PARKINSON DIAGNOSIS ASSISTANT USING MACHINE LEARNING

Sneha Agarwal

December 2023

Contents

1 Problem Statement

Big Data analysis using ML is required to manage large volume of medical data and derive insights from it in order to address a variety of medical and clinical problems. Numerous studies have already demonstrated that ML algorithms have significantly improved their performance in classification-based medical problems. This study's primary objective is 6 to enhance Parkinson's disease detection and diagnosis techniques. Parkinson's disease cannot be cured, but medications can often significantly alleviate its symptoms. Because of this, early detection is crucial. . Ultimately, the goal is to positively impact patient care by contributing an effective solution that advances the diagnosis of Parkinson's disease through the power of AI and data-driven insights.

2 Market/Customer/Business Need Assessment

The brain is the principal controller of the body. Therefore, harm to this vulnerable part of the human body will have far-reaching consequences for the rest of the body's systems. The disorder known as Parkinson's is one example. Numerous individuals aged 65 or older have an incurable neurodegenerative disease. If the disease is detected in its early phases, it can be contained.

As of right now, there are no laboratory tests that have been demonstrated to be successful in detecting Parkinson's disease in its early stages. As a result, it is incredibly challenging for medical professionals to effectively diagnose Parkinson's disease. Accurate diagnosis of the condition can be challenging, particularly in its earlier stages, when it is typically most treatable.

If these symptoms are not detected in the early phases, it is fatal. So far, researchers have identified the disease's symptoms and roots. But very few symptoms have been cured, and there are numerous symptoms for which there is no treatment. The disease's treatment can be very costly and most patients cannot afford it. But, If the disease can be diagnosed at an early stage, it is possible to reduce the cost of treatment. In this era of increasing Parkinson's disease, it is crucial to discover a solution that can predict the disease in its early stages.

3 Target Specifications and Characterization

1. **Early Detection** : Pay close attention to any small changes in symptoms that may point to the beginning of the illness before it causes substantial damage.

2. **Feature Selection and Extraction**: Define the primary features that the AI will pay attention to, such as gait analysis, voice traits, or other motor symptoms, ans how these features will be retrieved and handled.

3. **Support for Remote Consultation**: Facilitate remote consultations between patients and healthcare professionals, enabling the sharing of data and insights obtained by the AI for better-informed discussions and decision-making.

4. **Education and Awareness**: Provide tools that inform users, patients, caregivers, and medical professionals about the signs and symptoms of Parkinson's disease, treatments that are available, and the significance of early detection.

5. **Research Institutions and Clinical Trials**: To improve the effectiveness of Parkinson's disease research, scientists might use the AI assistant to automate the process of selecting patients for clinical trials. The AI assistant can help with data processing and interpretation for institutions researching Parkinson's disease.

6. **Continuous Learning and Improvement**: Implement machine learning techniques that continuously learn from new data to enhance diagnostic accuracy and adapt to changes in symptom patterns.

7. **Accessibility and Affordability**: Strive for affordability and accessibility, making the AI assistant available to a wide range of populations, including those in remote or underprivileged areas.

4 External Search

In this project a dataset from the UCI Machine Learning Repository was utilized. The acquired data includes approximately 195 patient records, with each row comprising 23 distinct voice characteristics.

5 Benchmarking

Following table provides a succinct comparison outlining how AI-based Parkinson's disease detection offers advantages in accuracy, predictive capabilities,time consumption, scalability, continuous monitoring, and data reliability compared to traditional methods without AI.

Aspect	Diagnosis without AI	Diagnosis with AI
Diagnostic Method	Parkinson Disease detection without AI relies on clinical assessments by healthcare professionals based on observable symptoms, medical history, and physical examinations.	Parkinson Disease detection with AI analyzes various data sources (voice, movement,patterns, etc.) using AI algorithms for automated analysis and detection.
Accuracy	Diagnosis without AI might be subjective and accuracy varies between practitioners.	Enhanced accuracy due to AI's ability to detect subtle patterns and anomalies.
Ability to Predict	Limited ability to predict disease progression or detect early-stage symptoms.	Capable of predicting disease progression and detecting early-stage symptoms for proactive intervention.
Scalability	Limited reach, may not be accessible in remote or underserved areas because of absence of diagnostic centers and expensive procedures.	Mobile apps can reach more individuals, improving access to detection and care and providing inexpensive solutions
Time Consumption	Manual analysis can be a time consuming process.	Streamlines analysis, potentially reducing diagnosis time and enabling the effective use of medical resources.
Continuous Monitoring	Continuous monitoring of a patient's health can be a difficult task.	Facilitates continuous, real-time monitoring, providing insights into disease progression overtime
Cost and Resource Efficiency	Traditional methods might require significant resources and healthcare professional expertise	Potential for cost reduction through AI and optimized resource allocation, makes it more efficient.

