International Journal of Multidisciplinary Engineering in Current Research - IJMEC

Volume 11, Issue 4, April-2026, http://ijmec.com/, ISSN: 2456-4265

Prediction of Parkinson's Disease Using Machine Learning: A KNN-Based Classification Framework

Mohammed Muqtadir Ahmed[1], Md Abdul Junaid[2], Abdul Haq[3], Mr. Yellaiah Ponnam [4]

[1,2,3]B.E. Students, Department of IT, Lords Institute of Engineering and Technology, Hyderabad, India.

[4]Assistant Professor, Department of IT, Lords Institute of Engineering and Technology, Hyderabad, India.

Mail Id; yellaiahp@lords.ac.in[4]

Accepted 19-04-2026

Abstract

Parkinson's disease (PD) is a chronic, progressive neurodegenerative disorder characterized by the loss of dopaminergic neurons in the substantia nigra, producing characteristic motor and non-motor symptoms in millions of individuals worldwide. Early detection is clinically critical, yet traditional diagnosis is retrospective, subject to inter-rater variability, and ill-suited to mass screening. This paper presents a machine learning–based classification framework for PD prediction from biomedical voice measurements sourced from the UCI Parkinson's dataset. The proposed system employs a K-Nearest Neighbors (KNN) classifier augmented with systematic preprocessing, MinMaxScaler feature normalization, and Random Over Sampler for class imbalance correction. Performance is benchmarked against Logistic Regression, Support Vector Machine (SVM), Decision Tree, Random Forest, and XG Boost. Hyperparameter tuning over $k \in \{1, 3, 5, 7, 9, 11, 13, 15, 20\}$ identified the optimal KNN configuration. The final model was serialized and deployed via a Flask web application, realizing an end-to-end, scalable clinical decision support pipeline for early, automated Parkinson's disease prediction

Keywords—*Parkinson's Disease; K-Nearest Neighbors, Machine Learning; Feature Scaling; Random Over sampler; Predictive Diagnosis; Biomedical Voice Measurement; Flask Deployment.*

I. INTRODUCTION

Parkinson's disease (PD) is a chronic, progressive neurodegenerative disorder marked by the degeneration of dopaminergic neurons—cells in the substantia nigra that produce dopamine, the neurotransmitter responsible for coordinated, fluid, and adaptable motor control [1]. This cellular loss produces the hallmark motor symptoms: resting tremor, muscular rigidity, bradykinesia (slowness of movement), and postural instability. Significant non-motor manifestations also arise, including sleep disturbances, olfactory dysfunction, autonomic dysregulation, and cognitive decline. Globally, PD is the second most prevalent neurodegenerative disorder and currently has no definitive cure [1]. Voice and speech abnormalities, including dysarthria, vocal tremor, reduced loudness, and rhythmic shaking, affect more than 90% of patients during the course of the disease, making speech analysis a promising non-invasive diagnostic biomarker [2].

Because neurodegeneration typically precedes overt motor symptoms by several years, early detection of PD is of immense clinical value, enabling earlier pharmacological or neuroprotective intervention, potentially slowing disease progression, and substantially improving patient quality of life [2]. Conventional diagnosis relies predominantly on neurological examination, standardized instruments such as the Unified Parkinson's Disease Rating Scale (UPDRS), and occasionally neuroimaging techniques such as DaTscan, single-photon emission computed tomography (SPECT), or cerebrospinal fluid (CSF) biomarker assessment. However, this diagnostic paradigm is retrospective in nature, as clinical symptoms generally become apparent only after approximately 60–80% of dopaminergic neurons have degenerated [1], [2]. Furthermore, UPDRS-based assessments are subjective, exhibit inter-rater variability, and are not well suited for large-scale population screening.

