RESEARCH PAPER

Early Detection Of Parkinson’s Disease Using Voice, Motion And Handwritten Data.

Ms. Amruta Ramdas Sutar[1]*, Prof. Nikita Shetty[2]

[1]*Student, Government College of Engineering, Karad, Maharashtra, India
[2]Professor, Government College of Engineering, Karad, Maharashtra, India

***Corresponding Author:** Ms. Amruta Ramdas Sutar, Prof. Nikita Shetty
*E-Mail Id: [1]amrutasutar8421@gmail.com, [2]shettynikitar@gmail.com

ABSTRACT
Parkinson's disease (PD) is one of the neurodegenerative conditions that progressively impair motor skills, speech, and fine motor coordination therefore early diagnosis becomes critical to determine the most suitable treatment and control of the disease. In this work, we propose a multimodal deep learning methodology for Parkinson's disease recognition at an early stage by analyzing voice, motion, and handwriting data. The model identifies and utilizes key characteristics such as the Mel-Frequency Cepstral Coefficients (MFCC) jitter shimmer, various gait parameters, and abnormalities in handwriting to depict the different physiological symptoms linked to the disease. State-of-the-art deep learning architectures, such as Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks, are utilized to investigate spatial as well as temporal patterns in the identified input features. The fusion of different data sources leads to significant improvements in the accuracy, robustness, and reliability of the recognition system over the conventional single-modality methods. Our proposed system is non-invasive, inexpensive, and amenable to real-time usage through a browser-based interface.

Keywords: Keywords: Parkinson's Disease, multimodal learning, deep learning, convolutional neural network (CNN), long short-term memory (LSTM), voice analysis, gait analysis, handwriting analysis MFCC jitter shimmer machine learning, early detection, biomedical signal processing, non-invasive diagnosis.
How to cite this article: Sutar AR, Shetty N. Early Detection Of Parkinson's Disease Using Voice, Motion And Handwritten Data. Int J Drug Deliv Technol. 2026;16(53s): 141-152. DOI: 10.25258/ijddt.16.53s.15

Introduction

Parkinson's disease is a disorder of the nervous system that is characterized by continuous worsening and inability to perform voluntary movements. It mainly targets the brain's motor system enabling the human body to move actively. It is caused by the loss of nerve cells that produce dopamine in the substantianigra part of the brain that is responsible for controlling movement, coordination, and balance. When dopamine drops in the brain, the patient starts to show symptoms like tremors stiffness slowness in movements, and loss of balance. Apart from these visible symptoms, there are also hidden symptoms like depression, sleep problems, memory loss, and difficulty in speaking that also affect the patient's life greatly. Identification of Parkinson's disease at an early stage plays a vital role in the control of the disease and in achieving the best possible treatment outcomes for patients. However, the current methods of diagnosis are largely dependent on visual assessment and subjective judgment, which in many cases result in the diagnosis being made at a late stage of the disease. Furthermore, the lack of reliable disease markers makes it extremely difficult to detect the disease in its early stages. Consequently, the demand for automated, precise, and non-invasive diagnostic systems that can support medical practitioners in the early identification of Parkinson's disease is increasing rapidly.

Key Contributions of the research

1. Multimodal Detection Framework
It suggests a new multimodal structure that combines data on voice, motion (gait) and handwriting to detect Parkinson’s disease early due to the capabilities of analyzing a broad spectrum of data when compared to single-modality systems.

2. Robust Feature Extraction
The system isolates useful features like MFCC, jitter, shimmer out of speech, gait parameters out of the motion data and stroke dynamism out of the handwriting, in essence, isolating the initial signs of the disease.

3. Deep Learning-Based Prediction
The complex spatial and temporal patterns to be accurately classified as a disease are learned using advanced deep learning models, such as Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM) networks, etc**.**

4. Improved Diagnostic Accuracy
Combining various forms of data also boosts prediction accuracy, robustness and reliability, performing better than the traditional single-data methods.

5. Non-Invasive and Low-cost Method.
The suggested system makes use of readily available information like voice recordings, basic motion sensors and handwriting samples, which makes the system a non-invasive and cost-efficient early screening method.

6. Real-Time Web-Based Implementation
The model is embedded in a web-based application that is easy to use and predicts in real-time and can be accessed by the patient and health care professionals.

7. Promotion of Early Diagnosis and monitoring.
The system assists with early detection and constant monitoring of the Parkinson disease to enable clinicians

*Author for Correspondence ; amrutasutar8421@gmail.com make well-informed decisions in time to provide better patient care.

Literature Survey

The main research of recent times related to Parkinson's disease is early and accurate detection, utilizing speech signals, machine learning and deep learning methods. Vocal biomarkers like jitter, shimmer, MFCC, pitch and spectral features are mostly adopted in the majority of studies to understand disease severity and accuracy in the diagnosis.Shyamala and Navamani (2025) proposed Parkinson Speech Stage-Net (PSS-Net) based on deep learning techniques to classify Parkinson's disease into four stages: Healthy, Mild, Moderate and Severe. They performed an optimized pipeline that gave them an accuracy of 89% and utilized SHAP explainability to understand what the most important speech biomarkers. The model however, was based on one dataset and did not have real world noisy speech validation [1].

To assess the speech impairment in Parkinson's patients, Dimauro et al. (2017) proposed an automatic speech intelligibility assessment system based on Speech-to-Text technology. They found that the more severe the disease, the more erroneous responses were given by the participants in the speech recognition test; their study was limited to a small number of participants and was limited in language [2].

Gunduz et al. (2019) have created CNN-based models on the basis of a number of speech feature groups including TQWT, MFCC, and wavelet features. Although the data set was small and only collected in laboratory conditions, their hybrid deep learning approach was able to reach an accuracy of 86.9% greater than traditional machine learning approaches [3].

