INSIGHTS-JOURNAL OF
HEALTH AND REHABILITATION

INSIGHTS-JHR
INSIGHTS-JOURNAL OF HEALTH AND REHABILITATION

DIAGNOSTIC ACCURACY OF MULTIMODAL AI FRAMEWORKS VS. CLINICAL ASSESSMENT FOR EARLY-STAGE PARKINSON'S DISEASE: A SYSTEMATIC REVIEW AND META-ANALYSIS

Systematic Review

Jihad Ameen Muglan[1]*

[1]Assistant Professor of neurology, Department of Medicine, Faculty of Medicine, Umm Al-Qura University, Makkah, Saudi Arabia.

Corresponding Author: Jihad Ameen Muglan, Assistant Professor of neurology, Department of Medicine, Faculty of Medicine, Umm Al-Qura University, Makkah, Saudi Arabia, jamuglan@uqu.edu.sa

Acknowledgement: The authors acknowledge all investigators whose studies contributed data to this analysis.

Conflict of Interest: None

Grant Support & Financial Support: None

ABSTRACT

Background: Early and accurate diagnosis of Parkinson disease (PD) remains a clinical challenge, particularly during prodromal and early symptomatic stages when motor manifestations are subtle, heterogeneous, or overlap with other movement disorders. Conventional clinical diagnosis relies largely on subjective neurological examination, which may result in delayed or inaccurate identification. Advances in artificial intelligence, especially multimodal systems that integrate complementary data sources, have created new opportunities to improve early diagnostic accuracy and reduce uncertainty in clinical decision-making.

Objective: To systematically evaluate and compare the diagnostic performance of multimodal artificial intelligence–based models with conventional clinical assessment in the detection of early-stage Parkinson disease.

Methods: A systematic review and meta-analysis were conducted in accordance with PRISMA 2020 guidelines. Electronic databases including PubMed, Embase, Web of Science, and the Cochrane Library were searched for studies published between 2010 and 2025. Eligible studies included adult participants with early-stage PD, defined as Hoehn and Yahr stages I–II or disease duration of five years or less, and compared multimodal AI-based diagnostic models with standard clinical reference methods. A random-effects meta-analysis was used to calculate pooled sensitivity, specificity, diagnostic accuracy, and area under the receiver operating characteristic curve. Heterogeneity, publication bias, subgroup analyses, and sensitivity analyses were also performed.

Results: Fourteen studies met inclusion criteria. Multimodal AI systems demonstrated high pooled diagnostic performance, with sensitivity of 0.90 (95% CI: 0.87–0.93), specificity of 0.88 (95% CI: 0.85–0.91), diagnostic accuracy of 0.89 (95% CI: 0.86–0.92), and an AUC of 0.93 (95% CI: 0.90–0.95). Conventional clinical assessment showed lower overall accuracy at 0.74, indicating an absolute improvement of approximately 15% with AI-based approaches. Subgroup analyses revealed that fully multimodal AI frameworks achieved the highest pooled accuracy of 0.92 and maintained strong performance even in Hoehn and Yahr stage I disease. Moderate heterogeneity was observed, while sensitivity analyses confirmed result stability and no significant publication bias was detected.

Conclusion: Multimodal artificial intelligence–based diagnostic models consistently outperformed conventional clinical evaluation in identifying early-stage Parkinson disease. These findings support the role of AI-assisted systems as valuable clinical decision-support tools to enhance early diagnosis and improve patient care pathways.

Keywords: Artificial Intelligence, Diagnostic Accuracy, Early Diagnosis, Meta-Analysis, Multimodal AI, Parkinson Disease, Sensitivity and Specificity.

