**Parkinson’s Disease Diagnosis by Audio and Drawing samples using XGBoost, CNN Models**

Prof. Deepali Deshpande  
*Artificial Intelligence and Data Science Vishwakarma Institute of Technology* Pune, India   
 deepali.deshpande@vit.edu

Harshit Mundhra  
*Artificial Intelligence and Data Science Vishwakarma Institute of Technology* Pune, India   
harshit.mundhra20@vit.eduChetanya Rathi  
*Artificial Intelligence and Data Science Vishwakarma Institute of Technology* Pune, India   
chetanya.rathi20@vit.edu

Pushkar Helge  
*Artificial Intelligence and Data Science Vishwakarma Institute of Technology* Pune, India   
pushkar.helge20@vit.eduHarsh Dhabekar  
*Artificial Intelligence and Data Science Vishwakarma Institute of Technology* Pune, India   
harsh.dhabekar20@vit.edu

Kartik Rajput  
*Artificial Intelligence and Data Science Vishwakarma Institute of Technology* Pune, India   
kartik.rajput20@vit.edu

***Abstract—Parkinson's disease (PD) is a neurodegenerative disorder characterized by the progressive loss of dopaminergic neurons in the substantia nigra region of the brain. This research paper provides an in-depth examination of Parkinson's disease, including its clinical features, etiology, pathophysiology, diagnosis, and treatment strategies. The paper also explores current research advancements and emerging therapies aimed at improving the management and prognosis of this debilitating condition. The comprehensive review consolidates existing knowledge and highlights potential future directions for research and therapeutic interventions in Parkinson's disease.***

***The research paper aims to provide a comprehensive understanding of Parkinson's disease, covering its clinical manifestations, underlying mechanisms, diagnostic approaches, current treatment strategies, emerging therapies, and supportive care options. By consolidating existing knowledge and highlighting recent advancements, this review contributes to the ongoing efforts in improving the lives of individuals affected by Parkinson's disease. It also emphasizes the need for further research to unravel the intricacies of the disease and develop novel therapeutic interventions for better disease management and patient outcomes.***

***Keywords : Machine Learning, XgBoost, CNN, Pytorch, SVM, Decision tree.***

1. **INTRODUCTION**

The primary motor symptoms of Parkinson's disease include tremors, rigidity, bradykinesia (slowness of movement), and postural instability. These symptoms can significantly impair a person's ability to perform daily activities and affect their quality of life. However, Parkinson's disease is not limited to motor symptoms alone. It can also present with various non-motor symptoms, such as cognitive impairment, mood disorders, sleep disturbances, autonomic dysfunction, and sensory abnormalities.

The exact cause of Parkinson's disease remains elusive, but both genetic and environmental factors are believed to play a role. Mutations in certain genes, such as LRRK2, PARKIN, and PINK1, have been identified in familial cases of PD, while environmental factors like exposure to pesticides, heavy metals, and certain toxins have been associated with an increased risk of developing the disease. Additionally, oxidative stress, mitochondrial dysfunction, and the accumulation of abnormal protein aggregates, known as Lewy bodies, are thought to contribute to the neurodegenerative process in PD.

**II. LITERATURE REVIEW**

Parkinson's disease (PD) is a widely researched topic, Swith numerous studies focusing on various aspects of the disease, including its etiology, pathophysiology, diagnosis, and treatment strategies. This literature review aims to provide an overview of key findings from recent research in the field of Parkinson's disease.

Etiology and Risk Factors:

Genetic factors have been extensively studied in PD. Mutations in genes such as SNCA, LRRK2, PARKIN, PINK1, and DJ-1 have been identified in familial forms of PD. Additionally, genome-wide association studies (GWAS) have identified several genetic risk variants associated with sporadic PD, highlighting the complex interplay between genetic and environmental factors in disease development.

Environmental factors, including pesticide exposure, industrial toxins, and heavy metals, have been implicated in the pathogenesis of PD. Epidemiological studies have provided evidence of increased PD risk in individuals exposed to certain pesticides and chemicals. Furthermore, mitochondrial dysfunction and oxidative stress have been proposed as key mechanisms linking environmental toxins to neuronal damage in PD.

