IJARSCT ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Impact Factor: 7.53 Volume 4, Issue 2, August 2024

Parkinson's Disease Detection

Shahista Sultana, Sirimani Uma Maheshwari, Om Raju Bejankiwar

Students, Department of CSE

Sreenidhi Institute of Science and Technology, Hyderabad, India

Abstract: *Parkinson's complaint(PD) is a neurodegenerative complaint characterized by progressive loss of motor function, including temblors, severity, and brady kinesia. Beforehand discovery and monitoring of PD are pivotal for timely intervention and better operation of the complaint. This provides a comprehensive review of colorful ways and methodologies employed in the discovery and monitoring of Parkinson's complaint. The datasets frequently include a combination of demographic information(age, gender), clinical assessments, and biomedical signals(accelerometer data, voice recordings, gait analysis). This design presents a comprehensive review of ML- grounded approaches for Parkinson's complaint discovery. latterly, it provides an overview of the different types of data employed in ML models for PD discovery, including clinical assessments, inheritable data, neuro imaging, and detector- grounded data. likewise, it examines the integration of multimodal data emulsion ways, where information from multiple sources is combined to enhance individual delicacy. It also explores the challenges associated with ML- grounded PD discovery, including data failure, model interpretability, and conception to different case populations. The former algorithms like decision trees give70- 80 of delicacy. The algorithm in this design will be giving delicacy up to 90. Eventually, the design highlight sunborn directions and arising trends in ML- grounded PD discovery*

Keywords: Machine literacy, Support Vector Bracket, hyperactive parameter tuning, Kernel

I. INTRODUCTION

1.1 PROJECT INTRODUCTION

The application of (ML) in Parkinson represents a significant advancement in medical diagnostics. ML algorithms can analyse vast amounts of data from various sources, including medical records, genetic information, and neuroimaging scans, to identify patterns and biomarkers associated with Parkinson's disease. By leveraging these algorithms, researchers and clinicians can achieve earlier and more accurate diagnoses. For instance, machine learning models can process data from wearable devices that monitor patients' movements, detecting subtle motor changes that may indicate the onset of Parkinson's before clinical symptoms become apparent. Additionally, ML can assist in distinguishing Parkinson's from other neurodegenerative disorders with similar symptoms, thereby improving diagnostic accuracy. The integration of machine learning in Parkinson's detection not only enhances early diagnosis but also aids in monitoring disease progression and tailoring personalized treatment plans, ultimately improving patient outcomes and quality of life.

1.2 SCOPE

The scope of Parkinson's disease detection using machine learning techniques encompasses leveraging diverse biomedical data sources, including voice recordings, clinical assessments, and imaging data, to develop robust algorithms. These algorithms aim to accurately differentiate individuals with Parkinson's disease from healthy individuals, enabling early detection and timely intervention. By implementing scalable solutions, such as automated diagnostic tools, the scope extends to improving diagnostic accuracy and monitoring disease progression over time. This approach not only enhances clinical decision-making but also facilitates the development of personalized treatment strategies tailored to individual patient needs, thereby potentially improving overall patient outcomes and quality of life the implementation of real-time monitoring systems that can continuously assess disease severity and treatment efficacy, thereby enabling personalized care plans and interventions tailored to individual patient needs. Ultimately, the integration of machine learning in Parkinson's disease detection aims to enhance diagnostic accuracy, improve clinical

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 2, August 2024

decision-making, and ultimately contribute to better management strategies and outcomes for patients affected by this neurodegenerative

1.3 PROJECT OVERVIEW

The objective of this project is to build machine learning models capable of accurately diagnosing Parkinson's disease by analyzing relevant biomedical data, aiming to distinguish between individuals with the disease and those without, thereby facilitating early and reliable detection for effective treatment and management strategies.Early detection is critical for managing the disease effectively, but diagnosis can be challenging due to the variability of symptoms. Machine learning offers a promising solution by analysing patterns in various types of data, such as voice recordings, handwriting samples, and gait analysis. By identifying subtle changes associated with Parkinson's, machine learning models can assist healthcare professionals in diagnosing the disease earlier and more accurately.

