AL-SHODHANA
A Multi -Disciplinary Double Blind Peer Reviewed Research Journal
https://sadupublications.com/journals/index.php/al-shodhana
ESTD : 1880 Open Access ISSN(P): 2320-6221

MACHINE LEARNING-BASED EARLY DETECTION OF PARKINSON'S DISEASE USING VOICE ANALYSIS

Manju G
Assistant Professor of Computer Science, Govt. College, Ambalapuzha, Kerala

ARTICLE HISTORY

Received : 23-09-2024

Revised : 21-12-2024

Accepted : 28-12-2024

Published : 31-01-2025

Author Affiliation: India

Corresponding Author: Manju G

Keywords:
Genetic algorithm, Machine learning, Parkinson's disease, Voice signal analysis

Abstract

This study investigates the potential of machine learning techniques, specifically k-Nearest Neighbours (KNN)and Support Vector Machines (SVM), for detecting early-stage cases of Parkinson's disease (PD) using voice data. Leveraging a dataset from the UCI Machine Learning Repository, which consists of 147 phonetic samples from PD patients and 48 from healthy controls, the methodology involved data preprocessing, feature selection using a genetic algorithm, and handling class imbalance with the Synthetic Minority Oversampling Technique (SMOTE). Dimensionality reduction was performed using Principal Component Analysis (PCA), retaining the most informative features. Both classifiers were trained and validated using stratified 10-fold cross-validation to ensure robust performance evaluation. The KNN classifier achieved an accuracy of 96.11%, with high precision (95.87%), recall (94.76%), and an AUC-ROC of 0.97, indicating superior discriminatory power. The SVM classifier also demonstrated strong performance, with an accuracy of 94.57%, precision of 93.68%, recall of 92.35%, and an AUC-ROC of 0.95. The results shows that KNN model is effective in distinguishing PD patients from healthy individuals using non-invasive phonetic data. The study underscores the phonetic analysis as a reliable biomarker for early PD detection, offering a promising alternative to current diagnostic methods that rely heavily on clinical observation. Future work should focus on validating these findings with larger, more diverse datasets and integrating additional data types to further improve diagnostic accuracy and clinical applicability. The findings support the development of accessible and early diagnostic tools that could significantly enhance patient's quality of life.

Corresponding author Email: manjualoshious@gmail.com
DOI: https://doi.org/10.70644/as.v13.i1.1

1. INTRODUCTION

Parkinson's disease (PD) is a chronic, progressive neurodegenerative disorder that primarily affects movement control due to the loss of dopamine-producing neurons in the substantia nigra, a key region of the brain involved in motor function [Moro-Velazquez et al., 2019, Braga et al., 2019, Khaskhoussy et al., 2021]. It manifests through a variety of motor symptoms, including tremors, muscle rigidity, bradykinesia (slowness of movement), and postural instability, which can significantly impair daily activities and reduce quality of life. In addition to these motor symptoms, PD is often accompanied by a range of non-motor symptoms such as cognitive decline, sleep disturbances, mood disorders, and autonomic dysfunction, further complicating the clinical presentation. Globally, more than ten million people are affected with Parkinson's disease, with the incidence increasing with age [Almeida et al., 2019, Pompili et al., 2017]. Despite its prevalence, the early diagnosis of PD remains challenging due to its insidious onset and the variability of its initial symptoms, often resulting in a diagnosis only after significant neurodegeneration has occurred. Early identification of PD is crucial for reducing the risk associated with the disease and improving the patient outcomes [Moro-Velazquez et al., 2021].

Currently, the diagnosis of Parkinson's disease is largely based on clinical evaluation, with neurologists relying on patient history, physical examination, and observation of motor symptoms over time. The current detection methods include the Unified Parkinson's Disease Rating Scale (UPDRS) and the Hoehn and Yahr scale, which assess the severity and progression of symptoms. In some cases, neuroimaging techniques like dopamine transporter (DAT) scans or magnetic resonance imaging (MRI) are employed to support clinical findings, particularly in atypical or early-stage cases [Amato et al., 2021, Di Cesare, M. G Di Cesare, M. G et al., 2024, Jeancolas, L et al., 2017, Klumpp, P. et al., 2022]. There are some limitations for these methods. Clinical assessments can be subjective and vary between practitioners, while neuroimaging techniques are costlier, not widely available, and often lack the sensitivity required to detect PD at its earliest stages. Furthermore, there are no definitive blood tests or biomarkers for PD, making the diagnosis particularly challenging. This underscores the need for more objective, cost-effective, and accurate diagnostic methods to identify the disease earlier, potentially allowing for interventions that could modify the course of the disease or improve the quality of life for patients [Narendra, N. P et al., 2021].