Pathophysiology:

The hallmark pathological feature of PD is the presence of Lewy bodies, intracellular aggregates primarily composed of alpha-synuclein, in affected brain regions. Accumulation of alpha-synuclein and its subsequent aggregation is believed to contribute to neurodegeneration and dopaminergic cell death in PD. Additionally, glial activation, neuroinflammation, impaired protein degradation pathways, and mitochondrial dysfunction have been implicated in disease progression.

Biomarkers and Diagnosis:

Efforts to identify reliable biomarkers for early diagnosis and disease monitoring in PD are ongoing. Various imaging techniques, including dopamine transporter imaging (DAT-SPECT) and positron emission tomography (PET), can assess dopaminergic function and help differentiate PD from other parkinsonian syndromes. Cerebrospinal fluid (CSF) biomarkers, such as alpha-synuclein, tau, and DJ-1, are being investigated for their potential diagnostic and prognostic value.

Treatment Strategies:

Pharmacological interventions remain the mainstay of PD management. Levodopa, a precursor of dopamine, provides symptomatic relief and remains the most effective therapy for motor symptoms. Dopamine agonists, MAO-B inhibitors, and COMT inhibitors are also commonly used to enhance dopaminergic transmission. However, long-term levodopa use is associated with motor complications, such as dyskinesias and motor fluctuations.

Surgical options, particularly deep brain stimulation (DBS), have shown efficacy in alleviating motor symptoms and improving quality of life in selected PD patients. DBS involves the implantation of electrodes in specific brain regions to modulate abnormal neuronal activity.

Non-Motor Symptoms and Management:

Non-motor symptoms, including cognitive impairment, depression, sleep disorders, autonomic dysfunction, and pain, significantly impact the lives of individuals with PD. Non-dopaminergic pathways are involved in the manifestation of these symptoms. Managing non-motor symptoms often requires a multidisciplinary approach involving medications, behavioral interventions, and supportive care.

Emerging Therapies:

Advancements in research have led to the exploration of novel therapeutic approaches in PD. Stem cell-based therapies, including the use of induced pluripotent stem cells (iPSCs) and transplantation of dopaminergic neurons, hold promise for restoring lost neuronal function. Gene therapy approaches targeting alpha-synuclein accumulation or enhancing dopaminergic signaling are being investigated in preclinical and clinical studies.

Immunotherapy strategies, such as the development of vaccines or antibodies targeting alpha-synuclein, aim to prevent or slow disease progression by reducing protein aggregation. Additionally, efforts are underway to develop disease-modifying drugs that can target multiple pathways implicated

**III. OBJECTIVE**

The objective of this research is to provide a comprehensive understanding of Parkinson's disease (PD) by examining its clinical features, etiology, pathophysiology, diagnosis, and treatment strategies. The research aims to accomplish the following specific objectives:

Review the clinical manifestations of Parkinson's disease, including motor symptoms (such as tremors, rigidity, bradykinesia, and postural instability) and non-motor symptoms (such as cognitive impairment, mood disorders, sleep disturbances, and autonomic dysfunction).

Explore the etiological factors contributing to Parkinson's disease, including genetic factors, environmental exposures, oxidative stress, mitochondrial dysfunction, and alpha-synuclein aggregation.

Investigate the underlying pathophysiological mechanisms involved in Parkinson's disease, such as neurodegeneration, dopaminergic dysfunction, the role of Lewy bodies, glial activation, inflammation, and neurotransmitter imbalances.

Examine the current diagnostic approaches for Parkinson's disease, including clinical assessments, neuroimaging techniques (such as DAT-SPECT and PET), and the potential use of biomarkers for early detection and differential diagnosis.

Evaluate the existing treatment strategies for Parkinson's disease, including pharmacological interventions (such as levodopa, dopamine agonists, and MAO-B inhibitors), surgical options (such as deep brain stimulation), physical and occupational therapy, and management of non-motor symptoms.

