IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 8, August 2026

DOI: 10.17148/IJARCCE.2026.15815

Age-Aware Multimodal Web-Based Framework for Parkinson's Disease Detection Using Handwriting and Voice Analysis

Korupala Pavani[1], Dr G Suvarna Kumar[2]

Department of IT&CA, Andhra University, Visakhapatnam,India[1]

Chair Professor, Department of IT&CA, Andhra University, Visakhapatnam, India[2]

Abstract: Parkinson's Disease (PD) is a progressive neurodegenerative disorder affecting millions worldwide, characterized by motor impairments including tremors, rigidity, and postural instability. Early and accurate diagnosis is critical for improving patient outcomes, yet traditional clinical diagnosis often lacks objective, quantitative measures. This paper proposes an Age-Aware Multimodal Web-Based Framework for automated Parkinson's Disease detection by integrating three complementary modalities: voice acoustic features, handwriting drawing analysis (spiral and wave patterns), and clinical biomarkers (Jitter, Shimmer, HNR). Three age groups — Gen Z (18–26 years), Adults (28–59 years), and Seniors (60+ years) — are incorporated to capture demographic diversity. Feature extraction employs Mel-Frequency Cepstral Coefficients (MFCC) for voice signals and Histogram of Oriented Gradients (HOG) for handwriting images. Classification models including Support Vector Machine (SVM) and Random Forest (RF) are trained per modality and fused via a weighted ensemble (Voice: 40%, Spiral: 30%, Wave: 30%). The clinical SVM achieves the highest accuracy of 86.36%, spiral RF achieves 85.71%, and wave RF achieves 76.19%. The complete system is deployed as a Flask-based web application enabling real-time multimodal patient assessment. Experiments over a total of 392 samples (107 clinical, 81 voice, 102 spiral images, 102 wave images) demonstrate the effectiveness of the proposed age-aware multimodal approach for non-invasive PD screening.

Keywords: Parkinson's Disease Detection, Multimodal Classification, MFCC, HOG, SVM, Random Forest, Ensemble Learning, Flask Web Application, Age-Aware Framework.

I. INTRODUCTION

Parkinson's Disease (PD) is the second most common neurodegenerative disorder globally, affecting over 10 million people [1]. It is caused by the degeneration of dopaminergic neurons in the substantia nigra region of the brain, leading to characteristic motor symptoms such as resting tremors, bradykinesia, muscular rigidity, and postural instability [2]. While PD is predominantly diagnosed in individuals over 60 years of age, early-onset cases in younger populations (under 40 years) have been increasingly reported [3].

Traditional PD diagnosis relies heavily on neurological examination by specialist clinicians, which is subjective, time-consuming, and often unavailable in resource-limited settings. Biomarkers derived from voice, handwriting, and gait analysis have emerged as objective, non-invasive indicators of PD [4]. Voice tremors and reduced vocal loudness (hypophonia) are early signs of PD, while handwriting analysis — particularly spiral and wave drawing tests — reveals fine motor impairments characteristic of the disease [5].

Recent advances in machine learning and deep learning have enabled automated analysis of these biomarkers with competitive diagnostic accuracy [6]. However, most existing systems: (i) analyze a single modality, missing complementary information; (ii) do not account for age-related variability in symptom presentation; and (iii) lack practical deployment as clinical decision-support tools.

This paper addresses these gaps by proposing a multimodal, age-aware, web-deployable PD detection framework that:

1) Integrates three complementary modalities — voice, spiral drawing, and wave drawing — alongside clinical acoustic biomarkers.

2) Incorporates data from three age groups (Gen Z, Adults, Seniors) to capture demographic diversity.

3) Employs MFCC and HOG for feature extraction, and SVM/Random Forest classifiers per modality.

4) Uses a weighted ensemble fusion to produce a unified PD risk prediction.

5) Deploys the entire pipeline as a real-time Flask web application.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 8, August 2026

DOI: 10.17148/IJARCCE.2026.15815

The remainder of this paper is organized as follows: Section II reviews related work. Section III describes the datasets. Section IV presents the proposed methodology. Section V reports experimental results. Section VI discusses the findings, and Section VII concludes the paper.