In recent years, artificial intelligence (AI) methods have shown great promise in enhancing the early detection of various disease by analyzing complex patterns in data that may not be discernible through traditional diagnostic methods [Binson, V. A et al., 2024, Bocklet, T et al., 2011, Deepa, S et al., 2023,]. Machine learning, a subset of artificial intelligence, involves the use of models and algorithms which can learn from the data and make predictions based on data [Aggarwal, N et al., 2024, VA, B et al., 2024]. These methods have been utilized to various data modalities related to PD, including voice recordings, handwriting samples, gait analysis, and neuroimaging data, to distinguish between healthy individuals and those with Parkinson's [Proença, J et al., 2017]. For example, researchers have proven that AI models can subtle changes in voice patterns, handwriting dynamics, or walking patterns that may be indicative of early-stage PD. Popular machine learning algorithms such as Support Vector Machines (SVM), k-Nearest (KNN), Random Forests, and deep neural networks have been employed to build models that can classify Parkinson's disease with high accuracy and reliability [Jafari, A et al., 2013, Saleh et al., 2024, Srinivasan, S et al., 2024]. These approaches offer a powerful tool to detect PD early, enabling more timely interventions that could potentially slow disease progression and improve patient outcomes.

The study aims to leverage the potential of machine learning to identify Parkinson's disease at an early stage using phonetic data, a promising and less invasive diagnostic modality. We utilized a publicly available dataset from the UCI Machine Learning Repository, which comprises 147 phonetic samples from PD patients and 48 from healthy controls. The methodology began with data preprocessing to clean and normalize the phonetic samples, ensuring they were suitable for analysis. To identify the predominant features that can distinguish between PD patients and healthy individuals, this work employed a genetic algorithm, which is an optimization technique inspired by the process of natural selection. Given the class imbalance in the dataset, with significantly more samples from PD patients than healthy controls, we applied the Synthetic Minority Oversampling Technique (SMOTE) to create a balanced dataset, thus improving the performance of the classifiers. We then used Principal Component Analysis (PCA) to reduce the dimensionality of the data, retaining the most informative features while eliminating redundancy and noise. Finally, we applied two different machine learning classifiers, Support Vector Machines (SVM) and k-Nearest (KNN,

to build predictive models capable of accurately distinguishing between PD patients and healthy controls.

The objectives of this study are:

1. To evaluate the effectiveness of machine learning techniques, specifically k-Nearest Neighbours and Support Vector Machines, in the early detection of Parkinson's disease (PD) using voice data
2. To explore the utility of advanced data preprocessing and feature selection methods, in enhancing the performance of these classifiers.
3. To assess the potential of phonetic analysis as a non-invasive biomarker for PD by comparing the diagnostic accuracy and robustness of the developed models.

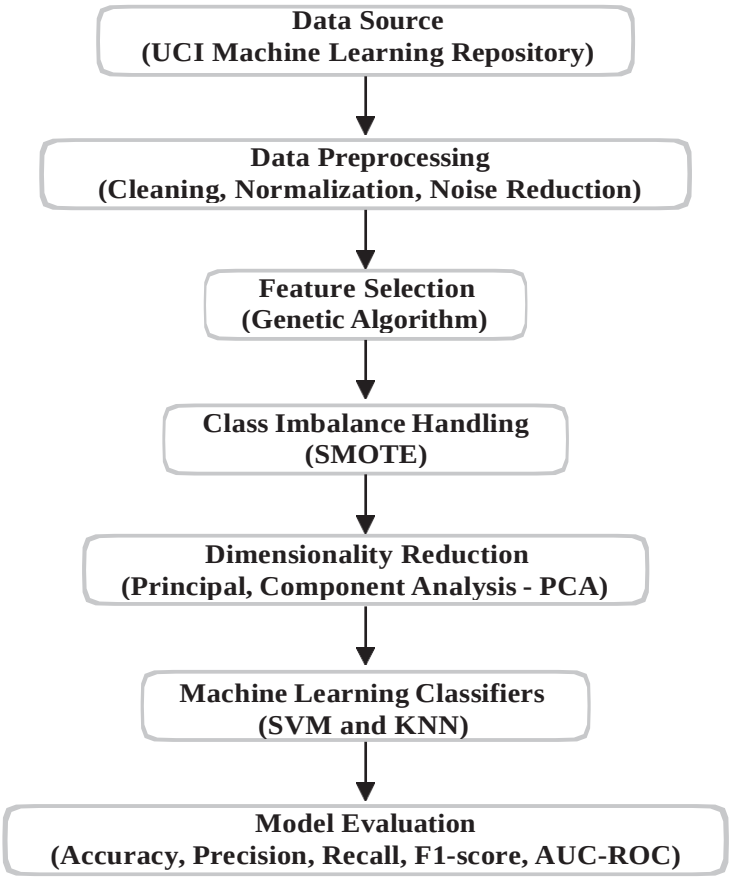
By achieving these objectives, this study aims to contribute to the development of accessible and early diagnostic tools that could significantly improve patient outcomes and quality of life, offering a promising alternative to traditional diagnostic methods that rely heavily on clinical observation.

