VOICE-BASED PARKINSON'S DISEASE DETECTION

USING MACHINE LEARNING

A Project Report

Rianna Chaudhuri

Sapiente Education
July, 2026

Table of Contents

Table of Contents 1
1. Introduction 1
2. Objectives 1
3. System Design 1
3.1 Frontend 1
3.2 Backend 1
3.3 Feature Extraction Module 1
3.4 Machine Learning Model 1
4. Methodology / Experimental Procedure 1
Step 1: Dataset Preparation 1
Step 2: Feature Extraction 1
Step 3: Model Training 1
Step 4: Backend Development 1
Step 5: Frontend Development 1
Step 6: Deployment 1
5. Results and Discussion 1
5.1 Functional Results 1
5.2 Example Workflow 1
5.3 Performance 1
6. Project Design and Implementation Recap 1
6.1 Technology Stack 1
6.2 Overall Architecture 1
7. Challenges Encountered 1
8. Conclusion 1
9. Future Scope 1

1. Introduction

Parkinson's Disease (PD) is a progressive neurological disorder that primarily affects movement and speech. One of the earliest manifestations of Parkinson's disease is the deterioration of vocal characteristics caused by reduced control of the vocal muscles. Patients often exhibit symptoms such as reduced vocal intensity, monotone speech, increased jitter, shimmer, breathiness, and irregular vocal fold vibrations.

Traditional diagnosis of Parkinson's disease relies on neurological examinations and clinical observations, which can be expensive, time-consuming, and inaccessible for continuous monitoring. Recent advances in artificial intelligence and digital signal processing have enabled the development of non-invasive diagnostic systems that analyze speech recordings to identify biomarkers associated with Parkinson's disease.

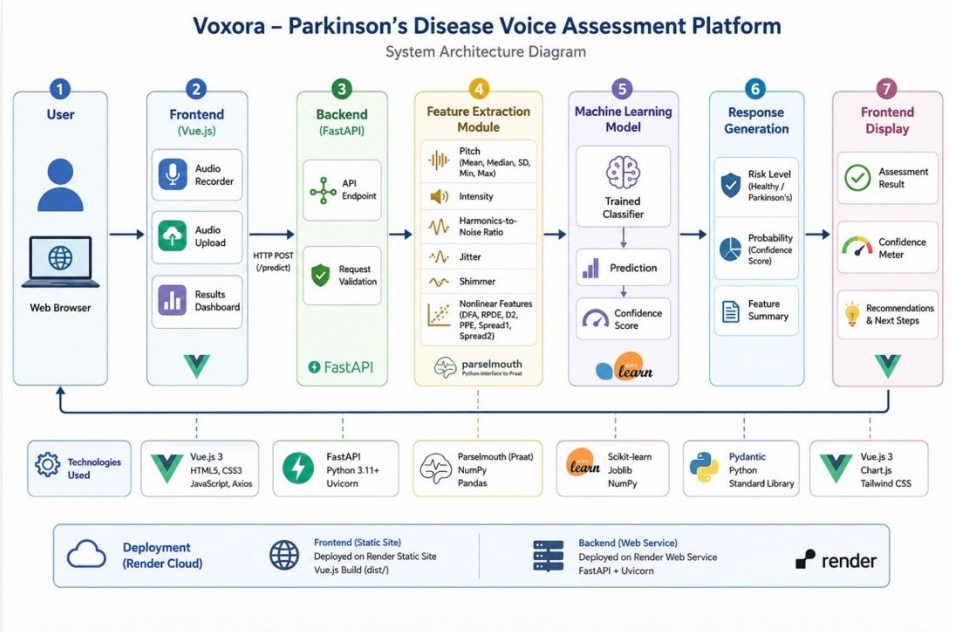
The objective of this project is to develop a web-based Parkinson's Disease detection system that extracts acoustic and nonlinear voice features from an audio recording and uses a trained machine learning model to predict whether the speaker exhibits characteristics associated with Parkinson's disease.