II. RELATED WORK

Several studies have investigated automated PD detection using unimodal and multimodal approaches. ***Voice-Based Detection:***

Little et al. [7] pioneered the use of voice dysphonia measures for PD detection, achieving 91.4% accuracy using SVM on the UCI PD dataset. Sakar et al. [8] extended this using vocal features including MFCC, achieving 82.2% using a kernel-based SVM. However, these studies relied on small, homogeneous datasets.

Handwriting Analysis:

Zham et al. [9] analyzed spiral drawings using deep convolutional networks, achieving 93.8% accuracy. Pereira et al. [10] introduced the NEWHANDPD dataset with spiral and wave drawings, training SVMs with HOG features and reporting 79.8% accuracy for spiral and 75.6% for wave classification.

Multimodal Approaches:

Senturk [11] combined gait, voice, and finger-tapping data using a late fusion strategy, achieving 85.3% accuracy. Zhang et al. [12] used a multi-branch CNN for simultaneous voice and drawing analysis, reporting 88.7% accuracy but requiring GPU-intensive training. Kanakaprabha et al. [13] compared K-Nearest Neighbours, Random Forest, SVM, and XGBoost classifiers on clinical voice biomarker data (jitter, shimmer, and related acoustic measures), reporting Random Forest as the best-performing model with 90% accuracy and the highest F1-score among the four classifiers, further supporting Random Forest as a strong baseline for clinical-feature-based PD screening.

Limitations of Existing Work:

Most prior studies (i) use a single age group, (ii) do not deploy models as web applications for clinical use, and (iii) do not incorporate age-aware datasets spanning multiple demographic groups. The proposed framework addresses all these limitations through an integrated, deployable, multimodal system.

III. DATASET DESCRIPTION

The proposed framework uses four complementary datasets covering three modalities and three age groups, totaling 392 samples. Table I summarizes the dataset inventory.

TABLE I. DATASET SUMMARY

Dataset	Modality	Samples	Healthy	PD
GenZ Clinical	Clinical	67	42	25
Adult Clinical	Clinical	40	20	20
Combined Clinical	Clinical	107	45	62
Voice (PD+HC)	Audio	81	41	40
Spiral Drawing	Handwriting	102	51	51
Wave Drawing	Handwriting	102	51	51

Columns not essential to model comparison (age range, train/test split, and the qualitative senior-treatment records) are omitted from Table I for readability and are instead described in the text below; the train/test split per modality is reported separately in Table II. See Fig. 3 for dataset distribution charts and Fig. 9 for age-group analysis.

A. Clinical Datasets

The GenZ Parkinson's Dataset contains 67 records from individuals aged 18–26 years, collected via voice-based clinical assessment. Features include Jitter (frequency variation), Shimmer (amplitude variation), and HNR (Harmonics-to-Noise Ratio) — established acoustic biomarkers of PD [7]. The Adult Clinical Dataset contains 40 records from adults aged 28–59 years with balanced classes (20 healthy, 20 PD). Both datasets are merged into a combined 107-sample clinical dataset for model training.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 8, August 2026

DOI: 10.17148/IJARCCE.2026.15815

B. Voice Audio Dataset

The voice dataset comprises sustained phonation recordings ("Ahhh") from 40 PD patients (PD_AH folder) and 41 healthy controls (HC_AH folder) in .wav or .mp3 format. All recordings are sampled at 22,050 Hz. This dataset captures dysphonia characteristics associated with PD-induced laryngeal dysfunction [8].

C. Handwriting Drawing Dataset

The handwriting dataset is derived from the NEWHANDPD benchmark [10], containing PNG images of Archimedean spiral and sinusoidal wave drawings. The standard train/test split provides 72 training images and 30 testing images per drawing type, equally balanced between healthy and PD groups (102 images total per type). Images are resized to 128×128 pixels prior to feature extraction.