Explore emerging therapies and research advancements in Parkinson's disease, including disease-modifying therapies, stem cell-based therapies, gene therapy, immunotherapy, and strategies targeting alpha-synuclein aggregation.

Discuss lifestyle modifications and supportive care approaches for individuals with Parkinson's disease, including exercise and physical activity, nutrition and diet, psychological support and counseling, and caregiver support.

Evaluate the prognosis and functional impairments associated with Parkinson's disease, as well as potential future directions for research in terms of improving disease management and patient outcomes.

**IV. AlGORITHM**

Data Collection:

Gather a dataset of audio recordings from individuals with and without Parkinson's disease.

Ensure the dataset includes a diverse range of samples, such as different age groups, genders, and disease stages.

Data Preprocessing:

Extract relevant features from the audio data, such as pitch, intensity, and spectral information.

Perform necessary data cleaning steps, including removing noise, normalizing the data, and handling missing values.

Feature Selection:

Apply feature selection techniques, such as correlation analysis or information gain, to identify the most discriminative features.

Select a subset of features that provide the best discriminatory power for differentiating Parkinson's disease from healthy individuals.

Model Training and Evaluation:

Split the dataset into training and testing sets, typically using a 70:30 or 80:20 ratio.

Train the following models using the training data:

XGBoost: Implement the XGBoost algorithm and train the model on the selected features.

Decision Tree: Build a decision tree classifier using the selected features.

Random Forest: Construct an ensemble of decision trees using the selected features.

SVM: Train a support vector machine classifier on the selected features.

Evaluate the performance of each model using appropriate evaluation metrics such as accuracy, precision, recall, and F1 score.

Perform cross-validation to assess the robustness of the models and tune hyperparameters if needed.

Model Comparison and Selection:

Compare the performance of the different models based on their evaluation metrics.

Select the model that achieves the highest accuracy or other relevant metrics as the primary Parkinson's disease detection model.

Model Testing and Deployment:

Use the selected model to predict Parkinson's disease in unseen audio samples from the testing set.

Evaluate the model's performance on the testing set and assess its generalization capability.

Once satisfied with the model's performance, deploy it for real-world applications, considering factors such as latency, scalability, and user-friendliness.

Ongoing Monitoring and Model Updates:

Monitor the model's performance in real-world scenarios and collect user feedback.

Continuously evaluate the model's accuracy and consider retraining or updating the model periodically to account for changes in data distribution or advancements in techniques.

It's important to consult with domain experts and follow ethical guidelines when conducting research or developing models for medical diagnosis purposes.