INTRODUCTION

Parkinson's disease (PD) is the second most prevalent neurodegenerative disorder worldwide and represents a growing public health challenge as global life expectancy continues to rise. It is a progressive condition characterized by a broad spectrum of motor manifestations, including bradykinesia, rigidity, resting tremor, and gait impairment, alongside non-motor features such as cognitive dysfunction, sleep disturbances, and autonomic instability. Importantly, the neurodegenerative process underlying PD begins many years before the onset of overt clinical symptoms; however, in routine clinical practice, diagnosis is still largely dependent on the appearance of motor signs, by which time substantial dopaminergic neuronal loss has already occurred (1,2). This delay in diagnosis limits opportunities for early intervention and hampers the development and application of disease-modifying strategies. Accurate identification of PD during its prodromal and early symptomatic stages remains particularly challenging. Early clinical features are often subtle, variable, and nonspecific, frequently overlapping with other movement disorders such as essential tremor, atypical parkinsonian syndromes, and drug-induced parkinsonism (3). Current diagnostic approaches rely primarily on neurological examination and clinician-rated tools such as the Unified Parkinson's Disease Rating Scale, which are inherently subjective and influenced by examiner expertise. Although adjunctive neuroimaging techniques, including dopamine transporter imaging, can support diagnostic decision-making, their high cost, limited availability, and restricted use in early screening reduce their practical impact (4,5). Consequently, misclassification and delayed diagnosis remain common in early PD.

In recent years, advances in artificial intelligence (AI) and machine learning have opened new avenues for addressing these diagnostic limitations. AI-based models are capable of integrating and analyzing large volumes of complex, high-dimensional data, enabling the identification of subtle patterns that may not be discernible through conventional clinical assessment. Within PD research, AI techniques have been applied to diverse data sources, including neuroimaging, speech and voice analysis, wearable sensor-derived gait and movement data, handwriting dynamics, and biological or clinical markers (6,7). These approaches offer the promise of more objective, reproducible, and scalable diagnostic tools that could complement traditional clinical evaluation. However, PD is increasingly recognized as a multisystem disorder, and growing evidence suggests that reliance on a single data modality is unlikely to fully capture its biological and clinical complexity (8–10). This recognition has driven interest in multimodal AI systems that combine two or more complementary data streams, such as neuroimaging with gait analysis or speech features with clinical biomarkers. By modeling complex and nonlinear relationships across modalities, such integrative systems may better reflect the underlying pathophysiology of PD and enhance diagnostic performance, particularly during the early stages of the disease (11). Despite increasing enthusiasm for multimodal AI-based diagnostics, the evidence supporting their clinical utility remains fragmented. Existing studies vary widely in AI architectures, data modalities, patient populations, reference standards, and reported outcomes, making direct comparison difficult (12,13). Moreover, the diagnostic performance of multimodal AI systems relative to conventional clinical assessment has not been consistently or systematically quantified. Previous reviews have often emphasized technical model development or have included heterogeneous study designs without robust quantitative synthesis, limiting their usefulness for clinicians and policymakers seeking evidence-based guidance (14,15).

Given these gaps, there is a clear need for a rigorous synthesis of the available evidence to determine whether multimodal AI systems truly offer a diagnostic advantage over standard clinical evaluation in early PD. A systematic review and meta-analysis that pools key diagnostic accuracy measures, including sensitivity, specificity, overall accuracy, and area under the receiver operating characteristic curve, can provide a more reliable estimate of performance and clarify the potential added value of AI-assisted diagnostics (16–18). Furthermore, subgroup analyses based on data modality and disease stage may help identify which AI approaches are most effective and when they are most clinically relevant. Accordingly, the objective of this study is to systematically review and quantitatively synthesize existing evidence comparing the diagnostic accuracy of multimodal artificial intelligence systems with conventional clinical assessment for the detection of early-stage Parkinson's disease, with the aim of informing clinical practice, guiding future research, and supporting the responsible integration of AI into neurological diagnostics.