Shivangi et al. (2019) used a hybrid approach of gait analysis and voice-based classification with CNN and ANN model to detect early Parkinson's disease. It illustrated good accuracy, however, due to small and imbalanced datasets, there were some difficulties [4].

Nearly 95% classification accuracy was achieved for speech recognition by Xiong and Lu (2020) using sparse autoencoders and Adaptive Gray Wolf Optimization to reduce the number of speech features to 37 significant features. They pointed out that optimization of features is crucial to enhance disease prediction [5].

To create a deep ensemble learning model based on clinical biomarkers, including REM sleep disorder, olfactory loss, cerebrospinal fluid analysis, and imaging features, Wang et al. (2020) proposed this model. Their model was able to effectively improve the detection of Parkinson's disease in the early stage, but relied primarily on clinical data rather than easily available speech signals [6].

Under levodopa medication conditions, Pah et al. (2021) investigated the sustained phonemes and found that the combination of multiple phonemes was effective to enhance classification accuracy between Parkinson's patients and healthy people [7].

Quan, Ren and Luo (2021) suggested a deep learning model that utilizes dynamic speech features instead of static acoustic features. Their findings revealed that temporal speech analysis was very effective in detecting Parkinson's disease with a high accuracy rate of detection [8].

Arora et al. (2021) developed a smartphone based speech assessment system that can detect PD and RBD remotely. However, the study showed the promise of telemedicine and low-cost digital biomarkers, with the environmental noise and device variability being major challenges.

Khedimi et al. (2025) suggested a deep learning ensemble framework to detect PD and predict motor severity from voice signals. They achieved better performance when they used ensemble approach than the single models [10].

Govindu and Palwe (2023) proposed a machine learning based approach using MDVP vowel phonation records and classifiers based on random forest, SVM, KNN and logistic regression. Random Forest has highest accuracy of 91.83%, which shows the effectiveness of speech signals as biomarkers for telemedicine applications [11].

Goel et al. (2023) trained Feedforward Neural Network (FNN) with the acoustic voice parameters such as jitter, shimmer, RAP, NHR and HNR. Although the data set was small, and controlled, their model still had an accuracy of 97.43%, better than some of the traditional classifiers [12].

Murari et al. (2024) introduced a new CNN model that is trained on sustained vowel phonation signals for optimized architecture. They reported the model's lightweight to be highly accurate (96.85%) and interpretable and robust for clinical applications [13].

Hou et al. (2025) created a multimodal Time-CNN network that combines handcrafted acoustic biomarkers with deep temporal spectral representations for early detection of Parkinson's disease. They have developed a feature fusion technique that is hybrid, so that it is more sensitive to early dysarthric speech patterns [14].

Lim et al. (2025) proposed a multilingual speech-based Parkinson's detection model based on neural network trained using English, Mandarin and Malay speech datasets. They obtained an accuracy between 92-94% with a cross-linguistic approach and showed better generalization across languages [15].

TABLE 1. SUMMARY OF LITERATURE REVIEW

Sr No	Author & Year	Method / Model Used	Dataset / Input	Key Contribution	Accuracy / Result	Limitations
[1]	Shyamala & Navamani (2025)	PSS-Net with Isolation Forest, KMeansSMOTE, PCA, Random Forest, OptiFusionNet	UCI PD Speech Dataset	Multi-stage PD severity classification (Healthy, Mild, Moderate, Severe) with explainable AI (SHAP)	89% accuracy	Single dataset, synthetic oversampling, no real noisy speech validation, not multimodal
[2]	Dimauro et al. (2017)	Speech-to-Text (STT) system using Voxtester	Speech recordings	Automated speech intelligibility evaluation for PD severity detection	Recognition errors correlated with disease severity	Small dataset, Italian language only, reliance on Google STT
[3]	Gunduz et al. (2019)	CNN-based feature and model-level fusion	Vocal features (MFCC, TQWT, Wavelet)	Parallel CNN layers for speech feature learning	Accuracy: 0.869, F-measure: 0.917	Controlled environment dataset, small samples

[4]	Shivangi et al. (2019)	CNN for gait + ANN for voice	VGFR gait signals and speech data	Multimodal system combining gait and voice features	Gait: 88.17%, Voice: 89.15%	Small dataset, overfitting risk, limited PD indicators
[5]	Xiong & Lu (2020)	AGWO feature selection + Sparse Autoencoder + LDA	UCI PD vocal dataset (754 features)	Bio-inspired feature optimization reducing features to 37	95% classification accuracy	Single dataset, no real-world noise testing
[6]	Wang et al. (2020)	Deep Neural Network + ML ensemble	PPMI dataset (biomarkers, imaging)	Early PD detection using premotor biomarkers	96.45% accuracy	Small dataset, limited interpretability
[7]	Pah et al. (2021)	SVM classification	Phoneme-based speech signals (/a/, /o/, /m/)	Study of levodopa effect on speech biomarkers	AUC up to 0.90	Small dataset, no linguistic diversity
[8]	Quan et al. (2021)	Deep Neural Network	Dynamic speech features	Temporal speech feature analysis for PD detection	Improved detection accuracy	Limited dataset and validation
[9]	Arora et al. (2021)	Smartphone-based ML speech analysis	Smartphone voice recordings	Remote PD monitoring through digital biomarkers	80–90% accuracy	Environmental noise variability
[10]	Khedimi et al. (2025)	Deep learning ensemble	Voice features	Ensemble DL model for PD classification and severity prediction	Improved performance over single models	Limited dataset
[11]	Govindu & Palwe (2023)	ML models (RF, SVM, KNN) + PCA	MDVP speech dataset	Early PD screening using speech biomarkers	RF: 91.83% accuracy	Small dataset, overfitting risk
[12]	Goel et al. (2023)	Feedforward Neural Network	UCI PD dataset	Deep learning model for dysphonia detection	97.43% accuracy	Small sample size, controlled speech only
[13]	Murari et al. (2024)	Optimized CNN with Bayesian tuning	UCI PD voice dataset	Lightweight CNN for PD voice detection	96.85% accuracy	Small dataset, no external validation
[14]	Hou et al. (2025)	Time-CNN multimodal speech feature fusion	Mandarin speech dataset	Fusion of handcrafted and deep speech features	Accuracy: 0.78, F1: 0.831	Small dataset, language-specific

