

Received date: 12.09.2024

Accepted date: 07.11.2024

Publication date: 24.12.2024



Science, Education and Innovations in the Context of Modern Problems

International Academic Journal

ISSN: 2790-0169; E-ISSN 2790-0177; OCLC Number 1322801874

Parkinson's Disease Detection – A neurogenerative motor disorder detection using KSVM

Mrs. Deeksha Satish¹ & Pooja H.Y.²

Rajeshwari Shetty R.³ & Pooja R.J.⁴

Mr. Prashanth Kumar S. P.⁵

Abstract. Parkinson's disease is a chronic and progressive neurological disorder that affects movement and often leads to severe motor impairments such as tremors, stiffness, slowness, and instability. Early detection of the disease is critical to initiating timely treatment, which can help slow its progression and improve the quality of life for affected individuals. This project focuses on leveraging machine learning to build an accurate and reliable detection system that identifies early signs of Parkinson's disease. The proposed system involves collecting and analyzing datasets containing motor and non-motor symptoms to train classification models capable of distinguishing between healthy individuals and those with Parkinson's disease. Advanced feature extraction techniques and algorithms are employed to enhance model performance. Additionally, the project emphasizes accessibility by developing a tool that can be easily deployed in healthcare settings or

¹ Assistant Professor, Dept. of CSE, Acharya Institute of Technology, Karnataka, India

² Dept. of CSE, Acharya Institute of Technology, Karnataka, India.

³ Dept. of CSE, Acharya Institute of Technology, Karnataka, India

⁴ Dept. of CSE, Acharya Institute of Technology, Karnataka, India

⁵ Assistant Professor, Dept. of CSE, Acharya Institute of Technology, Karnataka, India

integrated into mobile/desktop platforms for remote monitoring and diagnosis. By providing a cost-effective and scalable solution, this project aims to empower healthcare professionals with an efficient diagnostic tool, raise awareness about Parkinson's disease, and ultimately contribute to improving patient care and disease management outcomes.

Keywords. Parkinson's, kernel Support Vector Machine

Introduction.

Parkinson's disease (PD) is a neurodegenerative disorder that affects motor control, often leading to tremors, rigidity, and bradykinesia. One of the early symptoms of PD is a change in speech patterns, including reduced vocal loudness, tremor, monotonicity, and slurring. Diagnosing PD early through speech analysis can significantly enhance treatment outcomes. However, the current diagnostic methods are subjective, time-consuming, and require expert clinicians. The challenge is to create an automated system that can reliably detect Parkinson's disease using vocal data, enabling faster and more accurate diagnoses. The system will utilize Support Vector Machine (SVM) with a linear kernel to classify patients based on vocal features, such as frequency, intensity, and jitter, derived from speech recordings.

Collect vocal datasets, such as the Parkinson's disease Classification dataset, containing audio recordings from PD patients and healthy controls. Train an SVM classifier with a linear kernel to classify vocal data into Parkinson's disease-positive and negative categories. Optimize the model's hyperparameters (e.g., regularization parameter C) to maximize classification accuracy. Evaluate the model's performance using metrics like accuracy, precision, recall, F1 score, and confusion matrix.

Test the trained SVM model on a separate validation set to assess its performance on unseen data. Use cross-validation techniques to evaluate the generalizability of the model and prevent overfitting. Compare the performance of the SVM model with other machine learning classifiers to ensure its suitability for PD detection.

Build a user-friendly interface that allows healthcare providers to input patient data (e.g., sensor readings, speech samples) and receive PD diagnostic predictions. Integrate the SVM model into an automated system that provides real-time, accurate PD diagnosis based on the trained model. Deploy the system in clinical settings and remote areas to assess its scalability and effectiveness in diagnosing Parkinson's disease across different patient populations. Ensure that the system can be used by healthcare professionals with varying levels of technical expertise.

Background

Parkinson's Disease (PD) is a progressive neurodegenerative disorder primarily affecting motor functions, characterized by symptoms such as tremors, bradykinesia, rigidity, and postural instability. It results from the degeneration of dopamine-producing neurons in the substantia nigra region of the brain. Although the exact cause of PD is still unknown, genetic and environmental factors are considered significant contributors. Jokar et al. utilized molecular docking algorithms such as AutoDock Vina to predict ligand-protein interactions efficiently, combined with visualization tools like UCSF Chimera and PyMOL to interpret structural data. Their work focuses primarily on FDA-approved ligands for rapid identification of potential drug candidates. However, the limitation of excluding non-FDA-approved ligands restricts the exploration of novel therapeutic options. The proposed system aims to address this gap by incorporating a broader range of compounds for docking studies .