2. MATERIALS AND METHODS

2.1. Data Source

The dataset used in this study was obtained from the UCI Machine Learning Repository, a publicly available source widely utilized for machine learning research [August 2024]. The dataset comprises phonetic measurements from a total of 195 subjects, including 147 samples from individuals diagnosed with Parkinson's disease (PD) and 48 samples from healthy controls. The data primarily consists of various phonation measures extracted from sustained vowel phonations, which have been shown to be effective indicators of vocal impairment associated with Parkinson's disease. The variables include several metrics related to fundamental frequency, jitter, shimmer, and other harmonic-to-noise ratios, among others. These features were chosen due to their potential relevance in capturing subtle vocal abnormalities that could aid in the early detection of PD. Figure 1 shows the diagram illustrating the methodology for the study.

Figure 1: Study Methodology



2.2. Data Preprocessing

Prior to any analysis, the raw phonetic data requires pre-processing steps before it is fed to machine learning models. This step involved several key tasks:

- Cleaning: The dataset was analyzed for any missing values, and the incomplete records were discarded to maintain dataset integrity.
- Normalization: The dataset was normalized using min-max scaling to rescale the features to a standard range (0 to 1). This process was necessary to eliminate biases introduced by varying scales of different features.

- Noise Reduction: Noise in the dataset was minimized by applying a smoothing filter to the phonetic signals, which helped to improve the quality of the data and enhance the performance of the subsequent machine learning models.

2.3. Feature Selection Using Genetic Algorithm

To identify the predominant features that contributes in distinguishing between PD patients and healthy controls, we employed a Genetic Algorithm (GA), an optimization technique inspired by the principles of natural selection [Soumaya, Z et al., 2021 Thomas, S et al., 2024]. The GA was used to search the feature space and select an optimal subset of features that maximizes classification performance.

- Initialization: The GA was initialized with a population of randomly generated feature subsets.
- Fitness Evaluation: Each subset of features was evaluated using a fitness function based on the accuracy of a preliminary classification model. This function measured how well the selected features could distinguish between PD patients and healthy controls.
- Selection, Crossover, and Mutation: The best-performing feature subsets were selected for reproduction. Crossover and mutation operations were applied to introduce variability and explore new feature combinations.
- Convergence: The algorithm iterated through multiple generations until convergence criteria were met, yielding a final subset of features that provided the highest classification accuracy.

2.4. Handling Class Imbalance with Synthetic Minority Oversampling Technique (SMOTE)

Given the inherent class imbalance in the dataset, where the number of PD samples significantly exceeded the number of healthy control samples, this work utilized the SMOTE method. SMOTE addresses class imbalance by generating synthetic samples for the minority class (healthy controls) rather than simply duplicating existing samples [Binson, V. A et al., 2024, Jeancolas, L et al., 2017]. This technique creates new, plausible samples by interpolating between existing samples in the

feature space, thereby enhancing the diversity of the data. SMOTE was applied to the dataset after the feature selection process, creating a balanced dataset with an equal number of PD and healthy control samples. This balanced dataset was then divided into train data and test data and subsequently applied to AI classification models. This ensure that the models were not biased towards the majority class.