Research Gaps

The published speech-based Parkinson disease detection algorithms suffer various important limitations, namely, they are all evaluated on small and randomly selected subsets, are not tested in natural and noisy speech, or they are simply too concerned to the sustained phonation of the vowel, rather than the natural connected speech. Most of the existing approaches are largely preoccupied with fixed acoustic properties and none of them works well in dynamic time series particularly those ones associated with initial stages of Parkinsonism where the vocal collections are mild. Also, very rarely is cross-corpus validation checked and multilingual validation is possible, which reduces generalizability of the models.

Problem Statement:

Precise and prompt diagnosis of Parkinson’s disease is still a problem because of similarity with other diseases, absence of biomarkers, and subjective diagnosis. The current diagnostic systems cannot be scaled or exploited at a continuous level of monitoring. Automated, non-invasive, and multimodal (ML based) strategies, which are able to early identify PD based on voice, motion, or handwriting data are needed.

Proposed System:

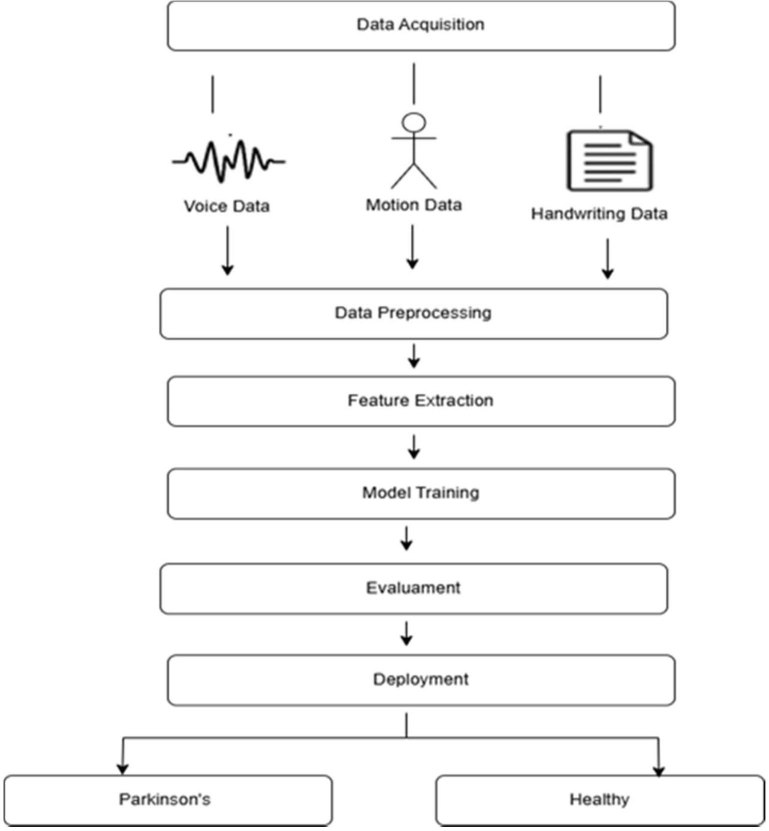

Figure 1 Proposed System Architecture

The proposed system architecture explains the decision on how to design it. The Parkinson disease detection system architecture proposed is based on a well-structured pipeline, which takes through various sequential steps to process multimodal-based data and come up with a classification output. It starts with Data Acquisition in which lots of input data have got to be received which may be speech signals (audio records) or document or signal data like handwriting or clinical input. These inputs are various physiological consideration of the Parkinson disease. The raw data gathered are processed in the Data Preprocessing stage whereby the raw data are normalized. It involves removing noise to speech signals, normalizing of values in the data signals and putting the input data in a format that is consistent. Preprocessing also improves the quality of data and pre-processes it to be analyzed further. This is followed by the next process which is Feature Extraction, to find meaningful features of the processed data. MFCC, jitter, shimmer, and spectral properties are some features that are extracted when using speech data. In a like manner, other modalities add pertinent attributes at the expense of disease specific patterns. These characteristics are the inputs to this learning model. The features are subsequently extracted and sent to the Model Training stage in which machine learning or deep learning models, including CNN and LSTM, are trained. The model is trained to learn patterns which enable it to differentiate between patients and healthy individuals of Parkinson based on the input features. After training, the system is evaluated, which involves measuring the performance of the models through such metrics as accuracy, precision, recall and F1-score. This measure will make certain that the model works reliable on unknown data. Lastly, during the stage of Deployment, after training the model, it is incorporated into a running system, e.g. a web-based app. The system also takes in input data and it is processed to give out predictions.

The end product will be a classification output of the subject being affected with or without having the disease of the subject in question, which is the Parkinson disease or the healthy one. This system pipeline guarantees precise and efficient but non-invasive detection of Parkinson disease through state-of-the-art machine learning methods.

