

Research Paper

A Comprehensive Framework for Parkinson's Disease Detection Using XG Boost Classifier and Data Driven Machine Learning Methods.

¹E. Ravindra Babu

¹M.Tech, Department of Computer Science and Engineering
DNR College of Engineering & Technology (Autonomous)
Bhimavaram – 534202, Andhra Pradesh, India
ravinderelipe@gmail.com

²Mr. Nandana Kumar

²Assistant Professor, Department of Computer Science and Engineering
DNR College of Engineering & Technology (Autonomous)
Bhimavaram – 534202, Andhra Pradesh, India
nandan@dnrcet.org

Abstract—Parkinson's disease (PD) is a progressive neurodegenerative disorder characterized by motor and non-motor manifestations that can make early and objective detection difficult. Conventional diagnosis primarily depends on clinical examination and longitudinal assessment, creating opportunities for data-driven machine learning (ML) approaches to provide quantitative decision support. This review examines the development of ML-based PD detection with particular emphasis on extreme gradient boosting (XGBoost), feature-engineered clinical models, speech and handwriting analysis, gait and wearable sensing, neuroimaging, and multimodal learning. Existing studies demonstrate that XGBoost is particularly suitable for heterogeneous tabular biomedical data because it can model nonlinear feature interactions, accommodate mixed feature distributions, provide feature importance estimates, and operate effectively with relatively structured datasets. However, reported performance is strongly influenced by dataset size, feature selection, class balancing, validation design, and data leakage. Recent systematic reviews further indicate inconsistent use of independent test sets, hyperparameter optimization, external validation, and direct model comparison. The review therefore synthesizes the current evidence into a data-driven PD detection framework integrating multimodal acquisition, preprocessing, feature selection, XGBoost-based classification, explainable artificial intelligence, uncertainty estimation, and external validation. Finally, the paper identifies research gaps involving dataset heterogeneity, clinical generalization, fairness, reproducibility, multimodal fusion, and real-world deployment, and proposes a roadmap toward clinically interpretable and robust PD decision-support systems.

Keywords— Parkinson's disease, XGBoost, machine learning, data-driven diagnosis, gait analysis, speech analysis, handwriting analysis, medical imaging, explainable artificial intelligence, multimodal learning.

I. INTRODUCTION

Parkinson's disease is a progressive neurological disorder associated with motor manifestations such as bradykinesia, rigidity, tremor, postural instability, and gait abnormalities, together with a broad range of non-motor symptoms. Diagnosis is primarily clinical and may become difficult when symptoms

are subtle, atypical, or overlap with other neurological conditions. This has motivated extensive research into computational methods that can transform measurable behavioural, physiological, and clinical characteristics into objective evidence for disease detection.

The rapid development of machine learning has created opportunities for analysing high-dimensional biomedical data without requiring manually defined diagnostic rules. Recent reviews have demonstrated that PD machine-learning research has expanded across voice, handwriting, gait, neuroimaging, EEG, wearable sensors, biomarkers, and multimodal clinical information. A 2024 systematic review identified substantial methodological variation across PD ML studies, particularly in validation strategies, hyperparameter optimization, and model comparison [1]. A comprehensive 2025 review further reported that voice, imaging, handwriting, gait, and EEG represent major categories of PD diagnostic data and emphasized persistent concerns regarding dataset quality and generalizability [2].

Voice analysis has received considerable attention because PD-related dysarthria can affect pitch, intensity, articulation, phonation stability, and vocal quality. Similarly, handwriting and drawing tasks can reveal micrographia, irregular stroke velocity, pressure variation, tremor, and impaired motor coordination. Gait and wearable-sensor signals provide complementary information about movement dynamics, while MRI, PET, SPECT, and EEG can provide information associated with neurological alterations. A 2024 review specifically examining handwriting and voice datasets reported extensive application of ML and DL algorithms and highlighted XGBoost as one of the strong-performing conventional classifiers for handwriting-based detection.

Among supervised machine-learning algorithms, XGBoost is particularly relevant to PD detection because many available datasets consist of structured numerical biomarkers rather than extremely large image collections. XGBoost constructs an additive ensemble of decision trees in which subsequent trees minimize the errors of previous trees. This allows nonlinear

relationships and interactions between clinical and behavioural variables to be represented efficiently.