**V. FLOWCHART**

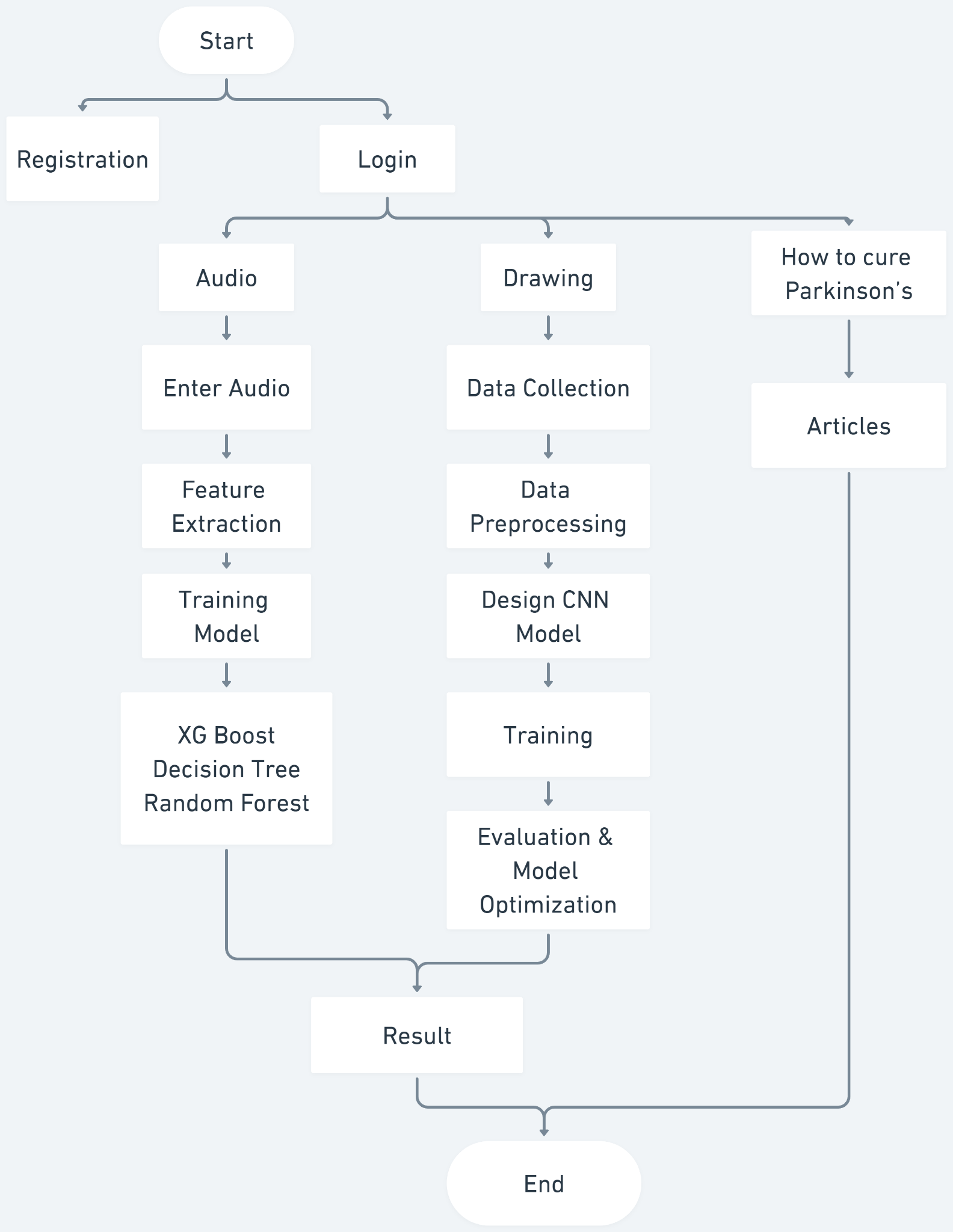
****

Fig 1a. Flowchart

**VI. RESULTS AND DISCUSSIONS**

For the audio dataset, the SVM classifier achieved an accuracy of 86.54% on the training data and 87.18% on the test data. The XGBoost classifier showed improved performance with 100% accuracy on the training data and 92.31% accuracy on the test data. The Decision Tree classifier achieved 100% accuracy on the training data but relatively lower accuracy of 76.92% on the test data. The Random Forest classifier obtained 100% accuracy on the training data and 92.31% accuracy on the test data. Lastly, the Logistic Regression classifier achieved an accuracy of 86.54% on the training data and 84.62% on the test data.