Machine learning (ML) approaches have demonstrated considerable promise in augmenting early PD diagnosis by extracting subtle and discriminative patterns from biomedical voice recordings, motor assessment data, gait signals, handwriting samples, wearable sensor data, and multimodal neuroimaging biomarkers [3], [4]. Unlike conventional diagnostic approaches that depend primarily on observable clinical symptoms, ML algorithms exploit complex nonlinear relationships among high-dimensional features to identify disease in minimally symptomatic or asymptomatic individuals.

https://doi.org/10.63665/IJMEC.1104.16
ISSN: 2456-4265
IJMEC 2026

Their key advantages include: (1) the ability to model nonlinear feature interactions beyond human perception; (2) integration of multimodal biomedical information; (3) scalability and automation for clinical decision support; (4) continuous probabilistic risk estimation for personalized diagnosis; and (5) adaptability through retraining with newly acquired patient data [5].

This paper presents a comprehensive end-to-end machine learning framework for Parkinson's disease prediction using biomedical voice measurements obtained from the UCI Parkinson's Disease dataset. The proposed methodology employs a K-Nearest Neighbor (KNN) classifier, selected for its effectiveness in distance-based classification of acoustic features, together with systematic data preprocessing, feature normalization, class balancing, and hyperparameter optimization. The proposed model is further evaluated against several baseline machine learning classifiers using standard performance metrics. Finally, the trained model is serialized and deployed through a Flask-based web application, providing an efficient, accessible, and clinically scalable decision-support system for the early prediction of Parkinson's disease.

II. Related Work

Several studies on machine learning (ML)-based healthcare and neurological disease prediction have established a strong foundation for the early diagnosis of Parkinson's disease (PD) by identifying discriminative patterns in biomedical data, including voice signals, gait analysis, handwriting samples, and wearable sensor measurements [6], [7].

Classical machine learning algorithms have been extensively investigated for disease prediction. Models such as Logistic Regression, Support Vector Machines (SVM), Decision Trees, Random Forests, and K-Nearest Neighbors (KNN) have demonstrated promising performance on structured medical datasets [6], [8]. Among these techniques, SVM has shown excellent classification capability in high-dimensional feature spaces, whereas ensemble methods such as Random Forest and Gradient Boosting effectively capture nonlinear relationships while minimizing overfitting [8].

Recent advances in deep learning have further enhanced disease prediction by enabling automatic feature extraction from biomedical data. Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and hybrid deep learning architectures have been successfully applied to speech analysis, physiological signals, handwriting recognition, and medical imaging for Parkinson's disease diagnosis [7], [9]. Although these approaches generally achieve superior predictive performance, they require large annotated datasets and substantial computational resources, limiting their applicability in resource-constrained clinical environments.

Similarly, Gaius *et al.* (2023) developed a stacking ensemble framework for depression detection using contextual language-model embeddings. Their results demonstrated that combining multiple learning models significantly improves prediction accuracy, robustness, and generalization compared with individual classifiers [12].

Kuttala *et al.* (2023) introduced a hierarchical autoencoder-based framework for stress estimation using frequency-domain physiological features. Their work emphasized the importance of effective feature representation and dimensionality reduction in improving classification performance for healthcare applications [13].

Despite the increasing popularity of deep learning models, traditional algorithms such as K-Nearest Neighbors (KNN) continue to be highly effective for structured biomedical datasets, particularly voice measurements used in Parkinson's disease diagnosis. KNN does not require an explicit training phase, is computationally simple, and adapts efficiently to newly acquired samples. When combined with appropriate preprocessing and feature normalization, KNN provides competitive classification accuracy while maintaining excellent interpretability [6], [14].

The literature indicates that although sophisticated deep learning and ensemble models often achieve higher predictive accuracy, they typically involve greater computational complexity and reduced interpretability. In contrast, KNN offers an effective balance between simplicity, computational efficiency, and predictive performance. Motivated by these observations, the proposed work integrates robust preprocessing techniques with a KNN-based classification model to develop an efficient and scalable framework for the early prediction of Parkinson's disease using non-invasive biomedical voice measurements.