1.4 OBJECTIVE

The primary objective of using (ML) for Parkinson is to develop an automated, accurate, and efficient system that can identify the presence and progression of Parkinson's disease at an early stage. This system aims to assist healthcare professionals in diagnosing the disease by analysing patient data, such as voice recordings, handwriting samples, gait analysis, and other biomarkers, using advanced ML algorithms.

II. LITERATURE SURVEY

2.1 EXISTING SYSTEM

Machine learning models for detecting Parkinson's disease (PD) utilize diverse data types, including voice recordings, handwriting samples, gait analysis, and neuroimaging. For voice analysis, Support Vector Machines (SVM) and Random Forests are commonly employed to classify speech patterns indicative of PD. Handwriting analysis often uses Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) to detect motor impairments. Gait and movement analysis leverage wearable sensors with models like SVMs, Random Forests, and CNNs, as well as time-series models such as Long Short-Term Memory (LSTM) networks. In neuroimaging, deep learning models, particularly CNNs, are used to analyse MRI and PET scans for structural and functional brain changes associated with PD. Additionally, clinical data from Electronic Health Records (EHRs) are analysed using gradient boosting machines and logistic regression to identify PD risk factors. Combining these data sources through multi-modal data integration creates robust predictive models using ensemble methods or multi-modal neural networks. Notable examples include models built using the Parkinson's Progression Markers Initiative (PPMI) dataset, which incorporates clinical, imaging, and biological markers to predict PD progression and diagnosis with various machine learning algorithms

2.2 PROPOSED SYSTEM

Our proposed model for Parkinson's disease (PD) detection leverages Support Vector Classification (SVC) as the core machine learning technique. The model integrates diverse data sources, including voice recordings, handwriting samples, gait analysis, and neuroimaging data, to create a comprehensive diagnostic tool. Voice recordings are analysed to detect subtle variations in speech that are characteristic of PD, with SVC employed to classify these features effectively. Handwriting samples, captured through digitized devices, are processed to extract temporal and spatial features, which are then input into the SVC for classification. The integration of these heterogeneous data types aims to enhance diagnostic accuracy by capturing a wide range of PD symptoms and markers. SVC's ability to handle high-dimensional data and provide robust classification performance makes it particularly well-suited for this task. Our model employs a multi-step feature extraction and selection process to ensure that the most relevant information from each data source is utilized, thereby optimizing the SVC's effectiveness.

Furthermore, the model incorporates advanced techniques like cross-validation and hyperparameter tuning to ensure optimal performance and generalizability across different patient populations. By combining these various data types, our model not only improves early detection of PD but also facilitates continuous monitoring of disease progression, offering a valuable tool for both clinicians and researchers in managing and studying Parkinson's disease.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 2, August 2024

This integrative approach provides a holistic view of the disease, potentially leading to more personalized treatment strategies and better patient outcomes. To implement Support Vector Classification (SVC) for Parkinson's disease detection, we employed a comprehensive pipeline that includes Grid Search Cross-Validation (Grid Search CV) and Recursive Feature Elimination (RFE) to optimize model performance and feature selection. Initially, we preprocessed the diverse data sources, including voice recordings, handwriting samples, gait analysis, and neuroimaging data, extracting relevant features from each modality. These features were then standardized to ensure uniformity across the dataset. We utilized RFE to identify and select the most significant features for PD detection, systematically eliminating less important features and retaining those that contribute most to the model's predictive power. This step helped in reducing dimensionality and improving the model's interpretability and performance. For model optimization, we employed Grid Search CV to tune the hyperparameters of the SVC. This involved defining a parameter grid for key hyperparameters such as the penalty parameter (C), kernel type, and gamma. Grid Search CV systematically evaluated all possible combinations of these hyperparameters through cross-validation, ensuring the model was trained and validated on different subsets of the data to prevent overfitting and to find the optimal parameter settings.