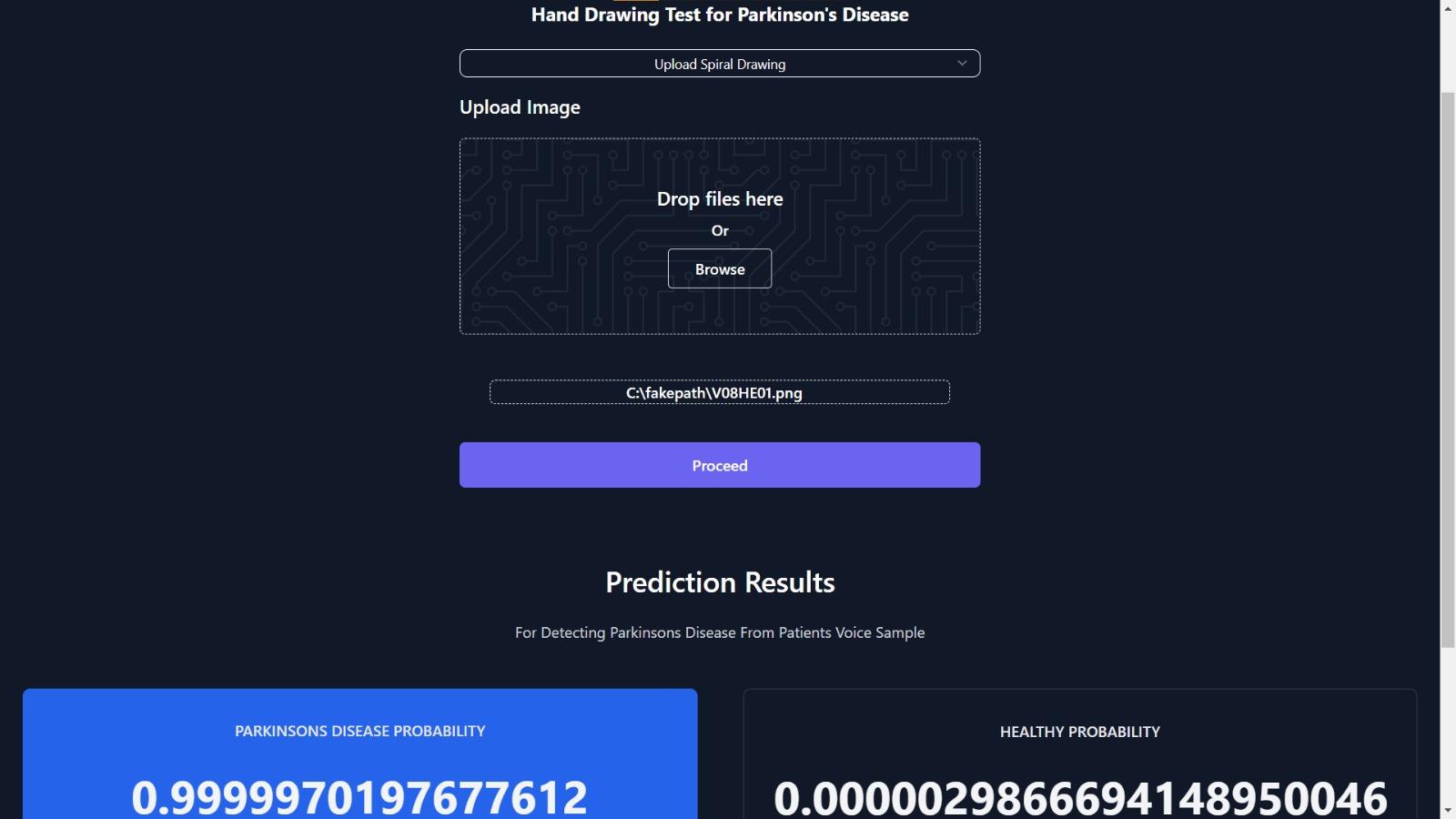


Fig 1b. Hand Drawing sample testing

For the hand drawing image dataset, the CNN model achieved an accuracy of 92%, indicating its effectiveness in classifying Parkinson's disease from hand-drawn images. Additionally, utilizing the PyTorch Lightning framework improved the accuracy to 97%, further demonstrating the efficacy of the approach.