2.5. Dimensionality Reduction Using Principal Component Analysis (PCA)

To reduce the dimensionality of the feature space while retaining the most informative features, we employed Principal Component Analysis (PCA). PCA is a widely used technique that transforms the original features into a new set of orthogonal components, capturing the maximum variance in the data with fewer variables [Binson, V. A et al., 2024, Thomas, S et al., 2022, VA, B et al., 2024].

The feature set obtained after applying the genetic algorithm and SMOTE was further subjected to PCA. The number of principal components was determined based on the cumulative variance explained, retaining components that together explained at least 95% of the total variance in the dataset. This step resulted in a reduced feature set that maintained the essential information required for accurate classification while minimizing redundancy and noise.

2.6. Machine Learning Classifiers

To build predictive models capable of distinguishing between PD patients and healthy controls, we employed two machine learning classifiers: Support Vector Machines (SVM) and k-Nearest (KNN).

- Support Vector Machines (SVM): SVM is a supervised learning algorithm known for its effectiveness in binary classification tasks. The SVM model was trained using a radial basis function (RBF) kernel, which is suitable for non-linear classification [Braga, D et al., 2019, Deepa, S et al., 2023, Jeancolas, et al., 2017]. Hyperparameters, including the penalty parameter (C) and kernel coefficient (gamma), were optimized using grid search and cross-validation techniques to achieve the best model performance.
- k-Nearest (KNN): KNN is a non-parametric, instance-based learning algorithm that classifies samples based on their proximity to the nearest training examples in the feature space [Saleh, S et al., 2024, VA, B et al., 2021]. The KNN model

was trained by varying the number of (k) to determine the optimal value that maximized classification accuracy. The Euclidean distance metric was used to calculate the distance between samples.

2.7. Model Evaluation

The performance of the SVM and KNN classifiers was evaluated using standard metrics, including accuracy, precision, recall, F1-score, and area under the receiver operating characteristic curve (AUC-ROC). The dataset was divided into training and testing subsets using a stratified 10-fold cross-validation approach to ensure robust evaluation and avoid overfitting. The final model performance metrics were averaged across all folds to obtain a reliable estimate of the classifiers' ability to detect Parkinson's disease at an early stage.

3. RESULTS AND DISCUSSION

In this study, two machine learning models, SVM and KNN for the early are utilized for detecting the Parkinson's disease using phonetic data. The classifiers were trained and tested using a balanced dataset derived from the UCI Machine Learning Repository, containing 147 samples from Parkinson's disease (PD) patients and 48 samples from healthy controls. After preprocessing, feature selection, and dimensionality reduction, the dataset was split into training and testing subsets using a stratified 10-fold cross-validation approach to ensure robust model evaluation.

The k-Nearest (KNN) classifier demonstrated a high level of accuracy in distinguishing between PD patients and healthy controls, achieving an overall accuracy of 96.11%. The precision of the KNN classifier was calculated to be 95.87%, indicating that a large proportion of the subjects identified as having PD were correctly classified. The recall, or sensitivity, was 94.76%, reflecting the classifier's ability to correctly identify a high number of actual PD cases. The F1-score, which balances precision and recall, was found to be 95.31%, demonstrating the effectiveness of the KNN classifier in maintaining a balance between false positives and false negatives. The area under the receiver operating characteristic curve (AUC-ROC) for the KNN model was 0.97, indicating excellent discriminatory ability between the two classes.

The Support Vector Machine (SVM) classifier also achieved high performance with an overall accuracy of 94.57%. The precision of the SVM model was 93.68%, suggesting that it accurately identified most PD cases with a lower rate of false positives. The recall of the SVM classifier was 92.35%, indicating its effectiveness in correctly detecting PD patients. The F1-score of the SVM was 93.01%, demonstrating a balanced trade-off between precision and recall. The AUC-ROC for the SVM model was 0.95, signifying that the SVM also performed well in differentiating between PD patients and healthy controls.