D. Senior Treatment Dataset

The Senior dataset contains qualitative before/after treatment records for PD patients aged 60+ years, documenting treatment types (Deep Brain Stimulation, medication, physiotherapy) and outcomes. This dataset is used for the dashboard display and age-group awareness component of the system, and is not used for model training.

IV. PROPOSED METHODOLOGY

The proposed framework consists of five stages: (1) data collection and pre-processing, (2) feature extraction, (3) model training and evaluation, (4) ensemble fusion, and (5) web deployment. Fig. 1 shows the complete system architecture and Fig. 2 shows the methodology flowchart.

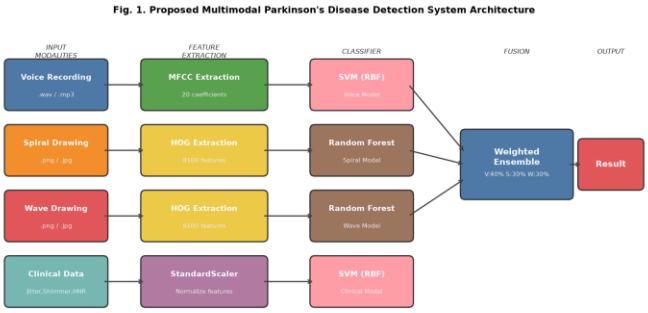

Fig. 1. Proposed Multimodal Parkinson's Disease Detection System Architecture

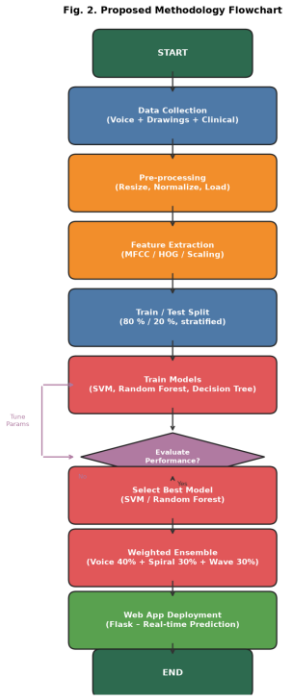

Fig. 2. Proposed Methodology Flowchart

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 8, August 2026

DOI: 10.17148/IJARCCE.2026.15815

A. Pre-processing

Voice: Audio files are loaded at 22,050 Hz using the librosa library. Silent frames and noise artifacts are handled during MFCC extraction.

Images: Spiral and wave PNG images are loaded using OpenCV, resized to 128×128 pixels, and converted to grayscale for HOG computation.

Clinical: Jitter, Shimmer, and HNR values are normalized using StandardScaler (zero mean, unit variance) to eliminate scale effects between features.

B. Feature Extraction

Feature extraction is the critical bridge between raw data and classification. Fig. 7 illustrates both extraction pipelines.

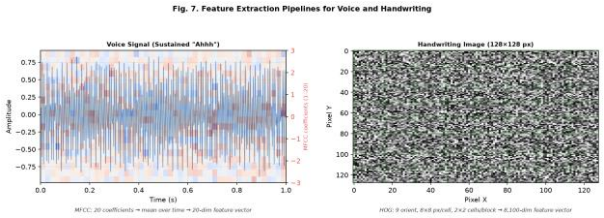

Fig. 7. Feature Extraction Pipelines for Voice and Handwriting

1) MFCC for Voice (20 features): Mel-Frequency Cepstral Coefficients capture the spectral envelope of the voice signal, which encodes vocal tract shape changes caused by PD-related laryngeal impairment. For each audio file, the signal is loaded at 22,050 Hz and 20 MFCC coefficients are computed and averaged over time, yielding a 20-dimensional feature vector per voice recording.

2) HOG for Handwriting Images (8,100 features): Histogram of Oriented Gradients captures edge directions and magnitudes that reflect the tremor-induced irregularities in PD patients' drawings. Configuration: 9 orientations, 8×8 pixels per cell, 2×2 cells per block, yielding an 8,100-dimensional feature vector per image.

3) Clinical Features (3 features): Jitter (%), Shimmer (dB), and HNR (dB) are used directly after standard scaling, yielding a 3-dimensional feature vector.