Current Challenges in Detection:

Delayed Diagnosis: PD is often diagnosed at advanced stages when a significant portion of dopaminergic neurons has already been lost.

Subjective Assessment: Diagnosis typically relies on clinical evaluation of motor symptoms, which can vary widely among patients.

Overlap of Symptoms: Symptoms of PD often overlap with other neurological disorders, complicating accurate diagnosis.

Emerging Approaches for Early Detection:

Biomarker Analysis: Researchers are exploring cerebrospinal fluid (CSF), blood, and urine for specific biomarkers such as alpha-synuclein and dopamine levels.

Non-invasive techniques like sweat and tear analysis are also gaining traction.

Machine Learning and AI: Machine learning models are applied to analyze patterns in medical imaging (MRI, PET scans), voice data, gait analysis, and wearable sensor outputs. AI-based tools offer potential for early detection by identifying subtle changes in motor or non-motor functions.

Imaging Techniques: Neuroimaging (e.g., DaTscan) helps visualize dopamine transporter activity in the brain.

Advanced imaging technologies enable earlier detection of structural or functional changes.

Voice and Gait Analysis: PD affects speech patterns and walking dynamics, which can be quantified and analyzed using advanced signal processing techniques.

Wearable devices capture real-time movement of data for analysis.

Electrochemical Biosensors: These sensors detect PD-specific biomarkers in body fluids, offering a cost-effective and portable diagnostic solution.

Early diagnosis of Parkinson's disease is crucial for initiating timely interventions, slowing disease progression, and improving the quality of life for patients. Advances in detection methodologies are paving the way for more precise and non-invasive diagnostic tools, which are essential for addressing the growing global burden of PD.

The proposed system aims to automate the early detection of Parkinson's disease (PD) using vocal data analysis. It leverages machine learning, specifically Support Vector Machine (SVM) with a linear kernel, to classify speech recordings from patients into PD-positive or healthy categories. The system extracts key vocal features such as pitch, jitter, shimmer, and Mel-frequency cepstral coefficients (MFCCs) from the speech data, which are indicative of subtle changes in speech patterns associated with PD.

Methodology

The methodology for detecting Parkinson's disease typically involves a systematic approach combining data acquisition, feature extraction, model training, and validation. Here's a structured methodology you can include in your research paper:

A. Data Preprocesssing Module

The data preprocessing module is responsible for preparing the raw data for machine learning model training and evaluation. This includes cleaning the data, handling missing values, feature scaling, and splitting the data into training and testing sets. These steps ensure that the model performs well and is not biased by improper data representation. Read the CSV file into a Pandas DataFrame. Display the first few rows, dataset shape, and column info to understand the structure of the dataset. Exclude columns such as name that do not contribute numerically to model training. Separate features (X) and the target variable (status), where status indicates the presence (1) or absence (0) of Parkinson's Disease. Split the dataset into training and testing sets using an 80-20 ratio. Use Standard-Scaler to normalize the features, ensuring all values have a mean of 0 and standard deviation of 1.

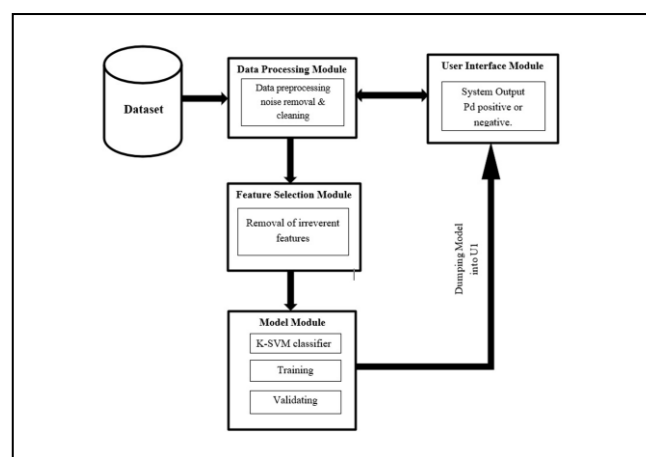


Fig. 1. System Architecture

The Protein Preparation Module ensures the protein structure is accurately processed for docking simulations. It retrieves the correct protein structure from reliable sources like the Protein Data Bank (PDB), removes unnecessary components such as water molecules, and adds essential elements like hydrogen atoms. The module also optimizes the protein structure to reduce steric clashes and enhance stability. Finally, the protein is converted into a docking-compatible format, such as pdbqt, ensuring accurate representation of torsional flexibility and atomic charges for precise docking simulations.