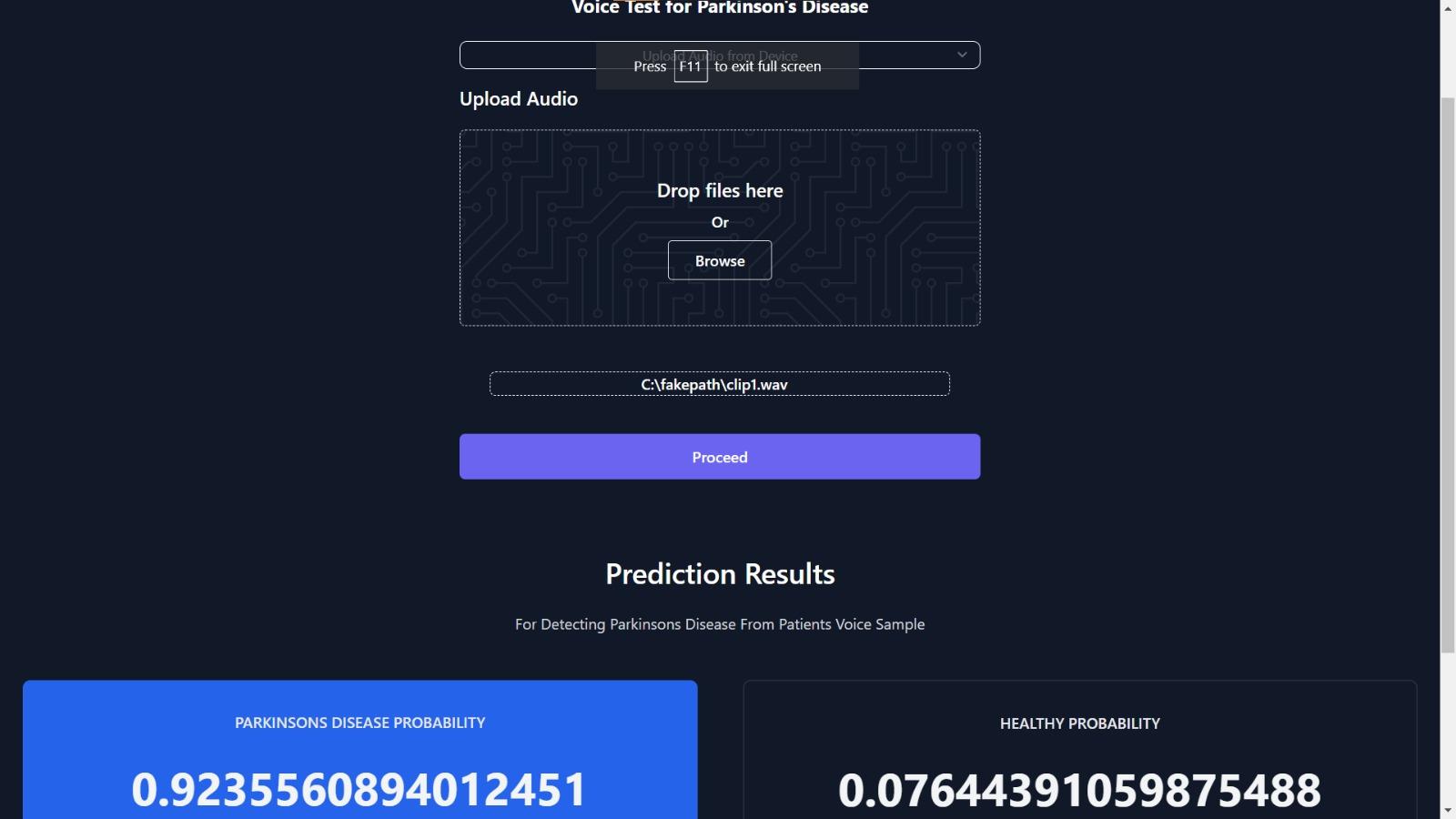


Fig 1c. Audio sample testing

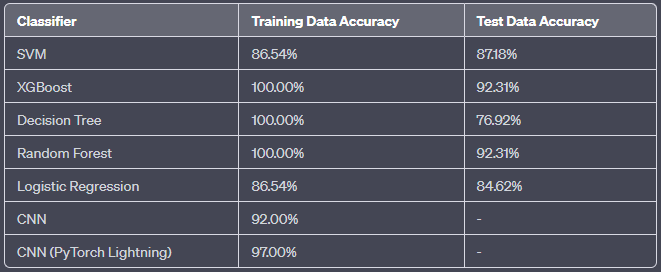
****

Fig 1d. Model Accuracy

**VII. CONCLUSION**

Machine learning algorithms have demonstrated high accuracy in distinguishing PD patients from healthy individuals or differentiating PD from other movement disorders. By analyzing various data sources, such as clinical assessments, neuroimaging data, and biomarkers, machine learning models can identify patterns and features that are indicative of Parkinson's disease. These models can aid clinicians in early and accurate diagnosis, potentially enabling timely intervention and improved patient outcomes.