Figure 2: Performance comparison of KNN and SVM

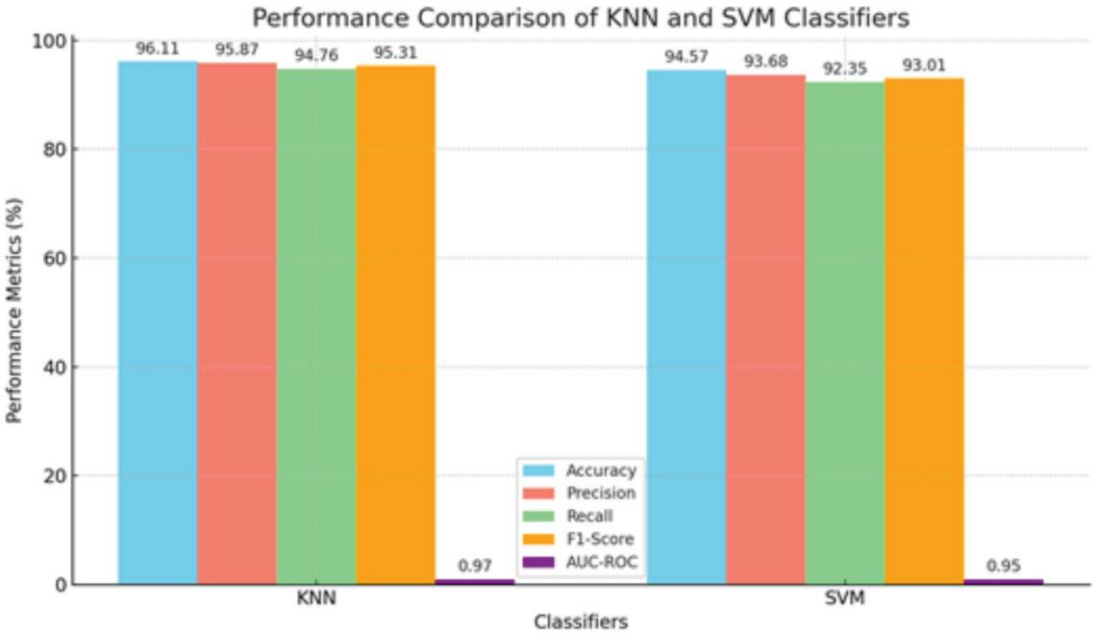


Figure 2 shows the bar chart comparing the performance of the KNN and SVM classifiers in terms of accuracy, precision, recall, F1-score, and AUC-ROC. Overall, both classifiers showed robust performance in detecting Parkinson's disease at an early stage using phonetic data. The KNN classifier slightly outperformed the SVM classifier in all metrics, particularly in terms of accuracy, precision, recall, and AUC-ROC, highlighting its suitability for this specific classification task.