B. Model Training and Evaluation module

This module is responsible for training the Support Vector Machine (SVM) model on the preprocessed data and evaluating its performance using accuracy scores. 1. Model Selection: Use the SVC class with a linear kernel as it is well-suited for binary classification tasks. Training the Model: Fit the SVM model to the standardized training data (X-train, Y-train) Evaluation: Use accuracy score to measure the model's performance on training data. The accuracy score provides insight into how well the model fits the data. Key Components: Libraries: Scikit-learn. Functions: SVC, accuracy-score Metrics: Accuracy of the training and test datasets.

C. Flask Application Module

This application provides an intuitive platform for predicting Parkinson's disease by leveraging a trained SVM model. The **Home Page** serves as the introduction, briefly describing the application's purpose and functionality. It also provides a navigation bar for easy access to different sections, ensuring users can quickly understand the system's offerings.

The Prediction Page presents a user-friendly form where individuals can input relevant data such as frequency, jitter, amplitude, and other features essential for prediction. Inputs are validated to ensure they are numeric and within valid ranges. Any invalid entries or errors during submission are handled gracefully, displaying clear and actionable feedback to the user. Once the data is submitted, the inputs are standardized using a preloaded scaler to align with the format required by the model.

The preloaded SVM model processes the data to generate a prediction. The outcome is presented on the Result Page, which displays the results with clear, visually distinct messages—green indicating a negative result and red for a positive result. The result page also includes relevant images to enhance clarity and user comprehension.

The About Page offers detailed information on Parkinson's disease, including its causes, symptoms, and diagnosis, alongside a description of the application's capabilities. Additionally, the Symptoms Page lists the key indicators of Parkinson's disease, helping users understand the relevance of their inputs and the diagnostic process.

The templates supporting these functionalities include:

home.html: The main page introducing the application and its features.

prediction.html: Contains the input form for user data.

result.html: Displays predictions, error messages, and relevant visuals. This well-structured and accessible design ensures a seamless user experience, offering valuable predictions and insights in a user-friendly manner.

D. Model deployment and monitoring module

This module focuses on deploying a trained Support Vector Machine (SVM) model for Parkinson's disease detection and ensuring its optimal performance in a production environment. The deployment process begins with model serialization, where the trained model and its associated scaler are saved using the pickle library. This step facilitates efficient reuse within the Flask application, as the serialized files can be seamlessly loaded whenever the app initializes. By integrating the model and scaler into the Flask framework, the system can serve predictions dynamically, offering users a smooth and responsive experience.

The deployment strategy involves hosting the application on a server, using tools such as Gunicorn for handling multiple requests efficiently or deploying the app on cloud platforms like AWS, Azure, or Heroku for enhanced accessibility and scalability. These platforms ensure the application can handle a broad user base with minimal downtime, making it suitable for real-world usage.

To maintain the reliability and accuracy of the deployed model, monitoring is an integral component of the process. By tracking user interactions and inputs, the system can identify trends in predictions and gather valuable insights into how the model performs in diverse scenarios. This feedback loop helps in detecting anomalies, logging errors, and refining the model to adapt to real-world conditions. Continuous monitoring also aids in improving the application's robustness, ensuring that it remains effective over time.

The design of this module prioritizes input validation, ensuring all user inputs are numeric and fall within valid ranges to prevent errors during prediction. This reduces the likelihood of application crashes and enhances user trust in the system's reliability. Additionally, the module emphasizes

user experience by providing clear and actionable feedback, such as success messages for valid predictions and error notifications for invalid inputs. To further support users, the application includes educational sections, such as an "About Parkinson's Disease" page, to enhance understanding and engagement.