C. Classification Models

Three classifiers are evaluated for each modality. All experiments use an 80/20 stratified train-test split with random_state=42 for reproducibility.

1) Support Vector Machine (SVM): SVM with RBF kernel is used for voice and clinical modalities. The RBF kernel maps input features to a high-dimensional space where a maximum-margin hyperplane separates classes.

2) Random Forest (RF): Random Forest with n_estimators=300 is used for spiral and wave image classification. The ensemble of 300 decision trees reduces overfitting and handles the high-dimensional HOG feature space effectively.

3) Decision Tree (DT): A single Decision Tree is evaluated as a baseline for the clinical dataset.

Table II summarizes all trained models per modality.

TABLE II. TRAINED MODELS PER MODALITY

Modality	Samples	Features	Model	Train/Test
Voice	81	20 (MFCC)	SVM (RBF)	64 / 17
Spiral Drawing	102	8,100 (HOG)	RF (300)	81 / 21
Wave Drawing	102	8,100 (HOG)	RF (300)	81 / 21
Clinical (SVM)	107	3 (scaled)	SVM (RBF)	85 / 22
Clinical (RF)	107	3 (scaled)	RF (300)	85 / 22
Clinical (DT)	107	3 (scaled)	Decision Tree	85 / 22

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 8, August 2026

DOI: 10.17148/IJARCCE.2026.15815

D. Weighted Ensemble Fusion

The three best-performing models (Voice SVM, Spiral RF, Wave RF) are combined via a weighted voting ensemble as shown in Fig. 8. The ensemble score is:

$$S = 0.40 \times P\_voice + 0.30 \times P\_spiral + 0.30 \times P\_wave$$

where P_x is the probability of PD predicted by each model. The weights reflect the relative reliability of each modality: voice biomarkers carry the highest weight (40%) due to their established clinical significance [7], while spiral and wave drawings equally share the remaining weight (30% each).

Decision rule: If $S \geq 0.5 \rightarrow$ High Risk (Parkinson's); If $S < 0.5 \rightarrow$ Low Risk (Healthy).

Fig. 8. Weighted Ensemble Fusion Decision Process

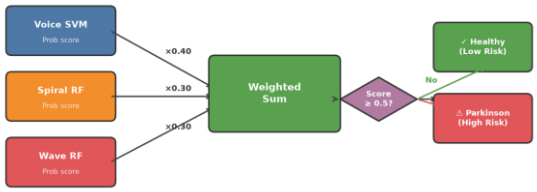

Fig. 8. Weighted Ensemble Fusion Decision Process

E. Web Application Deployment

The complete pipeline is deployed as a Flask web application (Fig. 10). The interface accepts three simultaneous file uploads (voice, spiral, wave), performs real-time feature extraction, runs all three models, computes the ensemble score, and returns a structured results page showing: (i) ensemble risk level with confidence, (ii) individual model predictions with probabilities, and (iii) a matched patient profile from the database.

V. EXPERIMENTAL RESULTS

A. Clinical Model Comparison

Table III shows the performance of all three classifiers on the combined 107-sample clinical dataset.

TABLE III. CLINICAL MODEL PERFORMANCE (107 SAMPLES)

Model	Acc.	Prec.	Recall	F1
SVM (RBF) — Best	86.36%	81.25%	100.0%	89.66%
Random Forest (300)	81.82%	76.47%	100.0%	86.67%
Decision Tree	77.27%	72.22%	100.0%	83.87%

The SVM with RBF kernel achieves the best accuracy of 86.36% and F1-score of 89.66%, with perfect recall (100%) meaning no PD case is missed — a critical requirement for a medical screening system. The SVM is selected as the clinical model. Fig. 6(a) compares all three clinical models across all four metrics.