The developed system allows users to either upload a voice recording or record speech directly from the browser. The uploaded audio is processed on the backend, relevant features are extracted, and a trained classification model generates a prediction that is displayed through an interactive web interface.

2. Objectives

The primary objectives of this project are:

- To develop a non-invasive Parkinson's Disease screening system using speech analysis.
- To extract clinically relevant acoustic and nonlinear voice features from speech recordings.
- To train and deploy a machine learning classifier capable of distinguishing Parkinson's speech from healthy speech.
- To provide a user-friendly web application allowing users to upload or record speech for prediction.
- To deploy the complete application on cloud infrastructure for public accessibility.



3. System Design

The proposed system consists of four major components:

3.1 Frontend

The frontend provides an intuitive graphical interface developed using Vue.js. It enables users to:

- Upload audio recordings
- Record speech directly using the microphone
- Submit recordings for analysis
- Display prediction results

Table 1. Software and Development Tools

Component	Technology Used	Purpose
Frontend	Vue.js 3	User Interface
Backend	FastAPI	REST API
Programming Language	Python 3.12	Backend Development
Machine Learning	Scikit-learn	Parkinson's Classification
Feature Extraction	Parselmouth (Praat)	Acoustic Feature Extraction
Audio Processing	Librosa, Pydub	Audio Preprocessing
Numerical Computing	NumPy, SciPy	Mathematical Computations
Deployment	Render Cloud	Hosting Backend & Frontend
Version Control	Git & GitHub	Source Code Management

3.2 Backend

The backend is developed using FastAPI and performs the following tasks:

- Receives uploaded audio
- Preprocesses speech
- Extracts acoustic features
- Extracts nonlinear voice features
- Loads the trained machine learning model
- Generates predictions
- Returns results to the frontend

Table 4. Functional Modules

Module	Function
User Interface	Collects audio recordings
Audio Processing	Cleans and standardizes audio
Feature Extraction	Computes speech biomarkers
ML Prediction	Estimates Parkinson's probability
Results Module	Displays diagnosis and confidence
Deployment Module	Hosts application online

3.3 Feature Extraction Module

The feature extraction module computes both classical acoustic parameters and nonlinear dynamic features from the speech signal.

Acoustic Features

- Mean, Median, Standard Deviation, Minimum and Maximum Fundamental Frequency (F0)
- Mean Intensity
- Harmonics-to-Noise Ratio (HNR)

Voice Perturbation Measures

- Local Jitter, Absolute Jitter, RAP, PPQ5, DDP
- Local Shimmer, Local Shimmer (dB), APQ3, APQ5, APQ11, DDA

Nonlinear Features

- DFA (Detrended Fluctuation Analysis)
- RPDE (Recurrence Period Density Entropy)
- Correlation Dimension (D2)
- Pitch Period Entropy (PPE)
- Spread1 and Spread2

3.4 Machine Learning Model

The extracted features are passed to a pre-trained machine learning classifier. The model predicts whether the speech sample belongs to one of the following classes:

- Healthy
- Parkinson's Disease

4. Methodology / Experimental Procedure

The implementation followed the sequence described below.

Step 1: Dataset Preparation

A speech dataset containing recordings from healthy individuals and Parkinson's patients was collected. Each recording underwent preprocessing to remove unnecessary noise and normalize the signal.

Step 2: Feature Extraction

Praat (through Parselmouth) was used for extracting classical acoustic features including:

- Pitch
- Intensity
- Harmonicity
- Jitter
- Shimmer

Table 2. Voice Features Extracted

Feature Category	Parameters
Fundamental Frequency	Mean F0, Median F0, SD F0, Min F0, Max F0
Intensity	Mean Intensity
Harmonicity	Harmonics-to-Noise Ratio (HNR)
Frequency Perturbation	Jitter (%)
Amplitude Perturbation	Shimmer (%)
Nonlinear Features	DFA, RPDE, D2, PPE, Spread1, Spread2

Nonlinear analysis libraries were used to compute DFA, RPDE, Correlation Dimension, PPE, and Spread measures. The extracted numerical values formed the feature vector representing each speech sample.