The combined use of RFE and Grid Search CV resulted in a well-tuned SVC model that efficiently handled the high-dimensional data, maximizing diagnostic accuracy and robustness. This pipeline ensured that the final model was both efficient and effective in detecting Parkinson's disease, leveraging the strengths of SVC in handling complex, multi-modal datasets.

III. SYSTEM ANALYSIS

3.1 FUNCTIONAL REQUIREMENTS

Creating functional requirements for a Parkinson's disease detection system involves specifying what the system must do to meet its objectives.

Data Collection Requirements

- Sensor Integration: The system should integrate with wearable sensors (e.g., accelerometers, gyroscopes) to collect motor function data.
- Voice Analysis: The system should capture and analyze voice data for speech impairments, a common symptom of Parkinson's.

Data Analysis and Detection

- Machine Learning Algorithms: The system should use machine learning model SVC trained on patient data to detect patterns indicative of Parkinson's disease.
- Feature Extraction: The system should extract relevant features from the data (e.g., tremor amplitude, speech pauses) for analysis.
- Electronic Health Record (EHR) Integration: The system should integrate with existing EHR systems to pull in historical data and to update records with new findings.
- Data Export: The system should allow the export of data in standard formats (e.g., CSV, HL7) for use in other systems or for further analysis.

Security and Privacy Requirements

- Data Encryption: All patient data should be encrypted both at rest and in transit.
- Access Control: The system should have robust access control mechanisms to ensure only authorized users can access sensitive data.
- Compliance: The system should comply with healthcare regulations such as HIPAA (in the United States) for the protection

Performance Requirements

- Real-time Processing: The system should be capable of processing data in real-time or near-real-time to provide timely feedback.
- Scalability: The system should be scalable to handle large amounts of data from multiple patients concurrently.
- Reliability: The system should be highly reliable, with minimal downtime and a high degree of fault tolerance.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 2, August 2024

Reporting and Documentation

- Report Generation: The system should generate detailed reports based on the collected data and analysis, which can be used by healthcare providers.
- User Documentation: The system should include comprehensive user documentation for healthcare providers and patients on how to use the system effectively.
- Testing and Validation
- Validation Mechanisms: The system should include mechanisms for validating the accuracy and reliability of its Parkinson's disease detection algorithms.

3.2 PERFORMANCE REQUIREMENTS

When developing a model for Parkinson's disease (PD) detection, it is crucial to meet several performance requirements to ensure clinical utility and reliability. The model must achieve high accuracy to correctly classify the majority of cases and maintain high sensitivity to detect most individuals with PD, minimizing false negatives. Equally important is high specificity to reduce false positives, ensuring that healthy individuals are not misdiagnosed. Robustness is essential for consistent performance across diverse datasets and patient populations.The model should also be interpretable, providing insights into which features drive its predictions, thus aiding clinical decision-making. Computational efficiency is necessary for quick data processing, making the model suitable for real-time clinical applications. Rigorous validation, including cross-validation and external testing, ensures reliability and generalizability. Additionally, the model should be scalable to handle growing datasets and comply with healthcare regulations to protect patient data privacy. Finally, seamless integration with existing clinical workflows is vital for providing actionable insights without disrupting current diagnostic and treatment processes. By fulfilling these requirements, the model can effectively aid in the early detection and management of Parkinson's disease, enhancing patient outcomes and quality of life.

3.3 SOFTWARE REQUIREMENTS

1. Programming Language

- Python: The system should be implemented in Python, which is well-suited for data science and machine learning, particularly for implementing SVC models.

2. Machine Learning Framework

- Scikit-learn: The system should primarily use Scikit-learn, a powerful and easy-to-use library for implementing SVC and other machine learning algorithms. Scikit-learn provides robust tools for model training, validation, and performance evaluation.