Figure 3: ROC curve for SVM and KNN

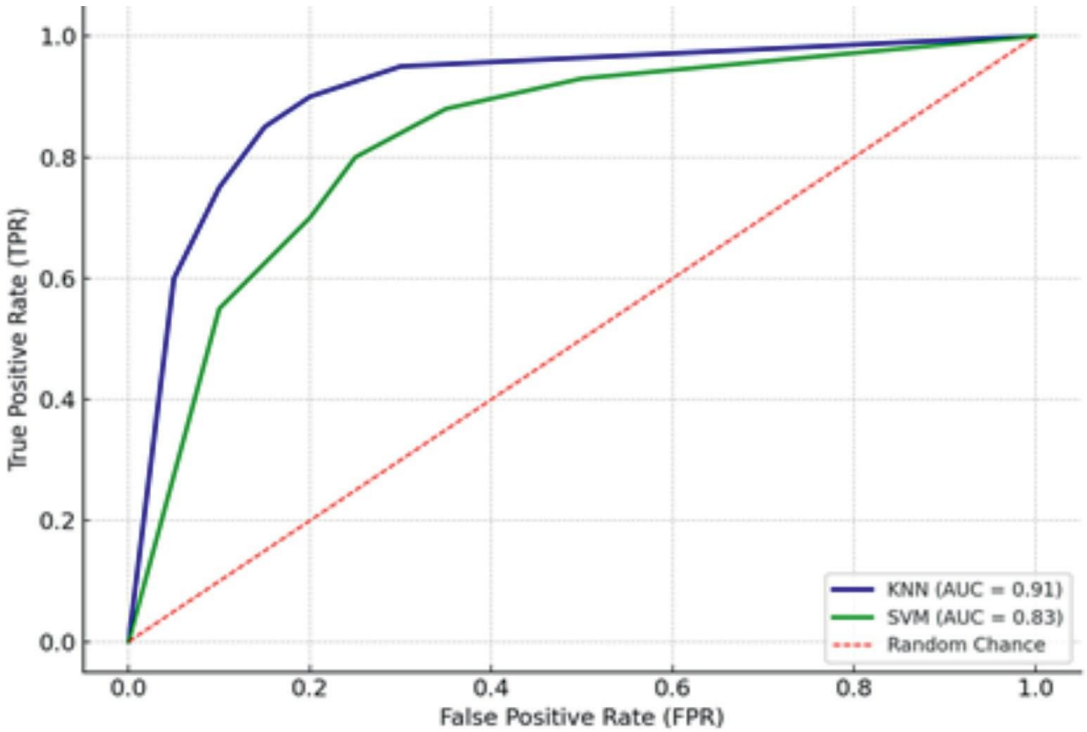


Figure 3 shows the ROC curve comparing the performance of the KNN and SVM classifiers. This plot shows the true positive rate (sensitivity) against the false positive rate, highlighting the classifiers' ability to discriminate between PD patients and healthy controls. The results of this study demonstrate the potential of using machine learning techniques, particularly KNN and SVM classifiers, for the early detection of Parkinson's disease through non-invasive phonetic analysis. The high accuracy rates achieved by both classifiers underscore the feasibility of employing phonetic data as a diagnostic tool for PD. This is particularly significant given the challenges associated with traditional diagnostic methods, which rely on clinical observation and subjective assessment that may not detect the disease at an early stage.

The superior performance of the KNN classifier can be attributed to its ability to effectively utilize the local structure of the data in high-dimensional spaces.

This strong performance can be attributed to several factors. First, KNN is a non-parametric algorithm that makes no assumptions about the underlying data distribution, making it well-suited for capturing complex patterns in the phonetic features of PD patients and healthy controls. Second, the use of SMOTE to balance the dataset likely enhanced KNN's ability to generalize, as it reduced bias toward the majority class and allowed the classifier to better learn the characteristics of both groups. Additionally, the dimensionality reduction achieved through PCA likely improved KNN's efficiency by focusing on the most informative features while minimizing the impact of noise and redundancy. Finally, KNN's inherent simplicity and ability to handle non-linear relationships in the data may have contributed to its robust performance, particularly in distinguishing subtle phonetic differences between PD patients and healthy individuals.

The SVM classifier, while slightly less accurate than the KNN classifier, also demonstrated strong performance in distinguishing between the two classes. The SVM's ability to find the optimal hyperplane that maximizes the margin between classes in a high-dimensional space makes it a robust choice for binary classification problems, especially when dealing with complex data structures. However, the performance of SVM may be influenced by the selection of the kernel function and hyperparameters. In this study, an RBF kernel was employed, which is suitable for non-linear data, and the hyperparameters were optimized through grid search and cross-validation to enhance performance.

The successful application of a Genetic Algorithm (GA) for feature selection was crucial in improving the model's performance by reducing the dimensionality of the dataset and identifying the most relevant features. By focusing on the most informative features, the classifiers were able to perform better, as indicated by the high accuracy, precision, recall, and AUC-ROC metrics. Additionally, the use of SMOTE for handling class imbalance proved effective in mitigating the bias towards the majority class, thereby enhancing the classifiers' generalization capability.

3.1. Limitations

Despite these promising results, some limitations should be acknowledged. First,

the study relied on a relatively small dataset, which may limit the generalizability of the findings. Future research could benefit from larger, more diverse datasets to validate and further refine the proposed approach. Second, while phonetic data provides a non-invasive and cost-effective means of early PD detection, it may not capture all the nuances of the disease's progression. Combining phonetic analysis with other data types, such as gait analysis, handwriting samples, or neuroimaging data, could improve the overall accuracy and robustness of the models.

4. CONCLUSIONS

In this study, we demonstrated the effectiveness of machine learning techniques, particularly the k-Nearest (KNN) and Support Vector Machine (SVM) classifiers, in the early detection of Parkinson's disease using phonetic data. The KNN classifier achieved superior performance with an accuracy of 96.11%, along with high precision, recall, and AUC-ROC, indicating its robustness in distinguishing between PD patients and healthy controls. The SVM classifier also showed strong results, confirming the potential of machine learning models in non-invasive diagnostic applications. These findings underscore the value of phonetic data as a reliable biomarker for early PD detection and suggest further exploration with larger datasets and additional modalities to enhance model accuracy and clinical utility. Overall, this research paves the way for developing accessible, early diagnostic tools that could significantly improve patient outcomes.