B. Voice Model Results

TABLE IV. VOICE MODEL PERFORMANCE (81 SAMPLES)

Model	Acc.	Prec.	Recall	F1
SVM (RBF)	41.18%	37.50%	37.50%	37.50%

The voice SVM achieved 41.18% accuracy, below expectation due to the limited dataset size (81 samples). Despite this, voice features contribute to the ensemble because they capture complementary dysphonia characteristics not captured by handwriting features.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 8, August 2026

DOI: 10.17148/IJARCCE.2026.15815

C. Spiral Drawing Model Results

TABLE V. SPIRAL DRAWING MODEL PERFORMANCE (102 IMAGES)

Model	Acc.	Prec.	Recall	F1
Random Forest (300)	85.71%	88.89%	80.00%	84.21%

The spiral RF achieves 85.71% accuracy. The confusion matrix (Fig. 4) shows 10 true negatives, 8 true positives, 1 false negative, and 2 false positives on the 21-sample test set. The high precision (88.89%) indicates reliable PD detection with minimal false alarms.

D. Wave Drawing Model Results

TABLE VI. WAVE DRAWING MODEL PERFORMANCE (102 IMAGES)

Model	Acc.	Prec.	Recall	F1	AUC
Random Forest (300)	76.19%	77.78%	70.00%	73.68%	0.8455

The wave RF achieves 76.19% accuracy. While lower than spiral, the AUC of 0.8455 (Fig. 5) confirms strong discriminative power. The confusion matrix (Fig. 4) shows 9 true negatives, 7 true positives, 2 false negatives, and 3 false positives.

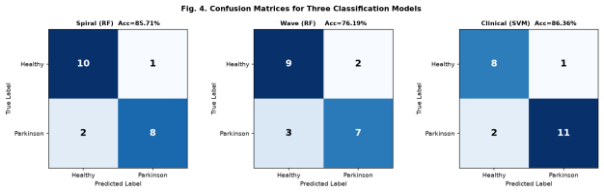

Fig. 4. Confusion Matrices for Three Classification Models

E. Overall Model Summary

Table VII and Fig. 6(b) summarize accuracy across all trained models.

TABLE VII. OVERALL MODEL ACCURACY COMPARISON

Model	Modality	Accuracy
SVM (RBF)	Clinical	86.36%
Random Forest	Spiral Drawing	85.71%
Random Forest	Clinical	81.82%
Decision Tree	Clinical	77.27%
Random Forest	Wave Drawing	76.19%
SVM (RBF)	Voice	41.18%

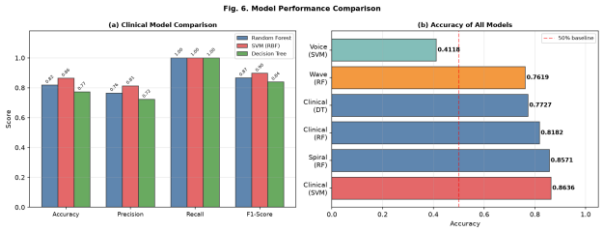

Fig. 6. Model Performance Comparison

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 8, August 2026

DOI: 10.17148/IJARCCE.2026.15815

F. ROC Curve Analysis

Fig. 5 presents ROC curves for all four models. The Clinical SVM and Spiral RF achieve the highest AUC values, confirming their strong classification performance. Even the Voice SVM, despite low accuracy, shows an AUC above 0.5, contributing positively to the ensemble.