Evidence from recent PD studies supports the importance of XGBoost while simultaneously demonstrating that algorithmic superiority cannot be inferred from accuracy alone. A 2024 decision-support study compared KNN, RF, SVM, XGBoost, and ensemble methods and reported 98% accuracy for a homogeneous XGBoost–RF approach under SMOTE and 10-fold cross-validation. A 2025 voice-based study reported 96.4% accuracy and 96.9% F1-score for XGBoost combined with chi-squared feature selection and SMOTE, illustrating the importance of preprocessing and feature-selection strategy in addition to the classifier itself.

Therefore, the central question of this review is:

How can XGBoost be integrated with rigorous data-driven machine-learning practices to develop reliable, interpretable, and clinically transferable Parkinson's disease detection systems?

The review is organized around seven objectives:

1. analysis of PD data modalities and biomarkers;
2. comparison of conventional ML, DL, and XGBoost approaches;
3. analysis of preprocessing and feature engineering;
4. examination of imbalance handling and validation;
5. explainability and uncertainty estimation;
6. clinical deployment and generalization; and
7. identification of open research challenges and future directions.

TABLE I
SCOPE OF THE REVIEW AND OBJECTIVE-TO-SECTION MAPPING

Research problem addressed	Section
PD data modalities, biomarkers and preprocessing	II
Comparison of ML, DL and XGBoost classifiers	III
Feature engineering, feature selection and imbalance handling	IV
Explainability, calibration and clinical trust	V
Validation, generalization and deployment	VI
Multimodal and longitudinal PD intelligence	VII
Open challenges and future research roadmap	VIII

II. PARKINSON'S DISEASE DATA MODALITIES AND COMPUTATIONAL BIOMARKERS

The performance of a PD detection model is fundamentally dependent on the quality and representativeness of its input data. Current literature can broadly be classified into voice, handwriting, gait and wearable sensing, neuroimaging, EEG, clinical variables, and biological biomarkers.

A. Voice-Based Biomarkers

Voice data are among the most frequently studied PD modalities because speech can be collected non-invasively using microphones and smartphones. PD-related vocal abnormalities can include reduced vocal intensity, altered fundamental frequency, increased jitter and shimmer,

dysphonia, articulation changes, and abnormal harmonic-to-noise characteristics.

Typical features include fundamental-frequency statistics, jitter, shimmer, harmonic-to-noise ratio, nonlinear dynamic descriptors, spectral characteristics, and Mel-frequency cepstral coefficients. These features produce structured numerical representations that are particularly suitable for tree-based ML algorithms.

The major advantage of voice analysis is low-cost and remote acquisition. However, microphone characteristics, environmental noise, language, age, gender, recording protocol, and speaking task can produce substantial distribution differences. Consequently, a model trained on controlled clinical recordings may perform differently on smartphone recordings collected in uncontrolled environments.

Recent work combining SMOTE and feature selection demonstrated that preprocessing substantially influences XGBoost performance for voice-based PD detection. This supports an important conclusion: XGBoost performance should be evaluated as part of a complete data-processing pipeline rather than as an isolated classifier.

B. Handwriting and Drawing Biomarkers

Handwriting provides a behavioural representation of fine motor control. Spiral and meander drawing tasks are commonly analysed because they can expose tremor, reduced movement amplitude, irregular trajectories, changes in velocity, and impaired coordination.

Two major strategies exist:

1. **Image-based analysis**, where handwriting images are analysed using CNNs or vision transformers.
2. **Kinematic analysis**, where pen position, velocity, acceleration, pressure, timing, and in-air movement are transformed into structured features.

The PaHaW and HandPD datasets are frequently used in the literature. A recent study using dynamic handwriting demonstrated the importance of feature selection and reported XGBoost-based results across multiple feature-selection strategies.

More recent multimodal research has combined handwriting images with engineered kinematic features and used XGBoost as the structured-data classifier. This direction indicates that XGBoost can serve as a complementary component rather than being restricted to single-modality classification.

C. Gait and Wearable-Sensor Data

Gait analysis provides information about whole-body motor dysfunction. Important variables include stride time, step duration, gait symmetry, acceleration, turning characteristics, cadence, and frequency-domain movement features.

Wearable inertial measurement units can continuously capture movement outside controlled laboratory environments. This creates opportunities for longitudinal monitoring rather than one-time diagnosis.