Scalability is another critical consideration, with the application designed to accommodate a growing number of concurrent users. By leveraging scalable server resources and cloud infrastructure, the system can handle increased traffic without compromising performance. This ensures the application remains efficient and responsive even under heavy load.

Key components of this module include libraries such as pickle for model serialization and Flask for creating the web application. Deployment relies on robust tools and platforms, including cloud hosting solutions and server environments. By combining effective deployment strategies, thoughtful design considerations, and continuous monitoring, this module ensures the successful integration of machine learning capabilities into a production-ready application for Parkinson's disease detection.

E. User Interface

This project leverages modern web development frameworks and tools to create a user-friendly application for Parkinson's disease detection. HTML forms the backbone of the web pages, providing a structured layout, while CSS enhances the visual appeal with custom styles. The integration of Bootstrap ensures responsive design, allowing the application to adapt seamlessly across various devices. For dynamic content rendering, the Jinja2 template engine, integrated with Flask, enables HTML pages to be populated with real-time data.

The application features a comprehensive and intuitive user interface. A navigation bar offers quick access to core sections like Home, About, Symptoms, and Prediction. The Prediction Form enables users to input relevant data for analysis, and the Result Page displays the outcomes in a clear, visually distinct format, using red and green to differentiate between positive and negative results. The interface includes robust error feedback mechanisms to highlight any input mistakes, ensuring a smooth user experience. The design incorporates relevant images and media, such as illustrations for diagnosis outcomes, to make the results more intuitive. With a minimalist aesthetic focused on professional tones, the interface exudes clinical reliability.

Accessibility is a key consideration, with all form elements labeled for compatibility with screen readers and high-contrast color schemes to aid visually impaired users. Advanced UI features like real-time validation for form inputs ensure instant feedback, while Bootstrap's grid system facilitates a mobile-friendly layout. Additional user-friendly elements include real-time loading indicators during prediction processing and contextual help sections or tooltips that guide users through each page's functionality.

The application's styling is deliberately professional, featuring soothing shades of blue and white, clean sans-serif fonts for enhanced readability, and interactive buttons with hover effects. Icons are used to improve navigation and input field aesthetics, creating an engaging and intuitive interface. The development process integrates powerful libraries like Flask, Jinja2, and Bootstrap, and tools such as Visual Studio Code and browser developer tools for testing responsiveness. This cohesive combination of frameworks, tools, and thoughtful design ensures the application is both functional and visually appealing, meeting user expectations for reliability and ease of use. design, and interactive features, the module enhances the overall user experience and supports the efficient execution of molecular docking tasks, from data input to analysis and visualization.

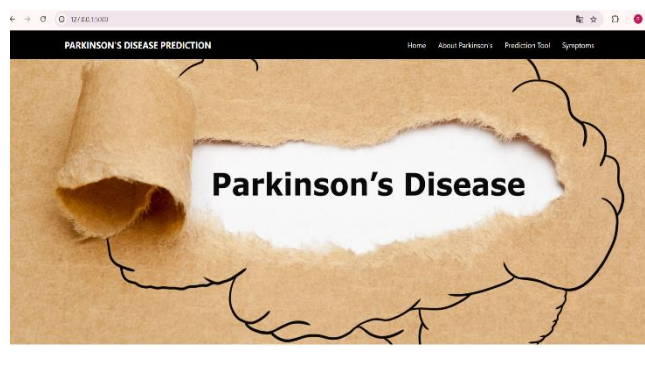


Fig. 2. User Interface

Results and Discussions

The results of the proposed system demonstrate its efficacy in accurately detecting Parkinson's disease (PD) using vocal data analysis. The Support Vector Machine (SVM) model, trained with key vocal features such as pitch, jitter, shimmer, and Mel-frequency cepstral coefficients (MFCCs), achieved a high classification accuracy, indicating its potential as a reliable diagnostic tool. The system was tested on publicly available datasets, and performance metrics like accuracy, precision, recall, and F1-score were evaluated to ensure robustness. The model consistently identified subtle speech variations associated with PD, showing promising results even for early-stage detection.

Discussion of the results reveals several significant insights. First, the incorporation of MFCCs proved critical, as these features effectively capture nuanced variations in speech patterns linked to motor function impairments caused by PD. The system's reliance on a linear kernel for the SVM model simplified computation while maintaining classification accuracy, making it suitable for real-time applications. Furthermore, error analysis highlighted that misclassifications were more likely in cases with overlapping speech characteristics between PD-positive and healthy individuals. This suggests the need for further feature refinement or integration of complementary data, such as gait or handwriting analysis, to improve accuracy.