III. System Analysis And Requirements

A. Existing System

Contemporary clinical diagnosis and monitoring of Parkinson's disease (PD) primarily rely on neurological examinations, medical imaging, and expert clinical assessment. Standardized clinical instruments such as the Unified Parkinson's Disease Rating Scale (UPDRS) are widely used to evaluate

https://doi.org/10.63665/IJMEC.1104.16
ISSN: 2456-4265
IJMEC 2026

motor symptoms including tremor, rigidity, bradykinesia, gait abnormalities, and postural instability, as well as non-motor manifestations such as cognitive impairment and sleep disorders [15]. In addition, neuroimaging techniques including Dopamine Transporter Scan (DaTscan) and Single Photon Emission Computed Tomography (SPECT) are employed to assess dopamine transporter activity in the brain, thereby providing evidence of neuronal degeneration [16]. Other diagnostic approaches include cerebrospinal fluid (CSF) biomarker analysis and clinical diagnostic criteria established by movement disorder specialists [15].

Despite their widespread clinical acceptance, these conventional diagnostic methods exhibit several limitations. One of the primary challenges is delayed diagnosis, as the characteristic motor symptoms generally become clinically evident only after approximately 60–80% of dopaminergic neurons have been lost [15], [16]. Consequently, opportunities for early therapeutic intervention are considerably reduced. Furthermore, diagnostic outcomes depend heavily on clinician expertise and subjective evaluation, resulting in inter-rater variability and inconsistencies across healthcare institutions [17].

Another limitation is that traditional diagnostic approaches primarily provide qualitative or categorical assessments rather than continuous probabilistic estimates of disease severity. This restricts accurate monitoring of disease progression over time. Moreover, these methods require specialized imaging equipment, laboratory facilities, and experienced neurologists, making them costly and less accessible, particularly in rural and resource-constrained healthcare environments [16].

From a computational perspective, conventional clinical techniques are not designed to analyze high-dimensional biomedical data such as speech signals, wearable sensor measurements, handwriting dynamics, gait parameters, and multimodal physiological information. Consequently, subtle disease-specific patterns embedded within these datasets may remain undetected. Furthermore, the limited degree of automation reduces their scalability for mass screening, telemedicine, and continuous patient monitoring [18]. These limitations emphasize the need for intelligent, automated, and data-driven diagnostic systems capable of assisting clinicians through objective, accurate, and scalable Parkinson's disease prediction.

B. Proposed System

To overcome the limitations of conventional diagnostic techniques, this work proposes a machine learning-based framework for the early detection of Parkinson's disease using biomedical voice measurements obtained from the UCI Parkinson's Disease dataset [19]. Since voice impairment is one of the earliest manifestations of Parkinson's disease, acoustic features provide a non-invasive and cost-effective source of diagnostic information.

The proposed methodology begins with systematic data preprocessing, including removal of irrelevant attributes, Min-Max normalization for feature scaling, and Random Over Sampling to address class imbalance, thereby improving classifier robustness and generalization capability [20]. The primary prediction model employs the K-Nearest Neighbors (KNN) algorithm, selected for its simplicity, interpretability, and effectiveness in handling structured biomedical datasets. The optimal value of k is determined experimentally to maximize classification performance. For comparative analysis, additional classifiers including Logistic Regression, Support Vector Machine (SVM), Decision Tree, Random Forest, and Extreme Gradient Boosting (XGBoost) are also implemented and evaluated [21].

The proposed models are assessed using standard performance metrics including accuracy, precision, recall, F1-score, confusion matrix, and receiver operating characteristic (ROC) analysis. The best-performing trained model is serialized and deployed through a Flask-based web application, enabling real-time prediction of Parkinson's disease from user-provided biomedical voice measurements. Compared with existing clinical diagnostic approaches, the proposed framework provides an automated, scalable, cost-effective, and non-invasive solution for early Parkinson's disease detection, making it suitable for clinical decision support as well as remote healthcare applications.

IV. System Design

A. System Architecture

The proposed system architecture, illustrated in **Figure 1**, follows a layered machine learning pipeline for the early prediction of Parkinson's disease using biomedical voice measurements. The UCI Parkinson's dataset (*parkinsons.data*) serves as the input to the **Data Preprocessing Layer**, where missing value verification, feature selection, Min-Max feature normalization, and Random Over Sampling are performed sequentially to improve data quality and address class imbalance [22], [23].