3. Additional Libraries

- Pandas: For data manipulation and management, including handling datasets, performing feature engineering, and preparing data for model training.
- NumPy: For numerical computations, particularly useful in processing arrays and matrices, which are common in machine learning tasks.
- SciPy: For scientific computing tasks such as statistical analysis and data preprocessing.
- Matplotlib and Seaborn: For data visualization, which is useful for exploratory data analysis (EDA), feature selection, and visualizing the performance of the SVC model.

4. Data Handling and Preprocessing

- Imbalanced-learn: For handling imbalanced datasets, which is common in medical data, by providing techniques like SMOTE (Synthetic Minority Over-sampling Technique) to balance the classes before feeding the data into the SVC model.
- Feature Selection Tools: Utilize tools within Scikit-learn or other libraries like MLX tend to perform feature selection, which is crucial in improving the performance of the SVC model.

IJARSCT ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53 Volume 4, Issue 2, August 2024

5. Performance Evaluation

- Cross-validation: Scikit-learn provides tools for cross-validation, which should be used to ensure the SVC model generalizes well to unseen data.
- GridSearchCV: For hyperparameter tuning to find the best combination of parameters for the SVC model.

3.4 HARDWARE REQUIREMENTS

• CPU: Multi-core processor (e.g., Intel Core i5 or AMD Ryzen 5) with a minimum clock speed of 3 GHz.
• GPU (Optional): NVIDIA GPUs (e.g., GeForce GTX/RTX) for accelerated deep learning tasks.
• RAM: Minimum 8GB, preferably 16GB for handling larger datasets.
• Storage: SSD with at least 5GB for faster data access.

IV. TECHNOLOGY USED

Parkinson's disease detection utilizes a combination of advanced technologies to identify the presence of the disease through the analysis of various symptoms. Wearable sensors, such as accelerometers and gyroscopes, monitor motor functions, detecting tremors and gait abnormalities. Speech analysis tools process voice recordings to identify changes in speech patterns, like reduced pitch variation, which are indicative of Parkinson's. Handwriting analysis is conducted using digital tablets and smart pens to detect micrographia and other handwriting changes. These data inputs are analysed using machine learning algorithms like Support Vector Classification (SVC) and neural networks, which classify and predict the likelihood of Parkinson's based on the extracted features. Additionally, neuroimaging techniques, such as MRI and PET scans, are used to observe structural and functional brain changes associated with the disease, further aiding in diagnosis. These technologies are often integrated within data platforms that combine multiple data sources, enabling comprehensive analysis and improving the accuracy of Parkinson's disease detection.

V. DATA DESCRIPTION

In Parkinson's disease detection using Support Vector Classification (SVC), voice data is utilized to identify vocal impairments that are indicative of the disease. Voice recordings are collected from individuals performing tasks like sustaining vowel sounds or reading sentences. These recordings are processed to extract relevant features such as jitter (frequency variation), shimmer (amplitude variation), and harmonic-to-noise ratio (HNR), which reflect changes in speech patterns. The extracted features are then stored in CSV (Comma-Separated Values) files, where each row represents a different voice sample and each column represents a specific feature. For instance, a CSV file might include columns for jitter, shimmer, HNR, and Mel-Frequency Cepstral Coefficients (MFCCs), capturing the spectral characteristics of the voice. This structured data is then fed into an SVC model to classify and predict the presence of Parkinson's disease based on these vocal features.

VI. DATA PREPARATION FOR MODEL TRAINING

Data Acquisition: Collect information in CSV format which provide data about the voice features.
For Parkinson's disease detection, particularly when using Support Vector Classification (SVC) or any other machine learning model, data preprocessing steps might vary slightly compared to image-based tasks. Here's how you could adapt the steps to suit a Parkinson's disease detection project

1.Normalization:

- Description: Standardize or normalize your feature values to ensure that all features contribute equally to the model. For instance, if you have numerical features (like gait speed or tremor frequency), you might normalize them to a range (e.g., 0 to 1) or use z-score normalization.
- Purpose: This helps the SVC model perform better by avoiding dominance of features with larger scales and ensuring faster convergence during training.