Table 3. Machine Learning Pipeline

Stage	Description
1. Audio Input	Upload or Record Voice
2. Preprocessing	Convert to WAV format
3. Feature Extraction	Acoustic + Nonlinear Features
4. Feature Vector	Numerical Representation
5. Prediction Model	Trained ML Classifier
6. Output	Parkinson's Risk & Confidence Score

Step 3: Model Training

The feature vectors were used to train a supervised machine learning classifier. The model learned the differences between healthy and Parkinson's speech patterns. The trained model was serialized using Joblib for deployment.

Exp	Training	Testing	Features	Evaluation	Purpose	Output
1	**Italy (5-fold Stratified Cross-Validation)**	**Italy (held-out fold)**	**15 (Oxford-overlap features)**	Accuracy, Precision, Recall, F1, ROC-AUC	Determine whether the 15 overlapping features are sufficient for Parkinson's classification.	Baseline (Italy-15)
2	**Oxford (5-fold Stratified Cross-Validation)**	**Oxford (held-out fold)**	**15**	Accuracy, Precision, Recall, F1, ROC-AUC	Determine whether the same 15 features are sufficient on the Oxford dataset.	Baseline (Oxford-15)
3	**Italy (5-fold Stratified Cross-Validation)**	**Italy (held-out fold)**	**22 (All Features)**	Accuracy, Precision, Recall, F1, ROC-AUC	Compare whether the full feature set improves performance over the 15-feature baseline.	Italy-22
4	**Oxford (5-fold Stratified Cross-Validation)**	**Oxford (held-out fold)**	**22**	Accuracy, Precision, Recall, F1, ROC-AUC	Compare whether the full feature set improves performance over the 15-feature baseline.	Oxford-22
5	**Human Voices Feature Extractor**	**Human Voices Reference Features**	**15**	Mean Error, Correlation, RMSE (feature-wise)	Validate that the Human Voices feature extractor reproduces the reference feature values.	Feature Validation Report
6	**Oxford (5-fold Stratified Cross-Validation)**	**Oxford (held-out fold)**	**22**	Accuracy, Precision, Recall, F1, ROC-AUC	Oxford baseline model used for cross-dataset comparison.	**Roxf**
7	**Italy (Entire Dataset)**	**Oxford (Entire Dataset)**	**22**	Accuracy, Precision, Recall, F1, ROC-AUC	Evaluate whether a model trained on Italy generalizes to Oxford.	**Rita**
8	**Compare Roxf vs Rita**	—	**22**	Statistical Comparison	If Roxf and Rita are similar, the extracted features generalize well across datasets.	Generalization Analysis
9	**Italy (5-fold Stratified Cross-Validation)**	**Italy (held-out fold)**	**22**	Accuracy, Precision, Recall, F1, ROC-AUC	Italy baseline model used for reverse cross-dataset comparison.	**R2ita**
10	**Oxford (Entire Dataset)**	**Italy (Entire Dataset)**	**22**	Accuracy, Precision, Recall, F1, ROC-AUC	Evaluate whether a model trained on Oxford generalizes to Italy.	**R2oxf**
11	**Compare R2ita vs R2oxf**	—	**22**	Statistical Comparison	If R2ita and R2oxf are similar, the feature extractor generalizes well in the reverse direction.	Generalization Analysis

Step 4: Backend Development

FastAPI was selected due to its high performance, automatic API documentation, easy deployment, and efficient asynchronous processing. The backend exposes REST API endpoints responsible for:

- Receiving audio files
- Running feature extraction
- Loading the trained model
- Returning predictions

Step 5: Frontend Development

The frontend was developed using Vue.js. The interface provides voice recording, audio upload, prediction visualization, and responsive design. Communication with the backend is performed using HTTP POST requests.