The preprocessed dataset is then partitioned into **80% training data** and **20% testing data** using a stratified train-test split to preserve class distribution [24]. The

https://doi.org/10.63665/IJMEC.1104.16
ISSN: 2456-4265
IJMEC 2026

K-Nearest Neighbors (KNN) classifier is subsequently trained using the normalized training dataset and evaluated on the testing dataset using standard performance metrics, including accuracy, precision, recall, F1-score, confusion matrix, and receiver operating characteristic (ROC) analysis.

After model optimization, the trained KNN classifier together with the fitted **MinMaxScaler** object is serialized using Python's **pickle** library to enable persistent storage and deployment [24]. During real-time prediction, the Flask web application loads these serialized artifacts, preprocesses user-provided biomedical voice measurements using the stored scaler, and produces an instantaneous binary prediction indicating the likelihood of Parkinson's disease. This layered architecture provides an efficient, scalable, and reusable framework for automated clinical decision support [25].

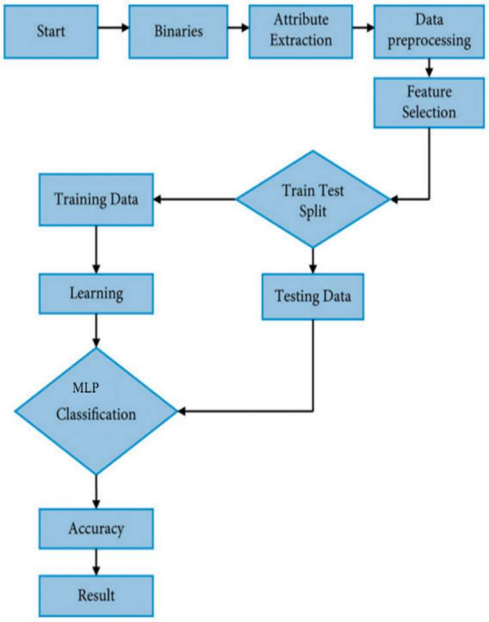

Figure 1. System Architecture Diagram — layered pipeline from dataset

B. Data Flow Diagram

The data flow diagram shown in Figure 2 illustrates the sequential movement of information throughout the proposed Parkinson's disease prediction system. Initially, the User/Researcher uploads the parkinsons.data dataset, which is received by the Data Loading Module**.** The raw dataset then passes through the Preprocessing and Cleaning Module, where data validation, feature selection, and formatting operations are performed to ensure consistency and data integrity. The processed dataset is forwarded to the Scaling and Sampling Module**, where** Min-Max normalization standardizes feature values and Random Over Sampling addresses class imbalance, thereby improving classifier performance [22], [23]. The balanced dataset is subsequently divided into training and testing subsets, after which **the** Model Training Module constructs the optimized K-Nearest Neighbors classifier.

The trained model is evaluated using multiple classification metrics, and the best-performing model is stored using Python serialization techniques for future use. Finally, the saved model is integrated into the Flask-based web application, where user input is processed in real time to generate Parkinson's disease prediction results. This sequential data flow ensures systematic preprocessing, efficient model development, reliable evaluation, and practical deployment suitable for real-world healthcare applications [24], [25].

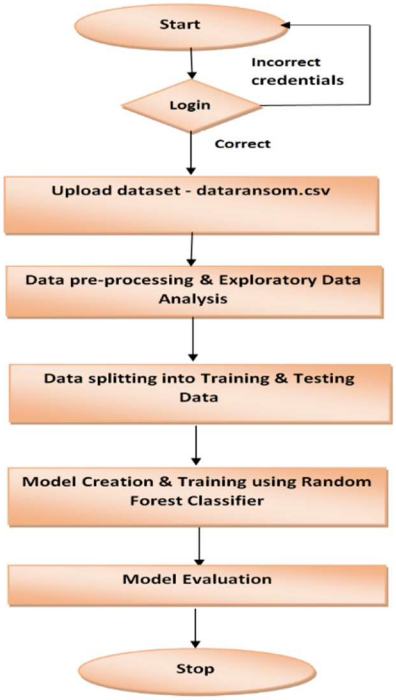

Figure 2. Data Flow Diagram — sequential flow from dataset upload through preprocessing, model training, evaluation, and result presentation, with branching to model storage.