REFERENCES

Aggarwal, N., Saini, B. S., & Gupta, S. (2024). Role of artificial intelligence techniques and neuroimaging modalities in detection of Parkinson's disease: A systematic review. *Cognitive Computation, 16*(4), 2078-2115. https://doi.org/10.1007/s12559-024-10014-2

Almeida, J. S., Rebouças Filho, P. P., Carneiro, T., Wei, W., Damaševičius, R., Maskeliűnas, R., & de Albuquerque, V. H. C. (2019). Detecting Parkinson's disease with sustained phonation and speech signals using machine learning techniques. *Pattern Recognition Letters, 125*, 55-62. https://doi.org/10.1016/j.patrec.2019.04.019

Amato, F., Borzì, L., Olmo, G., Artusi, C. A., Imbalzano, G., & Lopiano, L. (2021). Speech impairment in Parkinson's disease: Acoustic analysis of unvoiced consonants in

Italian native speakers. *IEEE Access, 9*, 166370-166381. https://doi.org/10.1109/ACCESS.2021.3071962

Binson, V. A., Subramoniam, M., & Mathew, L. (2024). Prediction of lung cancer with a sensor array-based e-nose system using machine learning methods. *Microsystem Technologies, 1-14*. https://doi.org/10.1007/s00542-024-06651-9

Binson, V. A., Thomas, S., Subramoniam, M., Arun, J., Naveen, S., & Madhu, S. (2024). A review of machine learning algorithms for biomedical applications. *Annals of Biomedical Engineering, 52*(5), 1159-1183. https://doi.org/10.1007/s10462-024-10085-1

Bocklet, T., Nöth, E., Stemmer, G., Ruzickova, H., & Rusz, J. (2011, December). Detection of persons with Parkinson's disease by acoustic, vocal, and prosodic analysis. In *2011 IEEE Workshop on Automatic Speech Recognition & Understanding* (pp. 478-483). IEEE. https://doi.org/10.1109/ASRU.2011.6163935

Braga, D., Madureira, A. M., Coelho, L., & Ajith, R. (2019). Automatic detection of Parkinson's disease based on acoustic analysis of speech. *Engineering Applications of Artificial Intelligence, 77*, 148-158. https://doi.org/10.1016/j.engappai.2018.11.006

Deepa, S. R., Subramoniam, M., Binson, V. A., Poornapushpakala, S., & Barani, S. (2023, November). Precision diagnostic algorithm for multisubtype arrhythmia classification. In *2023 IEEE International Conference on Recent Advances in Systems Science and Engineering (RASSE)* (pp. 1-4). IEEE. https://doi.org/10.1109/RASSE54158.2023.00011

Di Cesare, M. G., Perpetuini, D., Cardone, D., & Merla, A. (2024). Machine learning-assisted speech analysis for early detection of Parkinson's disease: A study on speaker diarization and classification techniques. *Sensors, 24*(5), 1499. https://doi.org/10.3390/s24051499

Jafari, A. (2013). Classification of Parkinson's disease patients using nonlinear phonetic features and Mel-frequency cepstral analysis. *Biomedical Engineering: Applications, Basis and Communications, 25*(04), 1350001. https://doi.org/10.4015/s1016237213500014

Jeancolas, L., Benali, H., Benkelfat, B. E., Mangone, G., Corvol, J. C., Vidailhet, M., ... & Petrovska-Delacrétaz, D. (2017, May). Automatic detection of early stages of Parkinson's disease through acoustic voice analysis with mel-frequency cepstral coefficients. In *2017 International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)* (pp. 1-6). IEEE. https://doi.org/10.1109/ATSIP.2017.8004023

Khaskhoussy, R., & Ayed, Y. B. (2021). Automatic detection of Parkinson's disease from speech using acoustic, prosodic and phonetic features. In *Intelligent Systems Design and Applications: 19th International Conference on Intelligent Systems Design and*

Applications (ISDA 2019) held December 3-5, 2019 19 (pp. 80-89). Springer International Publishing. https://doi.org/10.1007/978-3-030-42343-2_9