However, wearable-based datasets introduce challenges involving sensor placement, sampling rate, device heterogeneity, missing observations, and free-living behaviour. Recent research has also examined ML models for interpreting Parkinsonian motor states from wearable sensors,

demonstrating the potential of XGBoost alongside RF, SVM, logistic regression, and other models.

D. Neuroimaging and EEG

MRI, PET, SPECT, and EEG provide more direct information about structural, functional, or neural characteristics. These modalities can potentially capture disease-related changes that are not observable from behavioural measurements alone.

Their limitations are cost, acquisition complexity, preprocessing requirements, smaller sample sizes, and institutional variation. For these reasons, deep learning approaches are frequently used for raw imaging or signal representations, whereas XGBoost is more naturally applied after feature extraction or dimensionality reduction.

A recent comprehensive ML review categorized PD diagnostic literature into neuroimaging, voice, handwriting, gait, EEG, and other modalities, demonstrating the broadening of computational PD diagnosis beyond conventional voice datasets.

E. Multimodal Data

No individual modality completely describes PD. Voice can capture speech impairment, handwriting can capture fine motor dysfunction, gait can describe gross motor impairment, while imaging and clinical variables can provide complementary information.

Therefore, future systems should move from single-modality prediction toward reliability-aware multimodal fusion. The major research challenge is determining whether multimodal integration consistently improves external generalization rather than merely increasing performance on a single benchmark dataset.

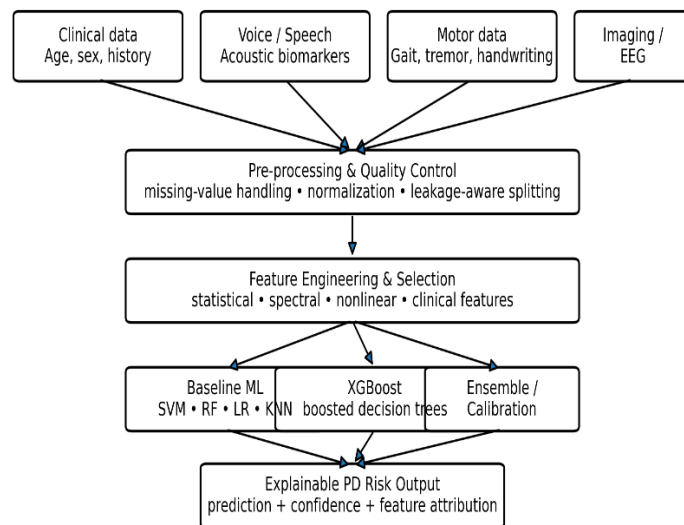


Fig. 1. Data-driven Parkinson's disease detection framework integrating behavioural, physiological, imaging and clinical data.

III. MACHINE LEARNING METHODS FOR PARKINSON'S DISEASE DETECTION

Conventional machine-learning algorithms remain highly relevant because many PD datasets contain relatively few participants but a large number of engineered features.

A. Classical Machine Learning

Logistic regression provides an interpretable linear baseline. SVM can model nonlinear boundaries through kernel functions. Random forest combines multiple randomized trees, while KNN determines class membership from local feature similarity.

These algorithms are important benchmarks because an advanced model should demonstrate a consistent benefit over simpler methods under the same data partition and preprocessing conditions.

Recent systematic reviews indicate that comparisons across PD studies are difficult because datasets, feature-processing pipelines, validation procedures, and evaluation metrics vary considerably.

B. Deep Learning

Deep learning is particularly useful when raw or minimally processed data are available. CNNs can analyse handwriting images and spectrograms, recurrent networks can process temporal signals, and transformer-based architectures can learn long-range dependencies.

However, deep networks may require larger datasets and careful regularization. High reported accuracy on small datasets should therefore be interpreted cautiously.

C. XGBoost

XGBoost is a gradient-boosted decision-tree framework particularly suited to structured biomedical data. Its principal strengths include:

- nonlinear feature interaction modelling;
- relatively strong performance on tabular datasets;
- regularization mechanisms;
- efficient training;
- handling of missing values;
- compatibility with feature-selection pipelines;
- support for feature importance and SHAP-based explanation.

The model can be represented conceptually as:

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i), f_k \in F \quad (1)$$

where (K) represents the number of trees and (f_k) denotes an individual regression tree.

The training objective can be expressed as:

$$L = \sum_i l(y_i, \hat{y}_i) + \sum_k \Omega(f_k) \quad (2)$$

- where $l(y_i, \hat{y}_i)$: prediction loss function
- $\Omega(f_k)$: regularization term controlling model complexity

The advantage of this formulation for PD data is that nonlinear interactions can be captured without requiring a manually designed decision boundary.