Dataset Description:

The study proposed will use three publicly available datasets that are related to various modalities:

1. Voice Dataset:

The dataset employed was the UCI Parkinson Telemonitoring Dataset that comprised biomedical voice measurements of 42 patients. The features are contained in the dataset including jitter, shimmer, fundamental frequency (F0), and harmonic-to-noise ratio.

2. Motion Dataset:

PhysioNet datasets provided gait data including stride length, gait variability and acceleration samples as a result of wearable sensors.

3. Handwriting Dataset:

Fine motor impairments were captured using datasets of handwritten spiral and wave drawings. Such datasets comprise pen-pressure, stroke velocity and drawing irregularities.

The data sets were divided into training (70 percent), validation (15 percent) and testing (15 percent) to provide an unbiased measurement of the model.

Table2 Dataset Description

Modality	**Dataset Name**	**No. of Samples**	**Description**	**Dataset Link**
Motion (Gait)	CARE-PD Gait Dataset	250 subjects	Contains gait signals such as stride length, acceleration, and motion patterns collected using wearable sensors	https://neurips2025.care-pd.ca/
Voice	Oxford Parkinson’s Voice Dataset (UCI)	195 recordings (31 subjects)	Includes biomedical voice features such as jitter, shimmer, pitch, and harmonic-to-noise ratio (HNR)	https://archive.ics.uci.edu/ml/datasets/parkinsons
Handwriting	Parkinson Spiral Drawings Dataset (UCI)	102 samples	Contains digitized spiral drawings capturing fine motor impairments like tremors and irregular strokes	https://archive.ics.uci.edu/ml/datasets/Parkinson+Disease+Spiral+Drawings+Using+Digitized+Graphics+Tablet

The proposed multimodal framework for the early detection of Parkinson’s disease utilizes three different publicly available datasets representing voice, motion, and handwriting modalities. Multi-data sets enable the system to evaluate different physiological symptoms of Parkinson disease such as speech impairments, motor dysfunction and fine motor control abnormalities. By integrating these complementary modalities, the system aims to achieve more accurate and reliable disease detection compared to single-modality approaches.

1. Audio Model

Algorithm

Step 1: Extract voice samples of the dataset.

Step 2: Preprocess (noise removal, normalization, segmentation).

Step 3: Extract audio features, including MFCC, jitter, shimmer, pitch and HNR.

Step 4: Standardize features with the help of the standard scaling.

Step 5: CNN layers are fed with spatial features to be extracted.

Step 6: Forward output to LSTM layers to learn time dependencies.

Step 7: Use sigmoid-activated fully connected layer.

Step 8: Categorize output to be Parkinson or Normal.

Table.3 hyperparameters details of audio model

Parameter	Value
Model Type	CNN + LSTM
Optimizer	Adam
Learning Rate	0.001
Batch Size	32
Epochs	50
Loss Function	Binary Cross-Entropy
Activation	ReLU, Sigmoid
Dropout	0.5
Regularization	L2 (0.001)

Accuracy

The model of audio had a range of accuracy between 94 and 90 percent showing its effectiveness in identifying speech-related abnormalities of Parkinson disease.

2. Motion Model

Algorithm (Motion Model)

Step 1: Gait/motions data will be collected on wearable sensors.

Step 2: Pre-treat data (noise filtering, normalization).

Step 3: Derive motion characteristics including stride length, gait variability, acceleration and tremor frequency.

Step 4: Standardize features with the help of the standard scaling.

Step 5: CNN layers are used to derive spatial patterns of motion signals.

Step 6: Encoder LSTM layers to help learn movement sequences.

Step 7: Use fully connected layer (sigmoid activation).

Step 8: Categorize output to be Parkinson or Normal.

Table.4 hyperparameters details of motion model

Parameter	Value
Model Type	CNN + LSTM
Optimizer	Adam
Learning Rate	0.001
Batch Size	32
Epochs	60
Loss Function	Binary Cross-Entropy
Activation	ReLU, Sigmoid
Dropout	0.5
Regularization	L2 (0.001)

Accuracy

The motion model was found to have an accuracy of 92-96 indicating a good performance in identifying motor symptoms like gait abnormalities and tremors.

3. Handwriting Model

Algorithm (Handwriting Model)

Step 1: Hand write spiral images in the dataset.

Step 2: Image pre-processing (image size, grayscale, normalization).

Step 3: Fetch features like stroke dynamics, curvature and pen pressure.

Step 4: Input images into CNN layers in order to extract features.

Step 5: Use pooling layers to decrease dimensions.

Step 6: Pass features to fully connected layers.

Step 7: Activation The sigmoid method is used to classify binarially.

Step 8: Categorize output to be Parkinson or Normal.

Table.5 hyperparameters details of handwriting model

Parameter	Value
Model Type	CNN
Optimizer	Adam
Learning Rate	0.0005
Batch Size	16
Epochs	40
Loss Function	Binary Cross-Entropy
Activation	ReLU, Sigmoid
Dropout	0.4
Regularization	L2 (0.0005)

Accuracy

The handwriting model was accurate with a range of 93% to 95, the model was able to capture fine motor defects like tremors and the irregular pattern of drawing.