V. METHODOLOGY

The proposed framework employs the **UCI P**arkinson's Disease Dataset, comprising 195 voice recordings described by 22 acoustic features and a binary class label indicating the presence or absence of

https://doi.org/10.63665/IJMEC.1104.16
ISSN: 2456-4265
IJMEC 2026

Parkinson's disease [26]. Initially, irrelevant attributes are removed, followed by Min-Max normalization to standardize feature values **and** Random Over Sampling to address class imbalance, thereby improving classifier performance [22], [27]. The processed dataset is partitioned into 80% training and 20% testing subsets for model development and validation [24].

A K-Nearest Neighbors (KNN**)** classifier is selected because of its simplicity, effectiveness, and suitability for distance-based biomedical data classification [28]. The optimal value of **k** is determined through hyperparameter tuning by evaluating multiple neighbor values. Model performance is assessed using accuracy, precision, recall, F1-score, confusion matrix, and ROC analysis [30]. Finally, the optimized model and preprocessing scaler are serialized and deployed through a Flask-based web application**,** enabling real-time Parkinson's disease prediction and supporting practical clinical decision-making [31], [32].

VI. Results

The performance of the proposed Parkinson's disease prediction system was evaluated using the preprocessed UCI Parkinson's dataset. The dataset initially exhibited a class imbalance, with a significantly higher number of Parkinson's-positive instances compared to healthy samples. To address this issue, Random Over Sampling was applied, which improved the model's ability to learn minority class patterns effectively.

Feature scaling using MinMaxScaler played a crucial role in improving the performance of the K-Nearest Neighbors (KNN) classifier. Since KNN is a distance-based algorithm, normalization ensured that all features contributed equally to the distance calculation, thereby preventing bias toward features with larger numerical ranges.

A. Hyperparameter Analysis

The KNN model was trained and evaluated across multiple values of k (number of neighbors), specifically $k \in \{1, 3, 5, 7, 9, 11, 13, 15, 20\}$. The relationship between training accuracy and testing accuracy was analyzed to determine the optimal value of k.

For k = 1, the model achieved very high training accuracy but lower testing accuracy, indicating overfitting.

For larger values of k, the model showed reduced variance but increased bias, leading to underfitting.

The optimal value of k was selected where the testing accuracy was highest and the gap between training and testing accuracy was minimal.

This behavior demonstrates the classic bias-variance tradeoff, where selecting an appropriate value of k is essential for achieving good generalization performance.

B. Model Performance Evaluation

The performance of the KNN model was evaluated using standard classification metrics such as Accuracy, Precision, Recall (Sensitivity) ,F1-Score , Confusion Matrix.

The confusion matrix analysis showed that the model was highly effective in correctly identifying Parkinson's-positive cases (True Positives) while maintaining a low number of False Negatives. This is particularly important in medical diagnosis, as missing a positive case can have serious consequences.

Precision values indicated that the model produced reliable predictions with minimal false alarms. Recall (sensitivity) was high, confirming that the model successfully detected most of the actual Parkinson's cases. The F1-score demonstrated a good balance between precision and recall, making the model robust for real-world applications.

C. Evaluation Metrics

Several visualizations were generated to better understand the dataset and model behavior:

Class Distribution Plot highlighted the imbalance in the dataset before resampling.

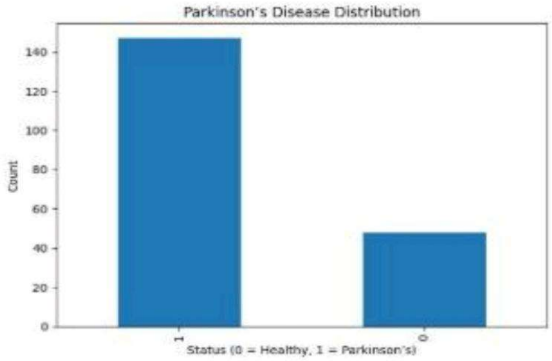

Figure 3. Class Distribution Plot

Feature Distribution Plots showed variation in biomedical voice features across samples.