A 2024 PD decision-support study found that an XGBoost–RF ensemble achieved 98% accuracy with SMOTE and 10-fold cross-validation, illustrating the potential of boosting-based ensemble approaches.

A separate clinical study reported an XGBoost validation AUC of 0.908 and accuracy of 0.819, emphasizing an important

lesson: performance can decrease considerably when a model is evaluated on validation data rather than training data.

IV. FEATURE ENGINEERING, SELECTION AND CLASS-IMBALANCE HANDLING

Feature engineering is one of the most important stages in data-driven PD detection. The same classifier can produce substantially different results depending on how raw measurements are transformed.

A. Feature Engineering

For voice data, useful features include:

- fundamental frequency;
- jitter;
- shimmer;
- harmonic-to-noise ratio;
- MFCCs;
- spectral descriptors;
- nonlinear dynamic measures.

For handwriting, features may include:

- stroke velocity;
- acceleration;
- pressure;
- displacement;
- stroke duration;
- pause duration;
- tremor frequency;
- trajectory irregularity.

For gait and wearable sensors:

- stride interval;
- cadence;
- acceleration statistics;
- gait asymmetry;
- frequency-domain energy;
- turning characteristics.

Clinical features may include age, disease duration, medication-related variables, motor scores, cognitive measurements, and other standardized assessments.

B. Feature Selection

High-dimensional feature sets can increase overfitting and reduce interpretability. Common approaches include:

- correlation filtering;
- mutual information;
- chi-squared selection;
- recursive feature elimination;
- LASSO;
- Relief-based methods;
- tree-based importance;
- SHAP-based selection.

Recent voice-based research demonstrated that combining SMOTE and feature selection substantially improved XGBoost performance, with the reported best configuration using chi-squared feature selection.

C. Class Imbalance

PD datasets may contain unequal numbers of patients and healthy controls. Imbalance can bias a classifier toward the majority class.

SMOTE is frequently applied to generate synthetic minority examples. However, oversampling must be performed **inside the training folds only**. If SMOTE is applied before train-test splitting, synthetic information derived from test observations can leak into training data.

Therefore, a correct pipeline is:

Split → preprocessing on training data → feature selection on training data → SMOTE on training data → model training → untouched test evaluation.

This methodological discipline is more important than simply selecting a sophisticated classifier.

TABLE II
COMPARATIVE REVIEW OF DATA MODALITIES AND MACHINE-LEARNING STRATEGIES

Modality	Representative features	Suitable ML methods	Major strength	Major limitation
Voice	Jitter, shimmer, MFCC, HNR, pitch	XGBoost, SVM, RF, DL	Low-cost and non-invasive	Language/device variability
Handwriting	Velocity, pressure, trajectory, spiral features	XGBoost, RF, SVM, CNN	Captures fine motor impairment	Small datasets
Gait	Stride, cadence, acceleration, asymmetry	XGBoost, RF, SVM, DL	Objective motor assessment	Sensor/pro tocol variability
Wearable sensors	IMU, tremor, movement statistics	XGBoost, RF, SVM, NN	Continuous monitoring	Missing/no isy observations
MRI/PET/SP ECT	Structural/functional imaging features	CNN, transformer, XGBoost after feature extraction	Neurobiological information	Expensive and complex
EEG	Spectral and temporal features	SVM, RF, XGBoost, DL	Neural activity information	Acquisition complexity
Clinical data	Demographic and clinical scores	XGBoost, LR, RF	Interpretable structured data	Clinical-site variation
Multimodal	Combined behavioural/clinical signals	Ensemble, fusion, XGBoost	Complementary evidence	Missing modalities and alignment

Recent systematic reviews similarly show that PD ML research spans voice, neuroimaging, handwriting, gait and EEG, while methodological heterogeneity limits direct comparison.

V. EXPLAINABILITY, CALIBRATION AND CLINICAL TRUST

A PD detection system should not only provide a predicted class; it should provide sufficient evidence for clinicians to understand why a prediction was generated.

Tree-based models are advantageous because their structured features can be examined using feature-importance methods and SHAP explanations. SHAP assigns contribution values to individual features and can be used to identify which biomarkers increase or decrease the predicted PD probability.