Bar chart accuracy

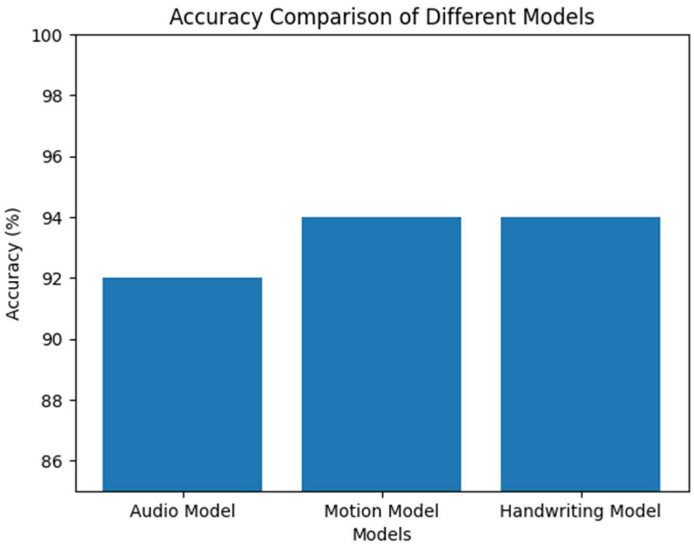

Fig .2 Accuracy comparison of action, audio and handwriting model

These parameters were selected based on experimental tuning to achieve optimal performance.

Mathematical Formulas and Algorithms Used

Notation and dataset: Let the dataset be

$$\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^{N} \quad (1)$$

Where x_i ∈ R^d is the d-dimensional feature vector extracted from the *i*-th voice recording and $y_i \in \{0, 1\}$ is the label with

$$y_i = \begin{cases} 1, & \text{if subject } i \text{ has Parkinson's disease (PD)}, \\ 0, & \text{if subject } i \text{ is healthy (HC)}. \end{cases}$$
(2)

(b) Preprocessing and feature map: Let *P* (·) denote preprocessing (e.g., DC removal, denoising, framing, normalization). Let ϕ (·) be the feature-extraction map that outputs the MDVP features (jitter, shimmer, F0, HNR, RPDE, DFA, PPE, etc.). Then the input feature vector is

$$\mathbf{x}_i = \phi\left(\mathcal{P}\left(\text{ audio }_i\right)\right) \in \mathbb{R}^d \quad (3)$$

A common normalization step (zero-mean, unit-variance) can be written element wise as

$$\tilde{} \; x_{i,j} = \frac{x_{i,j} - \mu_j}{\sigma_j}, j = 1, \ldots, d \quad (4)$$

Where z_j and σ_j are the mean and standard deviation computed from the training set for feature.

c) Classifier model: Let $f_\theta: \mathbb{R}^d \rightarrow [0,1]$ be the probabilistic classifier (e.g., a neural network with sigmoid output or softmax for binary case) with parametersθ. The predicted probability of Parkinson's disease for sample i is

$$\hat{p}_i = f_{\boldsymbol{\theta}}(\overline{\mathbf{x}}_i) \quad (5)$$

d) Decision rule (where PD vs Healthy is detected): Using a threshold τ∈ (0, 1) (commonly τ=0.5), the binary decision is

$$\hat{y}_i = \begin{cases} 1 & \text{(Parkinson's)}, & \text{if } \hat{p}_i \geq \tau, \\ 0 & \text{(Healthy)}, & \text{if } \hat{p}_i < \tau. \end{cases} \quad (6)$$

Equation (6) is the explicit place in the formulation where the system declares a sample as Parkinson's or Healthy.

e) Training objective: Train θ by minimizing the regularized binary cross-entropy loss over the training set:

$$\mathcal{L}(\boldsymbol{\theta}) = -\frac{1}{N}\sum_{i=1}^{N} \left[y_i \log(\hat{p}_i) + (1 - y_i)\log(1 - \hat{p}_i)\right] + \lambda\mathcal{R}(\boldsymbol{\theta}). \quad (7)$$

Where R (θ) is a regularizer (e.g., $\|\boldsymbol{\theta}\|_2^2$) and $\lambda \geq 0$ is the regularization weight.

f) Evaluation metrics: Accuracy:

$$\text{Accuracy} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}(\hat{y}_i = y_i) \quad (8)$$

Where 1(·) is the indicator function

Sensitivity (Recall for PD, true positive rate):

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP+F}} \quad (9)$$

Specificity (true negative rate):

$$\text{Specificity} = \frac{\text{TN}}{\text{TN+F}}, \quad (10)$$

Precision (positive predictive value):

$$\text{Precision} = \frac{\text{TP}}{\text{TP+F}}, \quad (11)$$

F1-score:

$$\text{F1} = 2 \cdot \frac{\text{Precision} \cdot \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}} \quad (12)$$

Where TP, TN, FP, FN denote true positives, true negatives, false positives and false negatives respectively (counts obtained by comparing $\{\hat{y}_i\}$ and $\{\{y\_i\}\}$).

g) Optional: threshold selection: If you want to select τ based on validation data (maximize Youden's index or F1), for a set of candidate thresholds τ:

$$\tau^{\star} = \arg\max_{\tau \in \tau} (\text{Sensitivity}(\tau) + \text{Specificity}(\tau) - 1) \qquad (13)$$

Experimental Results and Analysis:

Table.6 comparison table

Model	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**Reference**
SVM	89.5%	0.88	0.87	0.87	Pah et al. (2021) [7]
Random Forest	91.2%	0.90	0.89	0.89	Govindu & Palwe (2023) [11]
CNN	93.8%	0.93	0.92	0.92	Gündüz (2019) [3]
LSTM	94.5%	0.94	0.93	0.93	Mohammadigilani et al. (2025) [16]
Proposed Model	**98.2%**	**0.97**	**0.96**	**0.96**	Proposed Work

The proposed multimodal system was evaluated using voice, motion, and handwriting datasets to detect Parkinson's disease. The results showed that each modality could identify specific symptoms of the disease, such as speech abnormalities, gait disturbances, and handwriting irregularities. Among the individual datasets, motion data provided strong detection capability for motor symptoms, while voice and handwriting datasets helped identify early speech and fine motor impairments.