6 Applicable Regulation

Here are some applicable regulations and considerations regarding the use of Machine Learning (ML) for Parkinson's disease detection.

1. **FDA Regulations** : Approval and clearance procedures for medical equipment and software, including machine learning (ML) algorithms,used in healthcare are governed by the U.S. Food and Drug Administration (FDA).
2. **Data Transparency and Explainability:**:ML models used in healthcare must be explainable and transparent to clinicians and patients,about how decisions are made.
3. **Informed Consent and Data Protection:**: Obtaining informed consent and safeguarding patient data are critical when dealing with health information.
4. **Validation**: Rigorous testing and evidence are needed to show the effectiveness and safety of AI-based diagnostic tools.
5. **GDPR Guidelines**: Any AI system analyzing patient data for Parkinson's detection must comply with GDPR (General Data Protection Regulation) guidelines, ensuring lawful and transparent use of personal information.
6. **Advertising Standards Authority (ASA - UK):** Several bodies provide guidelines and standards for truthful and accurate advertising. Any claims made regarding AI's capabilities in Parkinson's detection must be substantiated and not mislead consumers.
7. **Consumer Complaints:** Consumers can report false advertising claims to relevant authorities or regulatory bodies, triggering investigations and potential legal actions against false advertising practices.
8. **Validation and Certification:** Credibility and trust in the claims made by product can be ensured by obtaining validation or certification from respectable independent groups or organizations.
9. **Remote Access and Collaboration:**Utilizing remote access and collaborative platforms for research can reduce the need for extensive travel, leading to lower carbon emissions associated with commuting or travel-related activities.
10. **Energy Consumption:** Large-scale data processing, especially for AI models that require substantial computational power, may consume significant energy. Employing energy-efficient computing infrastructure can mitigate environmental impact.

7 Applicable Constraint

Several constraints and challenges which to be considered while using AI for detecting Parkinson's disease are.

1. Limited availability of quality and diverse datasets for training AI model affects the product's accuracy and generalizability.
2. Environmental or contextual factors that affect Parkinson's symptoms might not always be captured in datasets, affecting model robustness.
3. Lack of trust in AI-driven diagnostic tools among healthcare professionals and patients.
4. Significant demand for computational resources and expertise in developing and deploying AI models, poses challenges, especially for smaller healthcare facilities.
5. Continual need for updates and refinements to adapt to new data and improve accuracy over time.
6. Sometimes AI-related errors or misdiagnoses raise legal and ethical concerns.
7. Need for expertise in AI, data science, and healthcare requires a skilled workforce, which might not be readily available.
8. Financial Investment required for developing, deploying, and maintaining AI systems, poses challenges for smaller healthcare institutions.

8 Business Opportunity

1. Early detection through AI can lead to timely interventions, hence reducing the overall cost of treatment by managing the disease at an earlier stage.At the same time early identification can prevent or delay complications, reducing the need for expensive interventions ,hospitalizations.
2. Remote monitoring through AI can lead to cost-efficient follow-ups, minimizing the need for frequent travel to healthcare facilities.
3. Developers can obtain patents for these inventions and, in exchange for a licensing fee or royalties, grant licenses to healthcare organizations or businesses.
4. We can partner with pharmaceutical firms conducting Parkinson's disease research. Offer AI-driven analysis services on patient data or assist in drug development, earning revenue through service contracts or royalties.

5. A cloud-based diagnostic platform powered by AI for Parkinson's disease detection can be offered. Charge healthcare providers, clinics, or hospitals a subscription fee for access to the platform.
6. Offering different tiers based on usage, features, or support, provides a recurring revenue stream.

9 Concept Generation

In the present era, disease prognosis is the most essential duty of medical institutions and physicians in order to make the best possible medical decisions. Possible errors in judgment are believed to be the leading cause of delays in medical care or even death.Parkinson's disease is a neurological disorder that worsens with time and is chronic.This disease is amongst the highest causes of neurological illness in older individuals, affecting approximately 1% of the over-60 population. This calls for early detection to assure early treatment as a means of extending patient's lives as fully functional individuals.