For example, an explanation may show that a prediction is strongly influenced by:

1. abnormal voice stability;
2. increased tremor-related features;
3. reduced gait regularity;
4. specific clinical measurements.

However, feature importance should not be interpreted as proof of causality. A highly predictive variable may represent correlation rather than a biological mechanism.

A. Calibration

A classifier may achieve high discrimination while producing poorly calibrated probabilities. For clinical decision support, the difference between a predicted probability of 0.60 and 0.95 is clinically meaningful.

Calibration methods such as Platt scaling and isotonic regression can be considered after model training. Calibration should be evaluated on data independent of model fitting.

B. Uncertainty

Future PD systems should estimate uncertainty and identify cases that are outside the training distribution.

A reliable system should be capable of communicating:

“The model predicts elevated PD risk, but confidence is low because the current case differs from the training population.” This is more clinically useful than presenting an apparently precise but unsupported prediction.

C. Explainability Evaluation

Future studies should evaluate explanations according to:

- faithfulness;
- stability;
- expert agreement;
- consistency across repeated measurements;
- robustness to feature perturbation.

A recent systematic biomarker review emphasizes the growing role of ML in identifying candidate biomarkers while also demonstrating that early diagnosis remains a challenging clinical problem.

VI. VALIDATION, GENERALIZATION AND CLINICAL DEPLOYMENT

The largest methodological weakness in PD machine-learning research is not necessarily the absence of advanced algorithms; it is insufficient validation.

A. Data Leakage

If multiple samples from the same participant appear in both training and testing subsets, the model can learn participant-specific characteristics rather than disease characteristics. Therefore, participant-level splitting is essential.

For longitudinal datasets, temporal leakage must also be avoided. Future observations should never influence predictions for earlier observations.

B. External Validation

A model trained at one hospital may encounter a different population, recording device, language, medication distribution, or clinical protocol at another hospital.

Consequently, external validation should involve:

- different institutions;
- different acquisition devices;
- different demographic distributions;
- independent patient cohorts;
- ideally prospective data.

A major 2025 review of 117 PD ML studies found that only 28 studies, or 23.9%, were judged to have low risk of bias using PROBAST, emphasizing the importance of methodological quality rather than accuracy alone.

C. Evaluation Metrics

Accuracy alone is insufficient, particularly when class imbalance exists.

- **Accuracy:**

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

- **Precision:**

$$Precision = \frac{TP}{TP+FP} \quad (4)$$

- **Recall:**

$$Recall = \frac{TP}{TP+FN} \quad (5)$$

- **F1-Score:**

$$F1 = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \quad (6)$$

Additional metrics should include ROC-AUC, PR-AUC, specificity, negative predictive value, calibration error, confidence intervals, and decision-curve analysis where appropriate.

D. Clinical Deployment

A practical architecture can be organized into three stages:

Stage 1 – Screening: Low-cost voice, handwriting, or wearable measurements identify individuals requiring further assessment.

Stage 2 – Data-driven risk estimation: XGBoost integrates selected behavioural, clinical, and physiological biomarkers.

Stage 3 – Clinical decision support: The system reports calibrated risk, influential features, uncertainty, missing information, and recommendations for further assessment.

The system should support clinicians rather than replace clinical judgment.

VII. MULTIMODAL, LONGITUDINAL AND DATA-DRIVEN PD INTELLIGENCE

The future of PD detection is likely to move beyond a single binary prediction.

A. Multimodal Fusion

A multimodal system can combine:

Voice + Handwriting + Gait + Wearable Sensors + Clinical Data

XGBoost can operate as the structured-data component while deep models process high-dimensional signals.

A possible architecture is:

Raw Data → Modality-Specific Preprocessing → Feature Extraction → Feature Selection → XGBoost / DL Models → Multimodal Fusion → Calibration → Explainability → Clinical Decision Support

Recent 2026 research demonstrates this direction by combining handwriting images with kinematic features and using XGBoost for structured features under subject-wise cross-validation.

B. Longitudinal Monitoring

PD is progressive. Therefore, repeated measurements may be more informative than a single cross-sectional classification.

Future models should investigate:

- disease progression;
- symptom severity;
- motor-state transitions;
- treatment response;
- diagnostic conversion;
- patient-specific baselines.

C. Personalized Learning

Rather than applying one global model to every patient, future systems may maintain patient-specific baselines.