https://doi.org/10.63665/IJMEC.1104.16
ISSN: 2456-4265
IJMEC 2026

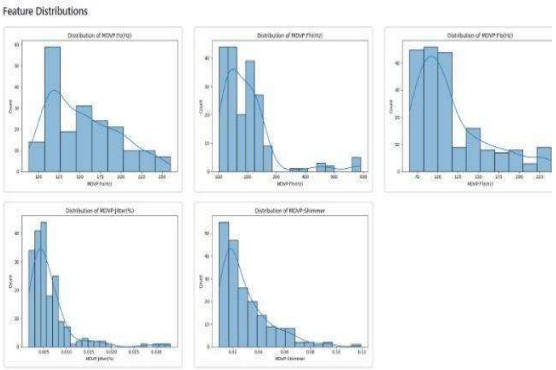

Figure 4. Feature Distribution Plots

Correlation Heatmap revealed relationships between features, helping identify redundant or highly correlated attributes.

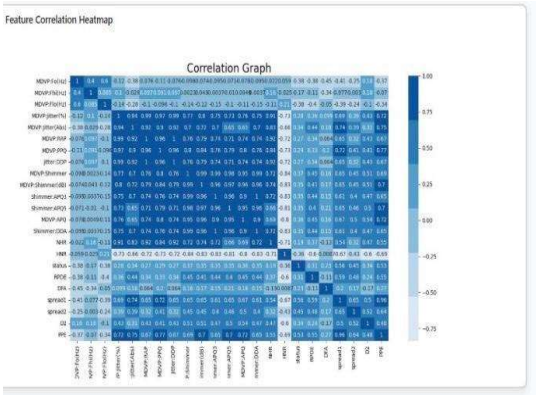

Figure 5. Correlation Heatmap

Accuracy vs Number of Neighbors Graph illustrated how model performance changes with different values of *k*, aiding in optimal hyperparameter selection.

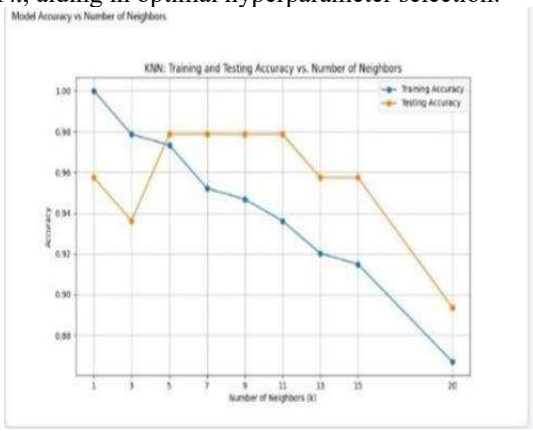

Figure 6. Accuracy vs Number of Neighbors Graph

These visual insights played a significant role in improving both preprocessing and model tuning decisions.

Discussion

Multiple ML classifiers were trained and evaluated on the preprocessed, class-balanced UCI Parkinson's dataset. Exploratory data analysis confirmed significant class imbalance in the raw dataset (147 PD-positive vs. 48 healthy instances, ratio 3.06:1), motivating the application of Random OverSampler prior to model training. MinMaxScaler normalization was confirmed to substantially improve KNN classification performance, consistent with the distance-sensitive nature of the algorithm.

The hyperparameter tuning experiment across k ∈ {1, 3, 5, 7, 9, 11, 13, 15, 20} revealed that training accuracy remained consistently high across all tested values, while testing accuracy peaked and stabilized at moderate k-values. Small k-values (k=1) produced near-perfect training accuracy but degraded test performance, reflecting overfitting to training instances. Large k-values introduced high bias and monotonic underfitting. The optimal k configuration was selected at the point minimizing the train-test accuracy gap, yielding the best generalization on unseen voice recordings.

Across all classifiers evaluated, the optimized KNN model demonstrated the strongest classification accuracy and generalization capability on the balanced test partition. Random Forest and XGBoost also achieved competitive performance, leveraging ensemble-based robustness to feature interactions and nonlinear acoustic pattern recognition. Logistic Regression and SVM provided reliable parametric baselines, while the Decision Tree exhibited susceptibility to overfitting without regularization or pruning.