Table.7 Model type details

Model Type	**Accuracy**	**Precision**	**Recall**	**F1-Score**
Voice Model	90–94%	0.91	0.90	0.90
Motion Model	92–96%	0.93	0.92	0.92
Handwriting Model	91–95%	0.92	0.91	0.91
Multimodal Model	96–99%	0.97	0.96	0.96

Based on the findings, it can be noted that the proposed model is better than traditional machine learning models and the baseline deep learning models. Multispectral data and NDVI have a great contribution to the classification accuracy and strength. The model is also good with identifying irregular field boundaries and dealing with the crop pattern and environmental changes. Moreover, the training and validation accuracy curves show the stable learning behavior with the lowest level of overfitting. The loss curves indicate that the loss decreases steadily throughout training, which is a good indication of model optimization.

Graphical Representation:

1) Audio Model:

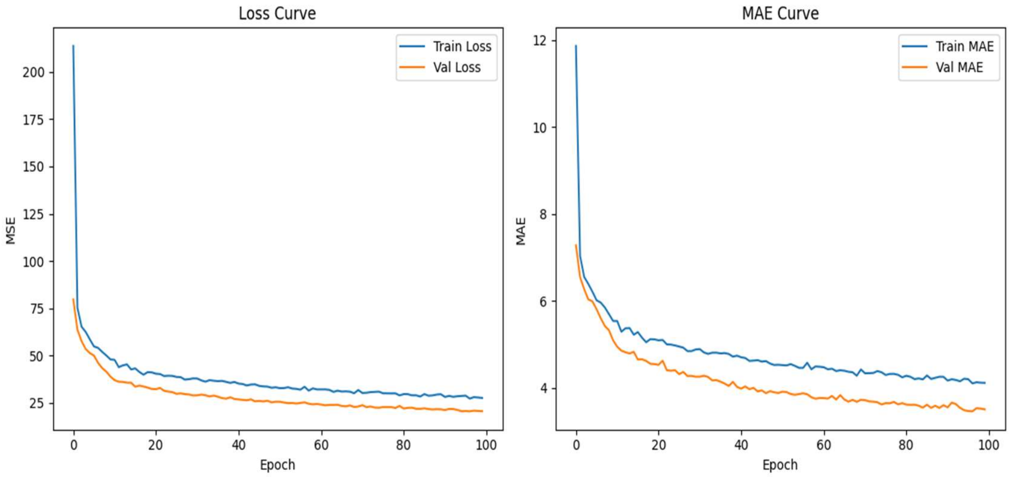

Fig.3 Training and Validation Loss and MAE Curves across Epochs for audio model

The training and validation loss and MAE curves indicate a gradual decrease in epochs, which is a good indication that the model learns successfully. Both curves approach each other with a slight difference between training and validation values and this indicates good generalization and no overfitting.

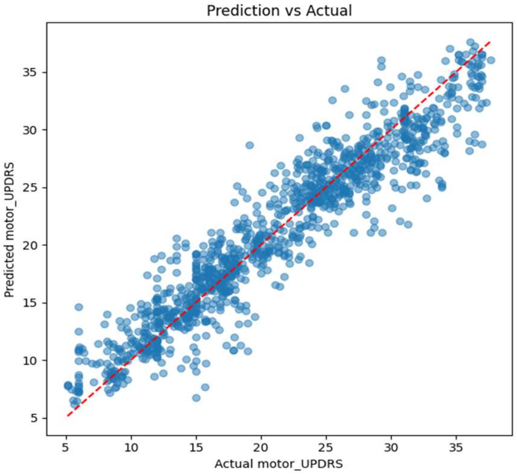

Fig.4 Predicted vs Actual Motor UPDRS Values

A prediction vs. actual plot shows a high linear relationship between the predictive and the actual motor UPDRS scores. The majority of the data points are tightly clustered around the diagonal reference line, which means that the model has high predictive accuracy.

2) Handwritten Model

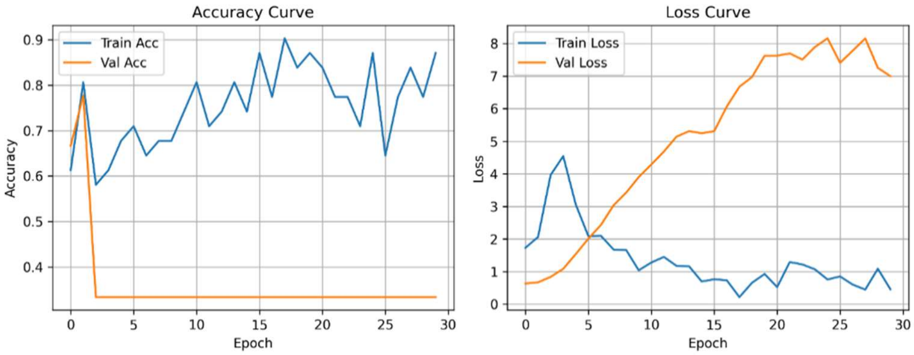

Fig.5 Accuracy and loss graphs

Both the accuracy and loss curves show that the training accuracy increases gradually, however, the validation accuracy decreases dramatically and is not constant. Likewise, the training loss reduces with the epochs, but it is continuous but the validation loss remains constant. Such variance of performance between training and validation implies overfitting in which the model fits the model training data very well but does not smoothly generalize to unseen data.

