Detection of Parkinson's Disease Using Machine Learning

Parkinson's – A Disorder of the Central Nervous System

Parkinson's disease is a neurological disorder that affects movement, often including tremors.
It can be **detected through speech**, as the *phonation* (the sound made when pronouncing vowels) is the most affected part of the speech pattern.

We have used a database of speech samples containing phonation recordings from both affected and healthy individuals. Several standard datasets are available from sources such as the **Journal of the Acoustic Society of America (JASA)** and the **UCI Machine Learning Repository**.

Project Overview

Speech signals or voice samples were taken from the **standard UCI voice dataset**, which contains recordings of 56 individuals. Each person was asked to pronounce the sustained vowels **"a"** and **"o"** three times respectively, resulting in a total of **336 recordings**. Healthy voice samples were also included for comparative study.

During the **training phase**, preprocessing of the voice signals was performed using **Praat** software to extract biomedical voice features. These extracted parameters include:

- Jitter
- Shimmer
- NHR (Noise-to-Harmonics Ratio)
- HNR (Harmonics-to-Noise Ratio)
- Mean and median pitch
- Number of pulses and periods
- Minimum and maximum period
- Standard deviation of pitch and period
- Number and degree of voice breaks

These features vary from patient to patient depending on the progression of Parkinson's disease.

After feature extraction, **dimensionality reduction** was performed using **Particle Swarm Optimization (PSO)**, an optimization method that mimics swarm behavior to select the most relevant features.

Finally, these optimized features were used to **train a Support Vector Machine (SVM) classifier**, which forms the core of the detection model.

🧠 Parkinson's Disease Detection Using Machine Learning

Tools & Technologies

Python, scikit-learn, Tkinter (GUI), Parselmouth (Praat), Pandas, NumPy, CSV, SVM, Pickle

---

Project Description

A machine learning–based system was developed to **detect Parkinson's disease from voice recordings** using biomedical signal analysis.

The model predicts whether a person is *healthy* or *Parkinsonian* by analyzing vocal parameters extracted from speech samples.

Key Components

1. Data Source:
Imported dataset containing approximately **1,000 voice samples (rows)** and **24 biomedical voice parameters**, such as:
MDVP:Fo(Hz), MDVP:Jitter(%), MDVP:Shimmer, HNR, RPDE, DFA, spread1, spread2, PPE, etc.

2. Feature Extraction:
Acoustic features (pitch, jitter, shimmer, harmonic-to-noise ratio) were extracted from .wav files using the **Parselmouth (Praat)** library for phonetic signal processing.

3. Model Used:
A **Support Vector Machine (SVM)** classifier was trained to distinguish between healthy and Parkinson's-affected individuals.

4. Model Storage:
The trained model was serialized using **pickle** and stored as svmclassifier.pkl for real-time prediction.

5. Testing Data:
The test.csv file was used for validation, ensuring consistent model performance on unseen data.

6. GUI Implementation:
A **Tkinter-based desktop interface** was developed to allow users to upload a .wav voice sample.
The system automatically:

- Extracts vocal features,

- Applies the trained SVM model, and

- Displays the diagnosis result ("Healthy" / "Parkinson's Detected").

7. Performance Metrics:
The SVM model achieved **90%+ classification accuracy** during validation testing.

Objective

To demonstrate how **machine learning and signal processing** can be applied in **healthcare diagnostics** for early, non-invasive detection of Parkinson's disease using voice analysis.

💼 **Skills Highlighted**

- Supervised Machine Learning (SVM Classification)

- Audio Signal Feature Extraction (Praat / Parselmouth)

- Data Preprocessing and Statistical Analysis (Pandas, NumPy)

- Model Deployment via GUI (Tkinter)

- Dataset Handling and Model Persistence using Pickle

- Optimization using Particle Swarm Optimization (PSO)

📄 **Short Resume Version**

- Developed a **Parkinson's Disease Detection System** using an **SVM classifier** trained on a voice dataset (~1,000 samples, 24 features).

- Implemented **feature extraction** from .wav files using **Parselmouth (Praat)** to derive acoustic parameters (jitter, shimmer, pitch).

- Built a **Tkinter-based GUI** for real-time diagnosis by uploading a voice file and predicting results using a pre-trained model (svmclassifier.pkl).

- **Technologies Used:** Python, scikit-learn, Pandas, Tkinter, Parselmouth, Pickle

📃 Dataset Source

Name: Parkinson's Disease Data Set
Repository: UCI Machine Learning Repository

Creators:

Max A. Little, Patrick E. McSharry, Eric J. Hunter, Jennifer Spielman, and Lorraine O. Ramig

---

📊 Dataset Details

- **Number of Instances:** 195 samples (in original UCI dataset)
  → Expanded to ~1,000 entries in this project by combining and augmenting datasets.

- **Number of Attributes:** 24 biomedical voice parameters + 1 target label

- **Target Column:** status

    o   1 → Parkinson's Disease Present

    o   0 → Healthy Individual