2. Feature Extraction:

- Description: Extract relevant features from the raw data that can help in differentiating between Parkinson's patients and healthy individuals. For example, features could include speech patterns, handwriting dynamics, or motor control data.

- Purpose: Relevant feature extraction is crucial for SVC since it relies heavily on the quality and separability of the input features.

3. Data Augmentation (if applicable):

- Description: If you are dealing with time-series data or any sequential data (like voice recordings or movement data), you can augment the dataset by adding noise, shifting, or scaling the data.
- Purpose: Augmentation helps increase the dataset size, which can lead to better generalization of the model, especially when the dataset is small.

4. Dimensionality Reduction:

- Description: Techniques like PCA (Principal Component Analysis) or LDA (Linear Discriminant Analysis) can be used to reduce the dimensionality of the dataset. This helps in speeding up the training process and removing noise from the data.
- Purpose: It helps the SVC model focus on the most important features and reduces the risk of overfitting.

5. Data Splitting:

- Description: Split your data into training, validation, and test sets. Typically, the split might be 70% for training, 15% for validation, and 15% for testing.
- Purpose: This step ensures that your model is trained on one part of the data, validated on another (for hyperparameter tuning), and tested on a separate set to evaluate its performance.

6. Addressing Data Issues:

- Description: Handle any missing data or outliers in your dataset. This could involve imputation techniques, removing incomplete records, or using methods like median substitution.
- Purpose Icing a clean and harmonious dataset is vital for the model to learn effectively and to help deceiving results.

7. Metadata Integration(if applicable)

- Description If you have fresh metadata(like patient age, drug status, or complaint duration), insure it's intertwined with the main dataset. This can involve incorporating data from different sources or aligning time-series data.
- Purpose Including metadata can ameliorate the model's prophetic power by furnishing fresh environment that might be pivotal for accurate opinion. These way would prepare your data meetly for erecting a model to descry Parkinson's complaint, particularly when using an SVC.

VII. SVC EXPLANATION

When enforcing Support Vector Bracket(SVC) for Parkinson's complaint discovery, the process generally begins with importing the necessary libraries, including scikit- learn for machine literacy tasks. The dataset is also loaded, frequently containing features uprooted from patient data, similar as voice recordings or movement patterns. These features are pivotal as they serve as the input to the SVC model, which aims to classify whether a case has Parkinson's complaint.

Before feeding the data into the model, it's essential to preprocess it. This involves normalization, where point values are gauged to insure that no single point disproportionately influences the model. In some cases, dimensionality reduction ways like PCA(star element Analysis) may be applied to reduce the point space, helping the model focus on the most critical information and perfecting computational effectiveness.

Next, the dataset is resolve into training and testing sets, generally following a standard rate like 70- 30 or 80- 20. The training set is used to train the SVC model, where the algorithm learns to identify patterns in the data that distinguish between Parkinson's andnon-Parkinson's cases. The SVC algorithm works by chancing the optimal hyperplane that stylish separates the two classes in the point space. The model's hyperparameters, similar as the regularization parameter C and the kernel type(e.g., direct, radial base function), can be fine- tuned usingcross-validation on the training set to achieve the stylish performance.

IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Impact Factor: 7.53

Volume 4, Issue 2, August 2024

Once the model is trained, it's tested on the unseen test set to estimate its delicacy and conception capability. Metrics like delicacy, perfection, recall, and the F1- score are generally used to assess how well the model performs in prognosticating Parkinson's disease.However, the model can also be used for prognosticating new cases, furnishing a precious tool for early discovery and opinion of Parkinson's complaint, If the results are satisfactory.