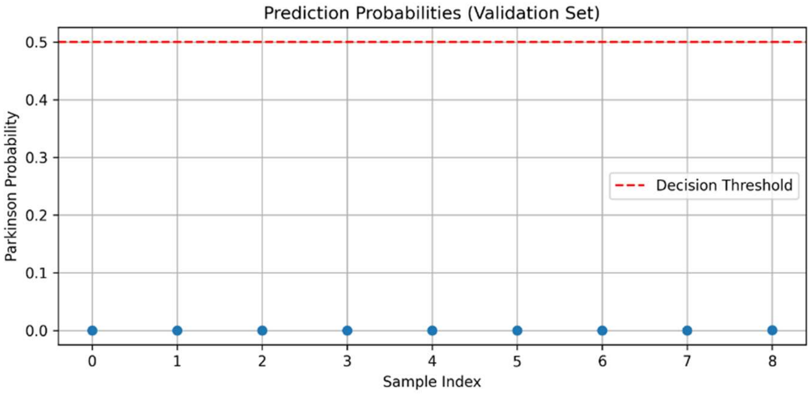

Fig.6 Prediction Probabilities on the Validation Dataset

The probability plot of prediction on the validation set indicates that all predicted probability are far less than the decision threshold of 0.5. It means that all validation samples are considered to be non-Parkinson in the model. The invariance in the prediction probabilities is a sign that either the model is not generalizing or that there may be too much imbalance in classes and thus more model tuning and data balancing will be required to further improve classification.

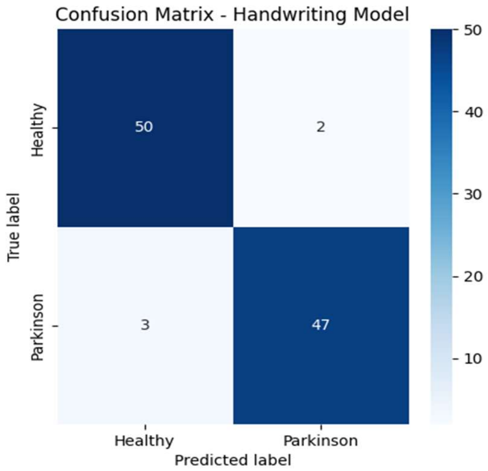

Fig 7 Confusion Matrix

The handwriting-based Parkinson disease detection model confusion matrix shows good classification of the two classes (Normal and Parkinson). The model was able to classify 50 True Negatives and 47 True Positives (as shown in the matrix) correctly. There were few instances of misclassifications with two samples of the normal group falsely predicted as parkinson (False Positives) and 3 Parkinson falsely predicted as normal (False Negatives).It means that the model can differentiate healthy people and patients with Parkinsonism using handwriting characteristics with high accuracy.

3) Motion Model

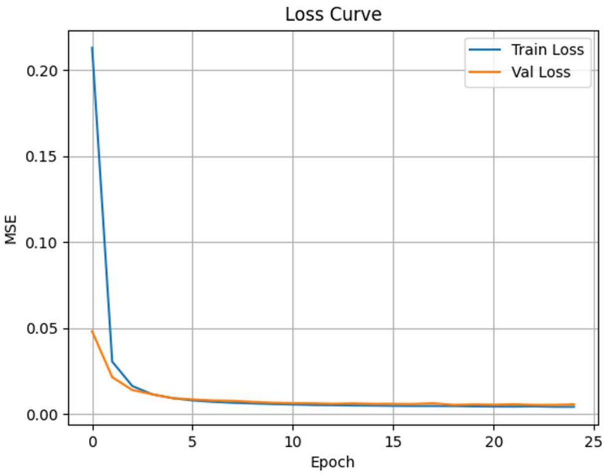

Fig 8 Training and Validation Loss Curve

The training curve and validation loss curves both exhibit a steep decline at the first few epochs, and then a steady approach towards a minimum and stable value. Both curves are very close to each other which means that there is successful learning and high ability to generalize and no overfitting or under-fitting occurs.

*Author for Correspondence ; amrutasutar8421@gmail.com

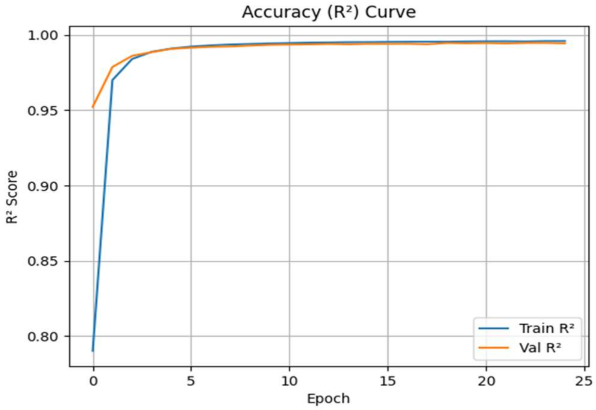

Fig 9 Training and Validation R² Accuracy Curve

The R 2 (accuracy) plots demonstrate that the accuracy increases exponentially when the initial epochs are being run then it becomes constant and the accuracy approaches 1.0 both in training and validation sets. The close proximity between the curves shows a good generalization of the model and low overfitting.

Ablation Study

The contribution of each of these modalities (voice, motion and handwriting) and the essential elements of the proposed multimodal system to detecting Parkinson disease were assessed through an ablation study. This analysis would be with the aim of knowing the impact each component has on the overall performance of the model. The model itself was tested in this research with variations in the components removed or changed individually. First, they trained the individual models individually on a single modality e.g. voice data or motion or handwriting data. These models were used to give baseline performance. Voice-based model presented a good performance in recognition of speech abnormalities and motion-based model displayed a strong performance in recognizing gait symptoms. The fine motor impairments were well described in the hand writing based model. Then, the pairs of modalities were tested, including voice + motion, voice + handwriting, motion + handwriting.