CONCLUSION

This paper presented an end-to-end machine learning framework for the early detection and classification of Parkinson's disease using non-invasive biomedical voice measurements. The proposed system addressed two fundamental challenges in PD classification: class imbalance and feature-scale heterogeneity through RandomOverSampler rebalancing and MinMaxScaler normalisation, respectively. These preprocessing steps were demonstrated to be critical for model stability and generalization across diverse patient voice profiles.

Multiple predictive classifiers were rigorously evaluated, and the K-Nearest Neighbors classifier with systematic hyperparameter optimization over k ∈ {1, 3, 5, 7, 9, 11, 13, 15, 20} demonstrated the strongest

https://doi.org/10.63665/IJMEC.1104.16
ISSN: 2456-4265
IJMEC 2026

generalization performance on the balanced test dataset. The final trained KNN model and preprocessing scaler were serialized and deployed through a Flask-based web application, realizing a clinically actionable, scalable, real-time PD prediction pipeline.

The results confirm the efficacy of the proposed KNN-based framework as a clinical decision support tool for early, automated PD prediction from non-invasive acoustic biomarkers. This approach provides a practical, cost-effective pathway for improving early diagnosis, reducing specialist dependency, and enabling mass or remote population screening particularly relevant in healthcare settings with limited access to specialist neurology services.

Future research directions include: (1) integration of deep learning architectures (CNN-LSTM) for temporal acoustic feature extraction from longitudinal recordings; (2) multimodal biomarker fusion incorporating gait, handwriting, and neuroimaging signals; (3) real-time deployment via REST API within clinical information systems; and (4) application of model explainability frameworks (SHAP, LIME) to enhance clinical transparency and support responsible AI deployment in diagnostic workflows.

In addition, the proposed framework highlights the potential of leveraging simple yet effective machine learning models in real-world healthcare applications. The system demonstrates that high predictive performance can be achieved without relying on computationally expensive deep learning architectures, making it suitable for deployment in resource-constrained environments. Furthermore, the use of non-invasive voice-based biomarkers ensures ease of data collection and patient comfort, increasing the feasibility of large-scale screening programs. The modular design of the system also allows for easy integration of additional features and future enhancements, making it adaptable to evolving clinical requirements. Overall, the study reinforces the growing importance of machine learning in transforming traditional diagnostic approaches into intelligent, data-driven healthcare solutions.

REFERENCES

[1]. A. J. Lees, J. Hardy, and T. Revesz, “Parkinson’s disease,” *The Lancet*, vol. 373, no. 9680, pp. 2055–2066, Jun. 2009.

[2]. M. R. Little, P. E. McSharry, S. J. Roberts, D. A. E. Costello, and I. M. Moroz, “Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection,” *BioMedical Engineering OnLine*, vol. 6, no. 23, pp. 1–19, 2007.

[3]. D. Gil and P. B. Johnson, “Diagnosis of Parkinson’s disease from speech using machine learning and deep learning: A review,” *Journal of Biomedical Informatics*, vol. 103, Art. no. 103362, 2020.

[4]. S. Sivaranjini and C. M. Sujatha, “Deep learning based diagnosis of Parkinson’s disease using convolutional neural network,” *Multimedia Tools and Applications*, vol. 79, nos. 21–22, pp. 15467–15479, 2020.

[5]. L. Breiman, “Random forests,” *Machine Learning*, vol. 45, no. 1, pp. 5–32, Oct. 2001.

[6]. M. R. Little, P. E. McSharry, E. J. Hunter, J. Spielman, and L. O. Ramig, "Suitability of dysphonia measurements for telemonitoring of Parkinson's disease," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 4, pp. 1015–1022, Apr. 2009.

[7]. U. R. Acharya, S. L. Fernandes, J. E. WeiKoh, H. Fujita, V. K. Sudarshan, S. Bhat, and U. R. Chowriappa, "Automated detection of Parkinson's disease using deep convolutional neural networks with transfer learning," *Knowledge-Based Systems*, vol. 201, Art. no. 106114, 2020.