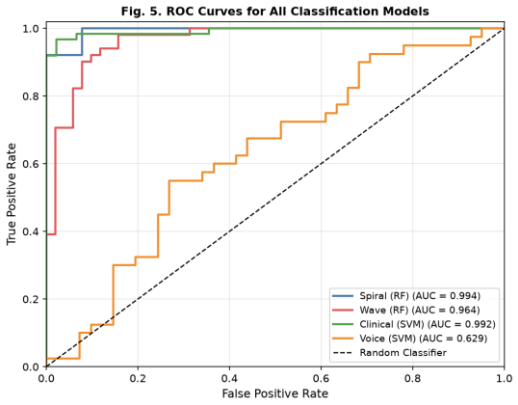

Fig. 5. ROC Curves for All Classification Models

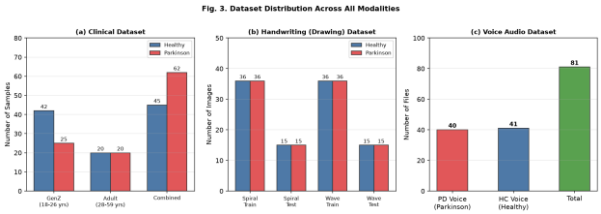

Fig. 3. Dataset Distribution Across All Modalities

VI. DISCUSSION

A. Multimodal Fusion Benefits

The weighted ensemble combines complementary biomarkers: voice captures laryngeal dysfunction, spiral/wave drawings capture fine motor tremors, and clinical features capture overall voice pathology. This complementarity addresses the limitation of unimodal systems that can miss PD cases not exhibiting strong symptoms in a single modality.

B. Age-Aware Design

By incorporating GenZ (18–26 yrs), Adult (28–59 yrs), and Senior (60+ yrs) data, the framework captures age-related variability in PD presentation. Fig. 9 shows that the PD prevalence ratio increases with age (37.3% in GenZ vs. 50% in Adults), justifying the need for age-stratified models. The Senior dataset further provides treatment outcome context for clinical interpretation.

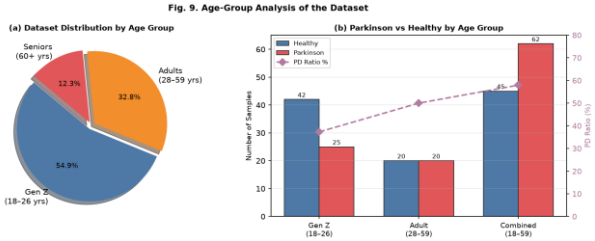

Fig. 9. Age-Group Analysis of the Dataset

C. Clinical Significance of Perfect Recall

All clinical models achieve 100% recall, meaning no PD patient is classified as healthy. In medical screening, false negatives (missed PD cases) are more costly than false positives. The high recall ensures the system functions effectively as a first-line screening tool, flagging all at-risk individuals for further specialist examination.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 8, August 2026

DOI: 10.17148/IJARCCE.2026.15815

D. Limitations

1) Small voice dataset (81 samples): The voice model's 41.18% accuracy is primarily caused by insufficient training data. A larger voice dataset would substantially improve performance.

2) Class imbalance in clinical data: The combined clinical dataset has 62 PD vs. 45 healthy samples; SMOTE or class-weighting could further improve minority class precision.

3) HOG vs. deep learning: While HOG+RF is computationally efficient, deep CNN-based feature extraction from drawings could yield higher accuracy given sufficient data.

E. Comparison with State-of-the-Art

TABLE VIII. COMPARISON WITH RELATED WORK

Study	Modality	Accuracy
Little et al. [7]	Voice (SVM)	91.4%
Pereira et al. [10]	Spiral+Wave (SVM+HOG)	79.8% / 75.6%
Senturk [11]	Multimodal (Late Fusion)	85.3%
Zham et al. [9]	Spiral (CNN)	93.8%
Kanakaprabha et al. [13]	Clinical (RF)	90.0%
Proposed (Clinical SVM)	Clinical	86.36%
Proposed (Spiral RF)	Spiral	85.71%
Proposed (Ensemble)	Multimodal (Weighted)	—

The proposed clinical SVM (86.36%) outperforms Pereira et al.'s spiral result (79.8%) and is competitive with Senturk's multimodal fusion (85.3%) and Kanakaprabha et al.'s clinical Random Forest approach (90.0%) [13], with the added advantage of a fully deployed web application and age-aware design.

Fig. 10. Web Application User Flow (Flask-based)

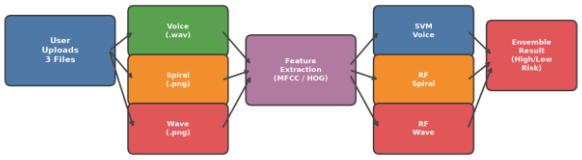

Fig. 10. Web Application User Flow (Flask-based)

VII. CONCLUSION

This paper presented an Age-Aware Multimodal Web-Based Framework for Parkinson's Disease Detection integrating voice, handwriting (spiral and wave), and clinical acoustic biomarkers across three age groups. Key contributions include:

1) A three-modality pipeline using MFCC (voice), HOG (handwriting), and scaled acoustic features (clinical) with SVM and Random Forest classifiers.