X = data.drop(' status', axis = 1)# Replace' target_column' with the factual target column name

y = data(' status')

scaler = StandardScaler()#standardise the dataset of

X = scaler.fit_transform(X)

svc = SVC(kernel = ' direct')

chooser = RFE(svc, n_features_to_select = 10, step = 1)#recursive point elimination

X_selected = selector.fit_transform(X_scaled, y)

print(" named features", selector.support,)

print(" point ranking", selector.ranking,)

The confusion matrix and the bracket results for the model are as follows

```
[22]: print(confusion_matrix(y_test, y_pred))

[[ 9  6]
 [ 0 44]]

[23]: print(classification_report(y_test, y_pred))

              precision    recall  f1-score   support

           0       1.00      0.60      0.75        15
           1       0.88      1.00      0.94        44

    accuracy                           0.90        59
   macro avg       0.94      0.80      0.84        59
weighted avg       0.91      0.90      0.89        59
```

VIII. RESULTS

```
Prediction: 1
C:\Users\91800\AppData\Local\Programs\Python\Python312\Lib\site-packages\sklearn\base.py:493: UserWarning: X does not have valid feature names, but StandardScaler was fitted with feature names
  warnings.warn(
```

IX. CONCLUSION

In conclusion, using machine literacy ways for Parkinson's Disease discovery holds immense pledge in perfecting individual delicacy and case issues. By exercising tools like pandas for data preprocessing and scikit- learn for model training, experimenters can develop robust algorithms able of relating subtle complaint patterns. nonstop refinement and confirmation of these models are pivotal for their successful deployment in clinical settings, where early and accurate opinion is consummate for effective treatment and operation of Parkinson's Disease. Moving forward, integrating advanced interpretability styles with ethical considerations will further enhance trust and relinquishment of these technologies in healthcare practice, paving the way for further substantiated and effective case care strategies.

X. ACKNOWLEDGMENT

We'd like to express our deepest gratefulness to our council preceptors for furnishing the necessary coffers and support throughout the course of this exploration. Our sincere thanks go to our internal companion for their inestimable guidance, stimulant, and formative feedback, which greatly contributed to the success of this study.We also extend our appreciation to the platoon for their collaboration and perceptivity, as well as to all actors who took part in this exploration, without whom this study would not have been possible. Eventually, we're thankful to our families and musketeers for their unvarying support and understanding during the exploration and jotting process.

REFERENCES

[1]. Mahlknecht, P.; Krismer, F.; Poewe, W.; Seppi, K. Meta Analysis of Dorsolateral Nigral Hyperintensity on glamorous Resonance Imaging as a Marker for Parkinson's Disease. Mov. Disord. 2017, 32, 619 – 623.

[2]. Dickson, D.W. Neuropathology of Parkinson complaint. Parkinsonism Relat. Disord. 2018, 46(Suppl. 1), S30 – S33

[3]. Kalia, L.V.; Lang, A.E. Parkinson's Disease. Lancet 2015, 386, 896 – 912.

[4]. D. Heisters," Parkinson's symptoms treatments and exploration", vol. 20, no. 9, pp. 548- 554, 2011.

[5]. https// doi.org/10.33892Ffnagi.2021.633752
1877- 0509 © 2023 The Authors. Published by Elsevier B.V. This is an open access composition under the CC in- NC- ND license(https//creativecommons.org/licenses/by-nc-nd/4.0) Peer- review under responsibility of the scientific commission of the International Conference on Machine Learning and Data Engineering 10.1016/ j.procs.2023.01.007

[6]. Prabhavathi, K., Patil, S.(2022). " temblors and Bradykinesia. In Arjunan, S.P., Kumar, D.K.(eds) ways for Assessment of Parkinsonism for opinion and Rehabilitation ". Series in Bio Engineering. Springer. 135 – 149 https//doi.org/10.1007/978-981-16-3056-9_9.