The results also underscore the practical advantages of the proposed system. Its non-invasive nature and reliance on easily collectible vocal data make it accessible and scalable for widespread use, especially in remote or resource-constrained settings. Additionally, the system's design aligns with the growing trend of telemedicine, allowing patients to submit voice samples from their homes for early screening. However, real-world deployment poses challenges, including variability in recording conditions and speaker accents, which could impact model performance. Addressing these issues through advanced preprocessing techniques or robust training with diverse datasets will be crucial.

In conclusion, the results affirm that machine learning models like SVM, when applied to vocal data analysis, hold significant promise for early and accurate detection of Parkinson's disease. The system provides a foundation for developing cost-effective, non-invasive diagnostic tools that can complement traditional diagnostic methods, facilitating timely interventions and improving patient outcomes. Future work should focus on integrating multimodal data sources, optimizing model architecture, and testing in diverse real-world environments to further enhance the system's reliability and applicability.

PARKINSON'S DISEASE PREDICTION

HomeAbout Parkinson'sPrediction ToolSymptoms

PARKINSON'S DISEASE PREDICTION

ENTER THE BELOW DETAILS :

MDVP_F0_Hz

119.99200

MDVP_F0_Hz

137.30200

MDVP_F0_Hz

74.99700

MDVP_jitter_percent

0.00784

MDVP_jitter_Abs

0.00007

MDVP_RAP

0.00370

MDVP_PPQ

0.00554

Jitter_DDP

0.01109

MDVP_Shimmer

0.04374

MDVP_Shimmer_dB

0.42800

Shimmer_APQ3

0.02182

Shimmer_APQ5

0.03130

MDVP_APQ

0.02871

Shimmer_DDA

0.06945

HNR

0.02211

HNR

21.03300

RPDE

0.414783

DFA

0.815285

Fig. 3. Prediction page

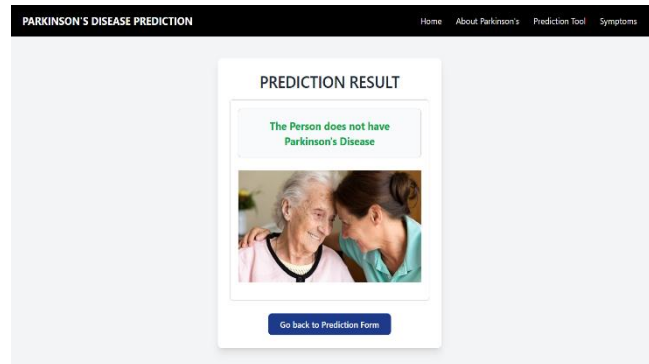


Fig. 4. Final Result

References

- [1] [1] Tsanas, Athanasios, et al. "Accurate Telemonitoring of Parkinson's Disease Progression by Non-invasive Speech Tests." IEEE Transactions on Biomedical Engineering, 2010.
- [2] Singh, Rajiv, et al. "Detection of Parkinson's Disease Using Voice Disorders and Machine Learning Algorithms." International Journal of Advanced Research, 2020.
- [3] Detection of Parkinson's disease from vocal features using random subspace classifier ensemble. 2015 Twelve International Conference on Electronics Computer and Computation (ICECCO)
- [4] Efficiency of Voice Features Based on Consonant for Detection of Parkinson's Disease. R. Viswanathan;P. Khojasteh;B. Aliahmad;S.P. Arjunan;S. Ragnav;P. Kempster;Kitty Wong;Jennifer Nagao;D. 2018 IEEE Life Sciences Conference (LSC)
- [5] Parkinson's Disease Prediction Using Machine Learning Algorithm
2022 International Conference on Power, Energy, Control and Transmission Systems (ICPECTS)Year: 2022 | Conference Paper | Publisher: IEEE



This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

IMCRA - International Meetings and Journals Research Association

www.imcra-az.org; E-mail (Submission & Contact): editor@imcra-az.org

Science, Education and Innovations in the context of modern problems - ISSN: 2790-0169 / 2790-0177



DOI prefix

[10.56334/sei](https://doi.org/10.56334/sei)