For example:

Patient Baseline → New Measurement → Deviation Analysis → Updated Risk → Clinical Review

This approach could identify clinically meaningful changes even when absolute values remain within population-normal ranges.

VIII. OPEN CHALLENGES AND RESEARCH ROADMAP

Several challenges must be resolved before XGBoost-based PD detection can become a reliable clinical technology.

1. Dataset fragmentation

Many datasets remain relatively small and originate from individual institutions. This limits generalizability.

2. Lack of standardized acquisition

Voice, handwriting, gait, EEG, and imaging protocols differ across studies.

3. Data leakage

Participant-level leakage can artificially inflate reported performance.

4. Limited external validation

Many studies report only internal cross-validation.

5. Explainability limitations

Feature importance does not establish causality.

6. Demographic fairness

Performance may differ according to age, sex, language, ethnicity, disease stage, or clinical site.

7. Multimodal missingness

A patient may not have every modality available.

8. Clinical integration

A technically accurate model may still fail if its output is difficult to interpret or integrate into clinical workflow.

9. Longitudinal evaluation

Cross-sectional datasets cannot fully represent disease progression.

10. Prospective validation

Future research should move from retrospective datasets toward prospective multi-site clinical evaluation.

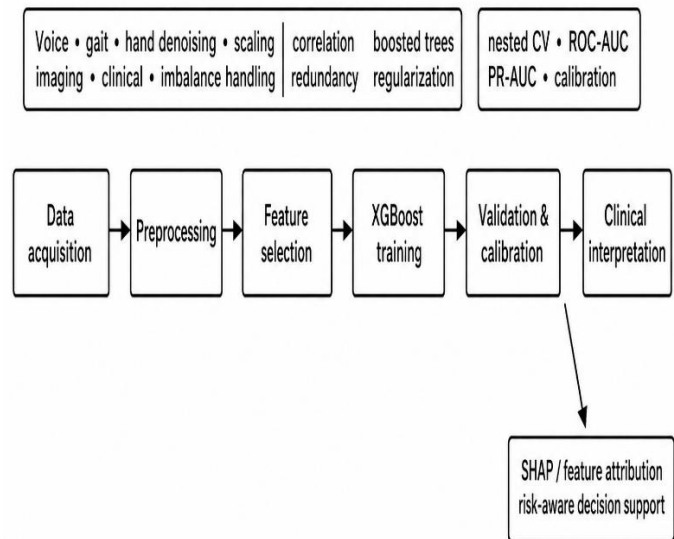


Fig. 2. Proposed XGBoost-centered data-driven Parkinson's disease detection pipeline.

The proposed review framework recommends that future studies follow the sequence:

Multi-site Dataset → Quality Control → Participant-Level Split → Preprocessing → Feature Engineering → Feature Selection → Imbalance Handling → XGBoost Benchmark → Multimodal Fusion → Calibration → Explainability → External Validation → Prospective Clinical Evaluation.

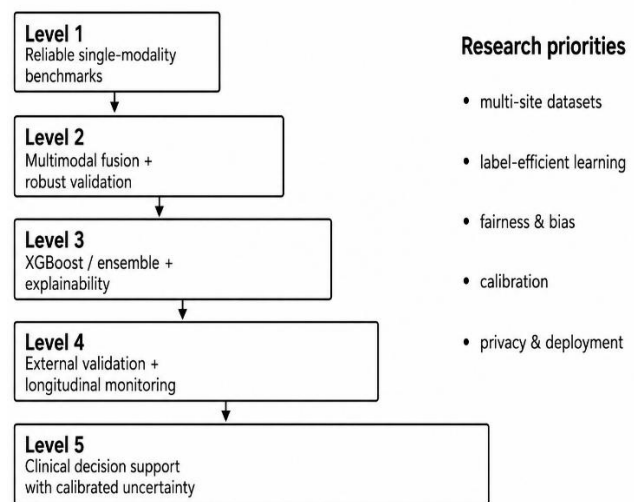


Fig. 3. Research roadmap from benchmark-level PD prediction toward externally validated clinical decision support.

The roadmap can be divided into three phases.

Short term: establish standardized datasets, participant-level validation, reproducible preprocessing, and consistent metrics.

Medium term: develop multimodal XGBoost-based ensembles, calibrated predictions, uncertainty estimation, and cross-site validation.

Long term: conduct prospective clinical studies and integrate PD AI systems into clinician-facing decision-support platforms.

