synuclein seed amplification: A cross-sectional study", Lancet Neurol., vol. 22, no. 5, pp. 407-417, May 2023.
6. M. K. Reddy and P. Alku, "Exemplar-Based Sparse Representations for Detection of Parkinson's Disease From Speech", IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 31, pp. 1386-1396, 2023.
7. C. Dong, Y. Chen, Z. Huan, Z. Li, B. Zhou, and Y. Liu, "Static-Dynamic Temporal Networks for Parkinson’s Disease Detection and Severity Prediction", IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 31, pp. 2205-2213, 2023.
8. S. Turaev et al., "Review and Analysis of Patients’ Body Language From an Artificial Intelligence Perspective", IEEE Access, vol. 11, pp. 62140-62173, 2023.
9. S. J. J. Jui, R. C. Deo, P. D. Barua, A. Devi, J. Soar, and U. R. Acharya, "Application of Entropy for Automated Detection of Neurological Disorders With Electroencephalogram Signals: A Review of the Last Decade (2012–2022)", IEEE Access, vol. 11, pp. 71905-71924, 2023.
10. H. Khachnaoui, B. Chikhaoui, N. Khlifa, and R. Mabrouk, "Enhanced Parkinson’s Disease Diagnosis Through Convolutional Neural Network Models Applied to SPECT DaTSCAN Images", IEEE Access, vol. 11, pp. 91157-91172, 2023.