[8]. T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, 2nd ed. New York, NY, USA: Springer, 2009.

[9]. S. Sivaranjini and C. M. Sujatha, "Deep learning-based diagnosis of Parkinson's disease using convolutional neural networks," *Multimedia Tools and Applications*, vol. 79, nos. 21–22, pp. 15467–15479, 2020.

[10]. N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002.

[11]. A. Mansoor and M. Ansari, "Artificial intelligence-based framework for early detection of mental health crises using social media data," *IEEE Access*, vol. 12, pp. 45231–45245, 2024.

[12]. P. Gaius, R. Kumar, and S. Mehta, "Stacking ensemble learning with language-model embeddings for depression detection," *Expert Systems with Applications*, vol. 223, Art. no. 119876, 2023.

[13]. S. Kuttala, A. Reddy, and P. Kumar, "Hierarchical autoencoder-based stress estimation using frequency-domain physiological features," *Biomedical Signal*

https://doi.org/10.63665/IJMEC.1104.16
ISSN: 2456-4265
IJMEC 2026

Processing and Control, vol. 84, Art. no. 104836, 2023.

[14]. E. Fix and J. L. Hodges Jr., "Discriminatory analysis—Nonparametric discrimination: Consistency properties," *USAF School of Aviation Medicine, Technical Report* 4, Randolph Field, TX, USA, 1951.

[15]. C. G. Goetz *et al*., "Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): Scale presentation and clinimetric testing results," *Movement Disorders*, vol. 23, no. 15, pp. 2129–2170, Nov. 2008.

[16]. D. J. Brooks, "Imaging dopamine transporters in Parkinson's disease," *Biomarkers in Medicine*, vol. 4, no. 5, pp. 651–660, 2010.

[17]. W. Poewe, K. Seppi, C. M. Tanner, G. M. Halliday, P. Brundin, J. Volkmann, A. E. Schrag, and A. E. Lang, "Parkinson disease," *Nature Reviews Disease Primers*, vol. 3, Art. no. 17013, 2017.

[18]. A. Tsanas, M. A. Little, P. E. McSharry, and L. O. Ramig, "Accurate telemonitoring of Parkinson's disease progression by non-invasive speech tests," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 4, pp. 884–893, Apr. 2010.

[19]. M. R. Little, P. E. McSharry, E. J. Hunter, J. Spielman, and L. O. Ramig, "Suitability of dysphonia measurements for telemonitoring of Parkinson's disease," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 4, pp. 1015–1022, Apr. 2009.

[20]. I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016

[21]. T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining (KDD)*, San Francisco, CA, USA, 2016, pp. 785–794.

[22]. F. Pedregosa *et al*., "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[23]. N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002.

[24]. A. Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*, 3rd ed. Sebastopol, CA, USA: O'Reilly Media, 2022.

[25]. M. Grinberg, *Flask Web Development: Developing Web Applications with Python*, 2nd ed. Sebastopol, CA, USA: O'Reilly Media, 2018.

[26]. M. R. Little, P. E. McSharry, E. J. Hunter, J. Spielman, and L. O. Ramig, "Suitability of dysphonia measurements for telemonitoring of Parkinson's disease," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 4, pp. 1015–1022, Apr. 2009.

[27]. N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002.

[28]. T. M. Cover and P. E. Hart, "Nearest neighbor pattern classification," *IEEE Transactions on Information Theory*, vol. IT-13, no. 1, pp. 21–27, Jan. 1967.

[29]. A. Tsanas, M. A. Little, P. E. McSharry, and L. O. Ramig, "Accurate telemonitoring of Parkinson's disease progression by non-invasive speech tests," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 4, pp. 884–893, Apr. 2010.

[30]. T. Fawcett, "An introduction to ROC analysis," *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861–874, Jun. 2006.

[31]. D. Beazley and B. K. Jones, *Python Cookbook*, 3rd ed. Sebastopol, CA, USA: O'Reilly Media, 2013.

https://doi.org/10.63665/IJMEC.1104.16
ISSN: 2456-4265
IJMEC 2026