A particularly important research direction is the use of XGBoost as an interpretable structured-data layer rather than claiming it as a universal replacement for deep learning. Recent work has demonstrated that XGBoost can function effectively within multimodal architectures, including systems where it serves as a meta-learner or processes engineered kinematic features.

IX. CONCLUSION

This review examined Parkinson's disease detection from the perspective of XGBoost and data-driven machine learning. Current literature demonstrates the value of voice, handwriting, gait, wearable sensors, EEG, imaging, biomarkers, and clinical variables as computational sources of PD-related information. XGBoost is particularly suitable for heterogeneous structured features because it captures nonlinear interactions, supports efficient learning, and can be integrated with feature-selection and explainability techniques. However, high internal accuracy does not establish clinical reliability. Future PD detection systems should therefore prioritize participant-level validation, class-imbalance control, calibration, uncertainty estimation, external testing, multimodal integration, fairness, and prospective clinical evaluation. The most promising direction is a human-centered, multimodal and externally validated framework in which XGBoost functions as a strong interpretable component of a broader clinical decision-support architecture.

REFERENCES

- [1] T. Tabashum, R. C. Snyder, M. K. O'Brien, and M. V. Albert, "Machine Learning Models for Parkinson Disease: Systematic Review," *JMIR Medical Informatics*, vol. 12, e50117, 2024.
- [2] S. Shokrpour, A. M. MoghadamFarid, S. Bazzaz Abkenar, M. Haghi Kashani, M. Akbari, and M. Sarvizadeh, "Machine learning for Parkinson's disease: a comprehensive review of datasets, algorithms, and challenges," *npj Parkinson's Disease*, vol. 11, art. no. 187, 2025.
- [3] M. A. Islam, M. Z. H. Majumder, M. A. Hussein, K. M. Hossain, and M. S. Miah, "A review of machine learning and deep learning algorithms for Parkinson's disease detection using handwriting and voice datasets," *Heliyon*, vol. 10, no. 3, e25469, 2024.
- [4] L. Di Biase, P. M. Pecoraro, G. Pecoraro, S. A. Shah, and V. Di Lazzaro, "Machine learning and wearable sensors for

automated Parkinson's disease diagnosis aid: a systematic review," *Journal of Neurology*, 2024.

[5] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, 2016, pp. 785–794.

[6] "An artificial intelligence-based decision support system for early and accurate diagnosis of Parkinson's Disease," *Decision Analytics Journal*, vol. 10, art. no. 100381, 2024.

[7] J. Zhang, Y. Zhang, Y. Weng, A. A. Hosseini, B. Wang, T. Dening, W. Fan, et al., "Applications of machine learning for computer-aided diagnosis of Parkinson's disease: progress and benchmark case study," *Artificial Intelligence Review*, vol. 58, art. no. 357, 2025.

[8] C. Altham, H. Zhang, and E. Pereira, "Machine learning for the detection and diagnosis of cognitive impairment in Parkinson's Disease: A systematic review," *PLOS ONE*, vol. 19, no. 5, e0303644, 2024.

[9] H. Rabie and M. A. Akhloufi, "A review of machine learning and deep learning for Parkinson's disease detection," *Discover Artificial Intelligence*, vol. 5, art. no. 24, 2025.

[10] "Enhancing Parkinson's Disease Detection by Combining SMOTE and Feature Selection for Improved Machine Learning Classification Using Voice Recordings," 2025.

[11] "A novel feature extraction method based on dynamic handwriting for Parkinson's disease detection," 2025.

[12] "Parkinson disease detection based on in-air dynamics feature extraction and selection using machine learning," *Scientific Reports*, 2025.

[13] J. Han, M. Sun, J. Yang, and Y. An, "Machine learning-based non-invasive Parkinson's disease diagnostic model using clinical blood biomarkers," *Neurological Sciences*, 2025.

[14] J. Lim and Y. Ay, "The Performance of Biomarkers for the Diagnosis of Parkinson Disease: A Systematic Review," *The American Journal of Medicine*, vol. 138, no. 12, pp. 1662–1670.e4, 2025.

[15] "Enhanced meta ensemble stacking approach with XGBoost and Optuna based detection of Parkinson's disease," *Frontiers in Digital Health*, 2026.

[16] A. Bensefia, N. Al Hashlamoun, and K. K. Napa, "Multimodal fusion of handwriting images and kinematic features for Parkinson disease detection," *Scientific Reports*, 2026.