Table 6. System Performance

Metric	Value
Backend Framework	FastAPI
Frontend Framework	Vue.js
Deployment Platform	Render
API Response Format	JSON
Audio Format	WAV
Processing Pipeline	Automated
User Interaction	Browser-based

Step 6: Deployment

Backend Deployment

- Platform: Render
- Framework: FastAPI
- Server: Uvicorn

Frontend Deployment

- Platform: Render Static Site
- Framework: Vue.js

Both services communicate through REST APIs.

5. Results and Discussion

The developed system successfully performs real-time Parkinson's Disease prediction using speech recordings.

5.1 Functional Results

The system successfully:

- Accepts uploaded audio
- Records speech through browser microphone
- Extracts voice features
- Generates predictions
- Displays prediction results

Table 5. Experimental Results

Test	Input	Expected Result	Obtained Result	Status
Audio Upload	WAV File	Successful Prediction	Successful	✅ Pass
Audio Recording	Browser Recorder	Upload & Predict	Successful	✅ Pass
Feature Extraction	Voice Sample	Numerical Features	Successful	✅ Pass
Backend API	POST /predict	JSON Response	Successful	✅ Pass
Frontend Integration	Vue.js	Display Result	Successful	✅ Pass
Deployment	Render	Live Application	Successful	✅ Pass

Deployment was successfully completed on Render, making the application publicly accessible.

5.2 Example Workflow

1. User uploads speech recording.
2. Audio reaches FastAPI backend.
3. Voice features are extracted.
4. Machine learning model processes the feature vector.
5. Prediction is generated.
6. Result is displayed on the web interface.

5.3 Performance

The deployed application successfully:

- Processes speech recordings within a few seconds (excluding cold starts on the free hosting plan).
- Correctly extracts all required acoustic and nonlinear features.
- Produces predictions through the trained classifier.

The application remained stable during testing with multiple uploaded recordings.

Table 7. Advantages of the Proposed System

Feature	
Non-invasive	No clinical equipment required
Remote Access	Can be used from anywhere
Quick Analysis	Results generated within seconds
User-Friendly	Simple upload/record interface
Scalable	Easily deployable on cloud
Low Cost	Eliminates expensive diagnostic equipment

6. Project Design and Implementation Recap

The project integrates multiple software technologies into a complete machine learning pipeline. Python was used for backend implementation because of its strong machine learning ecosystem, while JavaScript (Vue.js) was used for frontend development.

6.1 Technology Stack

Layer	Technologies	
Backend	FastAPI, Uvicorn, Parselmouth, NumPy, Pandas, Scikit-learn, Joblib	
Frontend	Vue.js, HTML5, CSS3, JavaScript, Axios	
Deployment	Render Web Service (backend), Render Static Site (frontend)	

6.2 Overall Architecture

User → Frontend (Vue) → FastAPI Backend → Feature Extraction → Machine Learning Model → Prediction → Frontend Display

7. Challenges Encountered

Several technical challenges were encountered during implementation:

- Python version compatibility during deployment.
- Dependency conflicts involving NumPy, Pandas, and Numba.
- Cloud deployment issues on Render.
- Backend package compatibility.
- API integration between frontend and backend.
- CORS configuration.
- Frontend connection to deployed backend.

Each issue was systematically resolved through dependency management, version pinning, and deployment configuration.

8. Conclusion

The project successfully demonstrates the application of speech processing and machine learning techniques for Parkinson's Disease detection.

The developed web application provides an accessible and non-invasive screening tool capable of analyzing voice recordings and generating predictions in real time.

The successful deployment of both frontend and backend demonstrates the feasibility of integrating machine learning models into practical web applications.