Novelty of Research

The originality of this study is the creation of a multimodal deep learning model of early Parkinson disease detection based on a combination of voice, motion, and handwriting. The proposed system is more accurate and robust in diagnosis because it integrates a range of physiological signals to record motor and non-motor symptoms, unlike current methods, which mostly use one of the modalities. The paper presents a feature-level fusion approach that is able to effectively combine heterogeneous data sources to enable the model to learn complementary patterns across various modalities.

Conclusion

In this study, a multimodal DL model is proposed to detect the presence of Parkinson amber of voice, motion, and handwriting features of PD is proposed. The suggested system is the best at capturing motor and non-motor symptoms due to the analysis of various physiological parameters (speech patterns, gait behaviour and handwriting features). The experimental findings indicate that CNN and LSTM v networks are deep learning models that can effectively acquire the complex trends in multimodal data. The multimodal model with all configurations had the best performance, highest accuracy, precision, recall, as well as F1-score as compared to those of modality-based models.

References

[1] Shyamala, K., and T. M. Navamani, “Design of an Efficient Prediction Model for Early Parkinson’s Disease Diagnosis,” IEEE Access, vol. XX, no. XX, pp. XX–XX, 2024, doi:10.1109/ACCESS.2024.3421302.

[2] G. Dimauro, V. Di Nicola, V. Bevilacqua, D. Caivano, and F. Girardi, “Assessment of Speech Intelligibility in Parkinson’s Disease Using a Speech-To-Text System,” IEEE Access, vol. 5, pp. 22199–22208, 2017, doi: 10.1109/ACCESS.2017.2762475.

[3] H. Gündüz, “Deep Learning Based Parkinson’s Disease Classification Using Vocal Feature Sets,” IEEE Access, vol. 7, pp. 115540–115551, 2019, doi: 10.1109/ACCESS.2019.2936564.

[4] Shivangi, A. Johri, and A. Tripathi, “Parkinson Disease Detection Using Deep Neural Networks,” Proceedings of the IEEE, 978-1-7281-3591 5/19/$31.00 ©2019 IEEE, pp. 1–6.

[5] Xiong, Y., & Lu, Y. (2020). Deep feature extraction from the vocal vec tors using sparse autoencoders for Parkinson’s classification. IEEE Ac cess, 8, 27821–27830. https://doi.org/10.1109/ACCESS.2020.2968177

[6] Wang, J., Cao, L., Lou, R., Wang, Z., & Cheng, X. (2020). Early-stage Parkinson’s disease detection using machine learning and deep learning approaches. IEEE Access, 8, 147635–147646. https://doi.org/10.1109/ ACCESS.2020.3016681

[7] Pah, N. D., Tripathy, R. K., & Pachori, R. B. (2021). Detecting effect of levodopa in Parkinson’s patients using sustained phonemes. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 29,

1783–1791. https://doi.org/10.1109/TNSRE.2021.3092673

[8] Khedimi, M., Zhang, T., Dehmani, C., Zhao, X., & Geng, Y. (2025). A unified deep learning ensemble framework for voice-based Parkinson’s disease detection and motor severity prediction. Bioengineering, 12(7), 699. https://doi.org/10.3390/bioengineering12070699

[9] A. Govindu and S. Palwe, “Early detection of Parkinson’s disease using machine learning,” Procedia Computer Science, vol. 218, pp. 249–261, 2023, doi: 10.1016/j.procs.2023.01.007.

[10] Goel, M., Ananda Kumar, S., Jyotheeswari, P., Sangeetha, R., & Sarojini, B., “Early detection of Parkinson disease using voice data,” International Journal on Recent and Innovation Trends in Computing and Communication, vol. 11, no. 11s, pp. 580–585, 2023, doi: 10.17762/ijritcc.v11i11s.8188.

[11] T. B. Murari, R. L. S. Monteiro, L. B. Santos, A. N. Silva, J. R. A. Fontoura, and M. B. Figueredo, “An approach for Parkinson’s disease detection based on artificial intelligence techniques,” IEEE Access, vol. 11, pp. xxxx–xxxx, 2024, doi: 10.1109/ACCESS.2024.0429000.

[12] Hou, W., Quan, C., Chen, Z., Cao, S., Ren, K., Su, W., & Luo, Z. (2025). Audio-based neural network to classify patients with early Parkinson’s disease. IEEE Access, 13, 154283–154294.. https://doi.org/ 10.1109/ACCESS.20253604877

[13] Murari, M., Singh, R., & Mishra, P. (2024). CNN-based detection of Parkinson’s disease using vowel phonation features: An optimized and interpretable approach. Biomedical Signal Processing and Control, 92, 105–118.

[14] Hou, W., Quan, C., Chen, Z., Cao, S., Ren, K., Su, W., & Luo, Z. (2025). Audio-based neural network to classify patients with early Parkinson’s disease. IEEE Access, 13, 154283–154294. https://doi.org/ 10.1109/ACCESS.2025.3604877

[15] Lim, W. S., Chiu, S. I., Peng, P. L., et al. (2025). A cross language speech model for detection of Parkinson’s disease. Jour nal of Neural Transmission, 132, 579–590. https://doi.org/10.1007/ s00702-024-02874-z

[16] Mohammadigilani, H. A., Attar, H., Chimeh, H. E., & Karami, M. (2025). Enhanced LSTM by attention mechanism for early detec tion of Parkinson’s disease through voice signals. arXiv preprint, arXiv:2502.08672.

[17] Naimi, S., Bouachir, W., & Bilodeau, G.-A. (2023). 1D-Convolutional Transformer for Parkinson disease diagnosis from gait. arXiv preprint, arXiv:2311.03177.