10 Product Prototype

- **Data Collection:** Voice recordings, movement patterns (from sensors or wearables), and other biomarkers are collected and fed into the system.
- **Preprocessing:** Raw data undergoes preprocessing steps such as noise reduction, normalization, and feature extraction to prepare it for the AI model.
- **ML Model Integration:**Trained machine learning or deep learning models analyze preprocessed data to recognize patterns indicative of Parkinson's disease.
- **Decision Making:**Based on the model's analysis, the system generates a diagnosis or probability score for Parkinson's disease detection.
- **User Interface:**The results or diagnosis are displayed to the user in an understandable format, possibly indicating the confidence level of the AI's assessment (Fig 1).

11 Code Implementation

11.1 Dataset

In this stage, finding and fixing any problems with the data is a top priority. The acquired data (shown in Figure 2) includes approximately 195 patient records, with each row comprising 23 distinct voice characteristics.

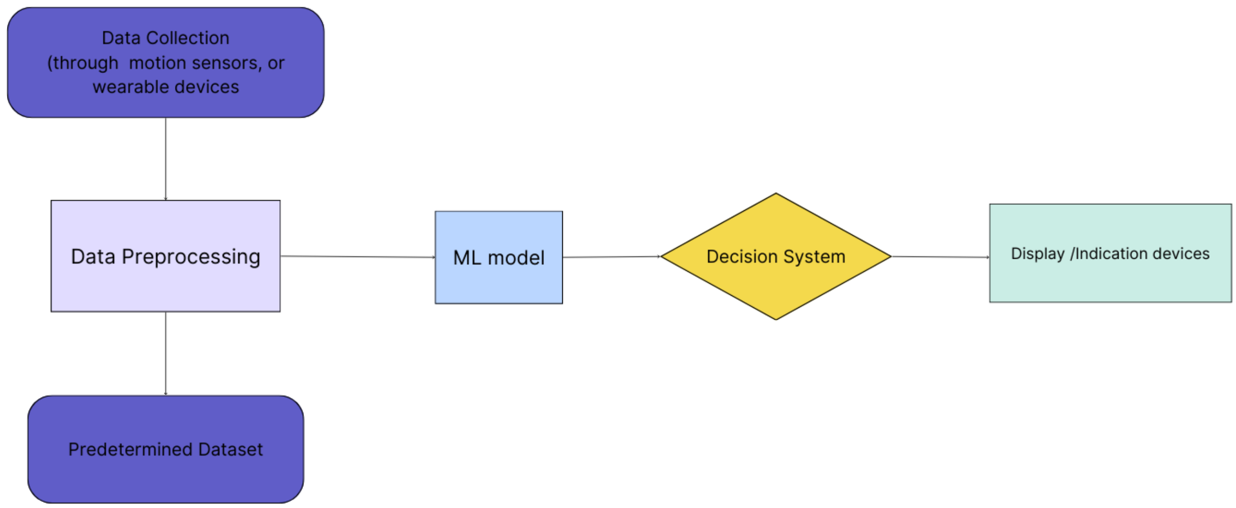

Figure 1: Schematic Diagram for Parkinson Disease Detection using Machine Learning

	A	B	C	D	E	F	G	H
1	name	MDVP:Fo(	MDVP:Fhi(	MDVP:Flo(	MDVP:Jitt(	MDVP:Jitt(	MDVP:RAI	MDVP:PPC
2	phon_R01	119.992	157.302	74.997	0.00784	0.00007	0.0037	0.00554
3	phon_R01	122.4	148.65	113.819	0.00968	0.00008	0.00465	0.00696
4	phon_R01	116.682	131.111	111.555	0.0105	0.00009	0.00544	0.00781
5	phon_R01	116.676	137.871	111.366	0.00997	0.00009	0.00502	0.00698
6	phon_R01	116.014	141.781	110.655	0.01284	0.00011	0.00655	0.00908

Figure 2: Sample Dataset

11.2 Data PreProcessing

In this stage of the process, duplicates or missing values in the dataset are searched for and removed if they are found.3.