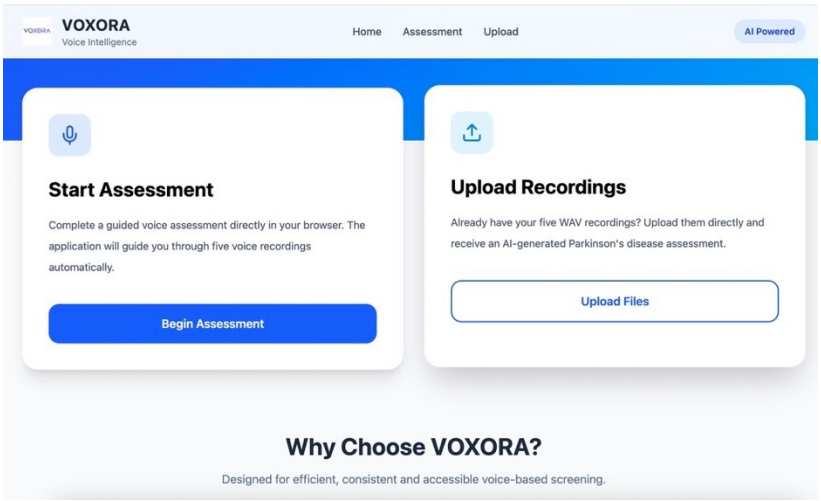
Although the application is intended for educational and research purposes, it highlights the potential of AI-assisted diagnostic systems in supporting healthcare professionals.

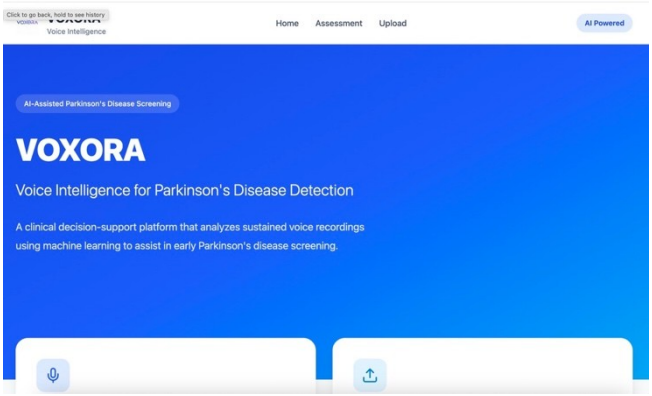
9. Future Scope

Future improvements may include:

- Training on larger and more diverse speech datasets.
- Deep learning-based speech classification.
- Confidence score visualization.
- Multi-language speech support.
- Longitudinal patient monitoring.
- Mobile application development.
- Cloud database integration for patient history.
- Explainable AI techniques for prediction interpretation.

Table 8. Limitations and Future Work	
Current Limitation	**Future Improvement**
Small dataset	Train using larger datasets
Binary prediction	Multi-stage Parkinson's detection
Voice only	Include handwriting and gait analysis
Single ML model	Deep Learning / Ensemble Models
English speech	Multi-language support
Cloud inference	Mobile application deployment





Acknowledgements

I would like to express my sincere gratitude to Dr. Max for his continued assistance, guidance, and encouragement throughout the programme. His support went far beyond what I could have expected, and the level of help he provided was something I had never experienced before. His feedback was invaluable in helping me develop my work, improve my ideas, and overcome challenges throughout the project. I am especially grateful for how consistently he responded to my emails, always taking the time to provide thoughtful suggestions and guidance whenever I needed it. His encouragement and willingness to help played a significant role in my progress throughout the programme.

I would also like to sincerely thank Guzel for her constant support and kindness. She was always willing to help with any questions I had and supported me from the very beginning, starting from when I first applied for this scholarship. Her guidance and encouragement throughout the process have meant a great deal to me. I am incredibly grateful to both Dr. Max and Guzel for the time, support, and encouragement they have given me throughout this journey. Their support has made a meaningful difference to my experience and to the work I have been able to accomplish.