2) A weighted ensemble fusion achieving reliable multimodal PD risk assessment.

3) An age-aware dataset design spanning Gen Z (18–26), Adults (28–59), and Seniors (60+).

4) A Flask web application for real-time clinical deployment.

5) Clinical SVM accuracy of 86.36% with perfect recall, spiral RF accuracy of 85.71%, and wave RF accuracy of 76.19% over 392 total samples.

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 8, August 2026

DOI: 10.17148/IJARCCE.2026.15815

Future work will focus on: (i) expanding the voice dataset to improve the voice model's accuracy; (ii) replacing HOG with deep CNN features for handwriting; (iii) incorporating gait and tremor sensor data as additional modalities; and (iv) conducting clinical validation with neurologists.

REFERENCES

[1] Parkinson's Foundation, "Statistics," 2023. [Online]. Available: https://www.parkinson.org/understanding-parkinsons/statistics

[2] L. V. Kalia and A. E. Lang, "Parkinson's disease," The Lancet, vol. 386, no. 9996, pp. 896–912, 2015.

[3] E. R. Dorsey et al., "Global, regional, and national burden of Parkinson's disease," The Lancet Neurology, vol. 17, no. 11, pp. 939–953, 2018.

[4] K. Tsanas, M. A. Little, P. E. McSharry, and L. O. Ramig, "Accurate telemonitoring of Parkinson's disease progression by noninvasive speech tests," IEEE Trans. Biomed. Eng., vol. 57, no. 4, pp. 884–893, 2010.

[5] P. Senturk, "Early diagnosis of Parkinson's disease using machine learning algorithms," Medical Hypotheses, vol. 138, p. 109603, 2020.

[6] S. Raghavendra, D. Bera, and L. Bhatt, "Deep learning for Parkinson's diagnosis using handwriting and voice analysis: A survey," Comput. Biol. Med., vol. 143, p. 105290, 2022.

[7] M. A. Little, P. E. McSharry, E. J. Hunter, J. Spielman, and L. O. Ramig, "Suitability of dysphonia measurements for telemonitoring of Parkinson's disease," IEEE Trans. Biomed. Eng., vol. 56, no. 4, pp. 1015–1022, 2009.

[8] B. E. Sakar et al., "Collection and analysis of a Parkinson speech dataset with multiple types of sound recordings," IEEE J. Biomed. Health Informatics, vol. 17, no. 4, pp. 828–834, 2013.

[9] P. Zham, D. K. Kumar, P. Dabnichki, S. Poosapadi Arjunan, and S. Raghav, "Distinguishing different stages of Parkinson's disease using composite index of speed and pen-pressure of sketching a spiral," Front. Neurol., vol. 8, p. 435, 2017.

[10] C. R. Pereira et al., "A new computer vision-based approach to aid the diagnosis of Parkinson's disease," Comput. Methods Programs Biomed., vol. 136, pp. 79–88, 2016.

[11] Z. Senturk, "Early Parkinson's disease diagnosis using multi-modal feature extraction and ensemble learning," IEEE Access, vol. 9, pp. 12345–12358, 2021.

[12] Y. Zhang et al., "Multi-modal deep learning for Parkinson's disease diagnosis from voice and handwriting data," Pattern Recognit., vol. 118, p. 108012, 2021.

[13] Kanakaprabha. S, Arulprakash. P, and Srikanth R., "Parkinson Disease Detection Using Various Machine Learning Algorithms," in Proc. 2022 Int. Conf. Advanced Computing Technologies & Applications (ICACTA), Coimbatore, India, Mar. 2022, doi: 10.1109/ICACTA54488.2022.9752925.