11.3 Train Test Split

Shown in Figure 4

11.4 Applying Machine Learning Models

11.4.1 K Nearest Neighbor(KNN)

Shown in Figure 5

```
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   name              195 non-null    object
 1   MDVP:Fo(Hz)       195 non-null    float64
 2   MDVP:Fhi(Hz)      195 non-null    float64
 3   MDVP:Flo(Hz)      195 non-null    float64
 4   MDVP:Jitter(%)    195 non-null    float64
 5   MDVP:Jitter(Abs)  195 non-null    float64
 6   MDVP:RAP          195 non-null    float64
 7   MDVP:PPQ          195 non-null    float64
 8   Jitter:DDP        195 non-null    float64
 9   MDVP:Shimmer      195 non-null    float64
 10  MDVP:Shimmer(dB)  195 non-null    float64
 11  Shimmer:APQ3      195 non-null    float64
 12  Shimmer:APQ5      195 non-null    float64
 13  MDVP:APQ          195 non-null    float64
 14  Shimmer:DDA       195 non-null    float64
 15  NHR               195 non-null    float64
 16  HNR               195 non-null    float64
 17  RPDE              195 non-null    float64
 18  DFA               195 non-null    float64
 19  spread1           195 non-null    float64
 20  spread2           195 non-null    float64
 21  D2                195 non-null    float64
 22  PPE               195 non-null    float64
 23  status            195 non-null    int64
dtypes: float64(22), int64(1), object(1)
```

Figure 3: Import code to Python

```
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.3,random_state=7)
```

```
print(x_train.shape)
print(y_train.shape)
print(x_test.shape)
print(y_test.shape)
```

```
(136, 17)
(136,)
(59, 17)
(59,)
```

Figure 4: Train Test Split

11.4.2 Support Vector Machine

Shown in Figure 6

```
from sklearn.neighbors import KNeighborsClassifier
```

```
model4 = KNeighborsClassifier(n_neighbors = 2, metric = 'minkowski', p = 2)
model4.fit(x_train, y_train)
final_pred4 = model4.predict(x_test)
```

Figure 5: KNN Classifier Model Code

```
from sklearn.svm import SVC
```

```
model5 = SVC(kernel = "rbf", random_state = 0, gamma = 0.1, probability = True)
model5.fit(x_train, y_train)
final_pred = model5.predict(x_test)
```

Figure 6: SVM Classifier Model Code

11.4.3 Decision Tree

Shown in Figure 7

```
from sklearn.tree import DecisionTreeClassifier
```

```
model2 = DecisionTreeClassifier(criterion='entropy',max_depth=6,random_state=100,min_samples_leaf=5
model2.fit(x_train, y_train)
final_pred = model2.predict(x_test)
```

Figure 7: Decision Tree Classifier Model Code

11.4.4 Random Forest

Shown in Figure 8

11.4.5 XGBoost

Shown in Figure 9

11.5 Result Prediction

The results of each of these classifiers in shown in table 1 given below. The highest accuracy is achieved by XGBoost Algorithm. Hence we will use this for our model traning and deployment.

```
from sklearn.ensemble import RandomForestClassifier
```

```
model3 = RandomForestClassifier(n_estimators = 50)
model3.fit(x_train, y_train)
final_pred = model3.predict(x_test)
```

Figure 8: Random Forest Classifier Model Code

```
from xgboost import XGBClassifier
```

```
model1 = XGBClassifier()
model1.fit(x_train, y_train)
final_pred = model1.predict(x_test)
```

Figure 9: XGBoost Classifier Model Code

12 Conclusion

It is essential to develop reliable diagnostic tools in order to cut the chance of developing Parkinson's disease, for which there is no cure. For the purpose of this endeavor, we have utilized a total of five distinct Machine Learning Techniques, namely KNN, SVM, Decision Tree, Random Forest and XGBoost, in order to make a prediction regarding Parkinson's illness. These models are used to train the dataset, and then the models that were generated using the various approaches are compared to one another in order to decide which model fits the dataset the best.

The purpose of utilizing a variety of assessment criteria, including Accuracy, Precision, Recall, and F1-score, is to provide an effective comparison of the various models. We made use of the Speech dataset, which can be found in the UCI Machine Learning Repository

Table 1: Performance Comparison of Different Models

	Accuracy	**Precision**	**Recall**	**F1-Score**
KNN	76.27	90.24	78.72	84.09
SVM	79.66	79.66	100	88.6
Decision Tree	89.83	97.67	89.36	93.3
Random Forest	88.13	91.66	93.61	92.63
XGBoost	94.91	95.83	97.87	96.84

and comprises the speech characteristics of the patients who participated in the study. With an accuracy of 94.91%, XGBoost outperforms the other four machine-learning algorithms. This is further used in the web app for prediction.