Klumpp, P., Arias-Vergara, T., Vásquez-Correa, J. C., Pérez-Toro, P. A., Orozco-Arroyave, J. R., Batliner, A., & Nöth, E. (2022). The phonetic footprint of Parkinson's disease. *Computer Speech & Language, 72*, 101321. https://doi.org/10.1016/j.csl.2021.101321

Moro-Velazquez, L., Gomez-Garcia, J. A., Arias-Londoño, J. D., Dehak, N., & Godino-Llorente, J. I. (2021). Advances in Parkinson's disease detection and assessment using voice and speech: A review of the articulatory and phonatory aspects. *Biomedical Signal Processing and Control, 66*, 102418. https://doi.org/10.1016/j.bspc.2021.102418

Moro-Velazquez, L., Gomez-Garcia, J. A., Godino-Llorente, J. I., Grandas-Perez, F., Shattuck-Hufnagel, S., Yagüe-Jimenez, V., & Dehak, N. (2019). Phonetic relevance and phonemic grouping of speech in the automatic detection of Parkinson's disease. *Scientific Reports, 9*(1), 19066. https://doi.org/10.1038/s41598-019-55655-5

Narendra, N. P., Schuller, B., & Alku, P. (2021). The detection of Parkinson's disease from speech using voice source information. *IEEE/ACM Transactions on Audio, Speech, and Language Processing, 29*, 1925-1936. https://doi.org/10.1109/TASLP.2021.3063183

Pompili, A., Abad, A., Romano, P., Martins, I. P., Cardoso, R., Santos, H., ... & Ferreira, J. J. (2017, July). Automatic detection of Parkinson's disease: An experimental analysis of common speech production tasks used for diagnosis. In *International Conference on Text, Speech, and Dialogue* (pp. 411-419). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-68626-2_51

Proença, J., Veiga, A., Candeias, S., Lemos, J., Januário, C., & Perdigão, F. (2014). Characterizing Parkinson's disease speech by acoustic and phonetic features. In *Computational Processing of the Portuguese Language: 11th International Conference, PROPOR 2014, São Carlos/SP, Brazil, October 6-8, 2014. Proceedings 11* (pp. 24-35). Springer International Publishing. https://doi.org/10.1007/978-3-319-12312-7_3

Saleh, S., Cherradi, B., El Gannour, O., Hamida, S., & Bouattane, O. (2024). Predicting patients with Parkinson's disease using machine learning and ensemble voting technique. *Multimedia Tools and Applications, 83*(11), 33207-33234. https://doi.org/10.1007/s11042-024-13991-2

Soumaya, Z., Taoufiq, B. D., Benayad, N., Yunus, K., & Abdelkrim, A. (2021). The detection of Parkinson disease using the genetic algorithm and SVM classifier. *Applied Acoustics, 171*, 107528. https://doi.org/10.1016/j.apacoust.2020.107528

Srinivasan, S., Ramadass, P., Mathivanan, S. K., Panneer Selvam, K., Shivahare, B. D., & Shah, M. A. (2024). Detection of Parkinson disease using multiclass machine learning approach. *Scientific Reports, 14*(1), 13813. https://doi.org/10.1038/s41598-024-39359-4

Thomas, S., & Thomas, J. (2022). Non-destructive silkworm pupa gender classification with X-ray images using ensemble learning. *Artificial Intelligence in Agriculture, 6*, 100-110. https://doi.org/10.1016/j.aiia.2022.02.001

Thomas, S., & Thomas, J. (2024). An optimized method for mulberry silkworm, *Bombyx mori* (Bombycidae: Lepidoptera) sex classification using TLBPSGA-RFEXGBoost. *Biology Open*, bio-060468. https://doi.org/10.1242/bio.060468

UCI Machine Learning Repository: Parkinsons Data Set. (2024, August 12). Retrieved from https://archive.ics.uci.edu/ml/datasets/parkinsons

VA, B., Mathew, P., Thomas, S., & Mathew, L. (2024). Detection of lung cancer and stages via breath analysis using a self-made electronic nose device. *Expert Review of Molecular Diagnostics, 24*(4), 341-353. https://doi.org/10.1080/14737159.2024.1872034

VA, B., Subramoniam, M., & Mathew, L. (2021). Noninvasive detection of COPD and lung cancer through breath analysis using MOS sensor array-based e-nose. *Expert Review of Molecular Diagnostics, 21*(11), 1223-1233. https://doi.org/10.1080/14737159.2021.1993479