Moreover, machine learning algorithms can predict the progression and severity of PD by integrating longitudinal data and clinical variables. These predictive models can help identify individuals who are at higher risk of developing motor complications or other adverse outcomes, allowing for proactive management and personalized treatment strategies.

Machine learning has emerged as a valuable tool in Parkinson's disease research, providing new avenues for diagnosis, prediction, biomarker discovery, treatment optimization, and uncovering disease mechanisms. By harnessing the power of machine learning, researchers and clinicians can improve patient care, advance our understanding of Parkinson's disease, and ultimately contribute to the development of more effective therapeutic strategies for this complex neurodegenerative disorder.

**VIII. FUTURE SCOPE**

Machine learning holds significant promise for the future of Parkinson's disease (PD) research and clinical management. The field of machine learning is continually evolving, and there are several areas where it can contribute to further advancements in understanding and treating PD. The future scope of Parkinson's disease in machine learning includes the following aspects:

Enhanced Diagnostic Accuracy:

Machine learning algorithms can continue to improve the accuracy and efficiency of PD diagnosis. By incorporating more diverse data sources, such as wearable sensors, smartphone apps, and voice analysis,

Personalized Treatment Approaches:

Machine learning can play a crucial role in developing personalized treatment strategies for individuals with PD.

Longitudinal Disease Progression Modeling:

Machine learning techniques can be employed to develop robust models for predicting disease progression trajectories in PD. By integrating longitudinal data from diverse sources, including clinical assessments, biomarkers, and digital health data.

Identification of Novel Biomarkers:

Machine learning can help identify novel biomarkers for PD that can aid in diagnosis, monitoring disease progression, and evaluating treatment response. By mining large-scale datasets and integrating various data modalities, including imaging, genetic, proteomic, and metabolomic data.

Real-time Monitoring and Management:

Machine learning combined with wearable sensors and continuous monitoring devices can enable real-time monitoring of PD symptoms and motor fluctuations.

**IX. REFERENCES**

1. Zhao, Wenbing, et al. "A human-centered activity tracking system: Toward a healthier workplace." *IEEE Transactions on Human-Machine Systems* 47.3 (2016): 343-355.
2. Stiefmeier, Thomas, et al. "Event-based activity tracking in work environments." *3rd International Forum on Applied Wearable Computing 2006*. VDE, 2006.
3. Kent, Karen Ann, and Murugiah Souppaya. "Guide to Computer Security Log Management:." (2006).
4. D. Ganapathy and V. K. S. Iyengar. "A Survey of Computer Activity Monitoring Tools" (2010)
5. A. R. Karim and M. A. Hasan. "Computer Activity Monitoring in the Workplace: An Overview"(2011).
6. M. E. S. El-Bendary, A. A. El-Sayed, and H. A. El-Bendary. "Design and Implementation of a Computer Activity Monitoring System" (2009).
7. J. K. Kim, J. H. Kim, and Y. K. Kim. "Monitoring and Analyzing User Activities on a Computer System" (2010).
8. M. J. Froomkin. "Surveillance and Privacy in the Digital Age: A Review of Computer Activity Monitoring Practices" (2000).
9. Krisha, Chatrati Sai, Naidu Sumanth, and C. Raghava Prasad. "RFID based student monitoring and attendance tracking system." *2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT)*. IEEE, 2013.
10. May, Madeth, Sébastien George, and Patrick Prévôt. "A closer look at tracking human and computer interactions in web‐based communications." *Interactive Technology and Smart Education* (2008).