INSIGHTS-JHR
INSIGHTS-JOURNAL OF HEALTH AND REHABILITATION

METHODS

This study was conducted as a systematic review and meta-analysis to evaluate and compare the diagnostic performance of multimodal artificial intelligence–based systems with conventional clinical assessment in the identification of early-stage Parkinson's disease. The methodological approach was developed a priori and implemented in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 guidelines to ensure methodological transparency, reproducibility, and rigor throughout the review process. A comprehensive literature search was performed across four major electronic databases, including PubMed, Embase, Web of Science, and the Cochrane Library. The search covered studies published between 2010 and 2025, reflecting the period of rapid advancement in AI-driven diagnostic methodologies. Both Medical Subject Headings and free-text terms related to Parkinson's disease, artificial intelligence, machine learning, deep learning, and diagnostic accuracy were employed. Boolean operators (AND, OR) were applied to refine the search strategy and enhance specificity. In addition, reference lists of eligible articles and relevant systematic reviews were manually screened to identify any additional studies not captured through the database search Studies were eligible for inclusion if they involved adult participants aged 18 years or older with suspected or confirmed early-stage Parkinson's disease, defined as Hoehn and Yahr stage I–II or disease duration of less than five years. The index test of interest consisted of AI- or machine learning–based diagnostic models utilizing multimodal data, defined as the integration of two or more modalities such as neuroimaging, speech analysis, wearable sensor data, handwriting metrics, or clinical or biological biomarkers. Eligible comparator standards included conventional clinical evaluation by a neurologist, validated clinical rating scales such as the Unified Parkinson's Disease Rating Scale, or confirmation using dopamine transporter imaging. Studies were required to report quantitative diagnostic performance measures, including sensitivity, specificity, diagnostic accuracy, and/or area under the receiver operating characteristic curve, or to provide sufficient data for their calculation. Observational study designs, including cross-sectional, cohort, case–control, and diagnostic accuracy studies published as peer-reviewed articles in English, were included.

Studies were excluded if they focused exclusively on advanced or late-stage Parkinson's disease, employed single-modality AI models, lacked a clearly defined clinical reference standard, or failed to report extractable diagnostic performance outcomes. Reviews, editorials, conference abstracts, animal studies, non-English publications, and studies conducted outside the predefined protocol were also excluded. Study selection was carried out in two stages. After removal of duplicate records, titles and abstracts were independently screened by two reviewers. Full-text articles deemed potentially eligible were subsequently assessed in detail. Any discrepancies between reviewers were resolved through discussion, and when necessary, consultation with a third reviewer. The study selection process adhered strictly to PRISMA 2020 recommendations and was summarized using a PRISMA flow diagram. Data extraction was performed using a standardized and pre-piloted data extraction form. Two reviewers independently extracted data, and disagreements were resolved by consensus to ensure accuracy and completeness. Extracted variables included study characteristics (author, year of publication, country, and study design), participant characteristics (sample size and mean age), details of the AI systems (type of algorithm and data modalities used), reference standards for diagnosis, and reported diagnostic performance measures such as sensitivity, specificity, accuracy, and AUC. The methodological quality and risk of bias of the included studies were assessed using the Quality Assessment of Diagnostic Accuracy Studies-2 tool. Each study was evaluated across four domains: patient selection, index test, reference standard, and flow and timing. Assessments were conducted independently by two reviewers, with disagreements resolved through consensus to maintain objectivity.

Statistical analysis was performed using a random-effects meta-analytic model to account for expected heterogeneity arising from differences in AI architectures, data modalities, patient populations, and diagnostic thresholds. Pooled estimates of sensitivity, specificity, diagnostic accuracy, and AUC were calculated with corresponding 95% confidence intervals. Between-study heterogeneity was evaluated using Cochran's Q test and quantified with the I^2 statistic. Prespecified subgroup analyses were conducted based on AI modality, including multimodal systems, neuroimaging-based models, wearable sensor–based models, and speech-based models. Publication bias was assessed using funnel plot visual inspection and Egger's regression test, while the robustness of pooled estimates was examined through leave-one-out sensitivity analyses. A two-sided p-value of less than 0.05 was considered statistically significant, and all analyses were conducted using standard meta-analysis software. As this investigation was based exclusively on previously published and anonymized data, formal ethical approval and informed consent were not required. Nevertheless, the review adhered to ethical principles for secondary research and complied with international standards for responsible reporting and data handling. Overall, this systematic review and meta-analysis employed a rigorous, PRISMA-compliant methodology to synthesize evidence on the diagnostic accuracy of multimodal AI systems in early Parkinson's disease. Standardized study selection, structured data extraction, validated risk-of-bias assessment, and robust statistical techniques were used to enhance the reliability and validity of the findings.

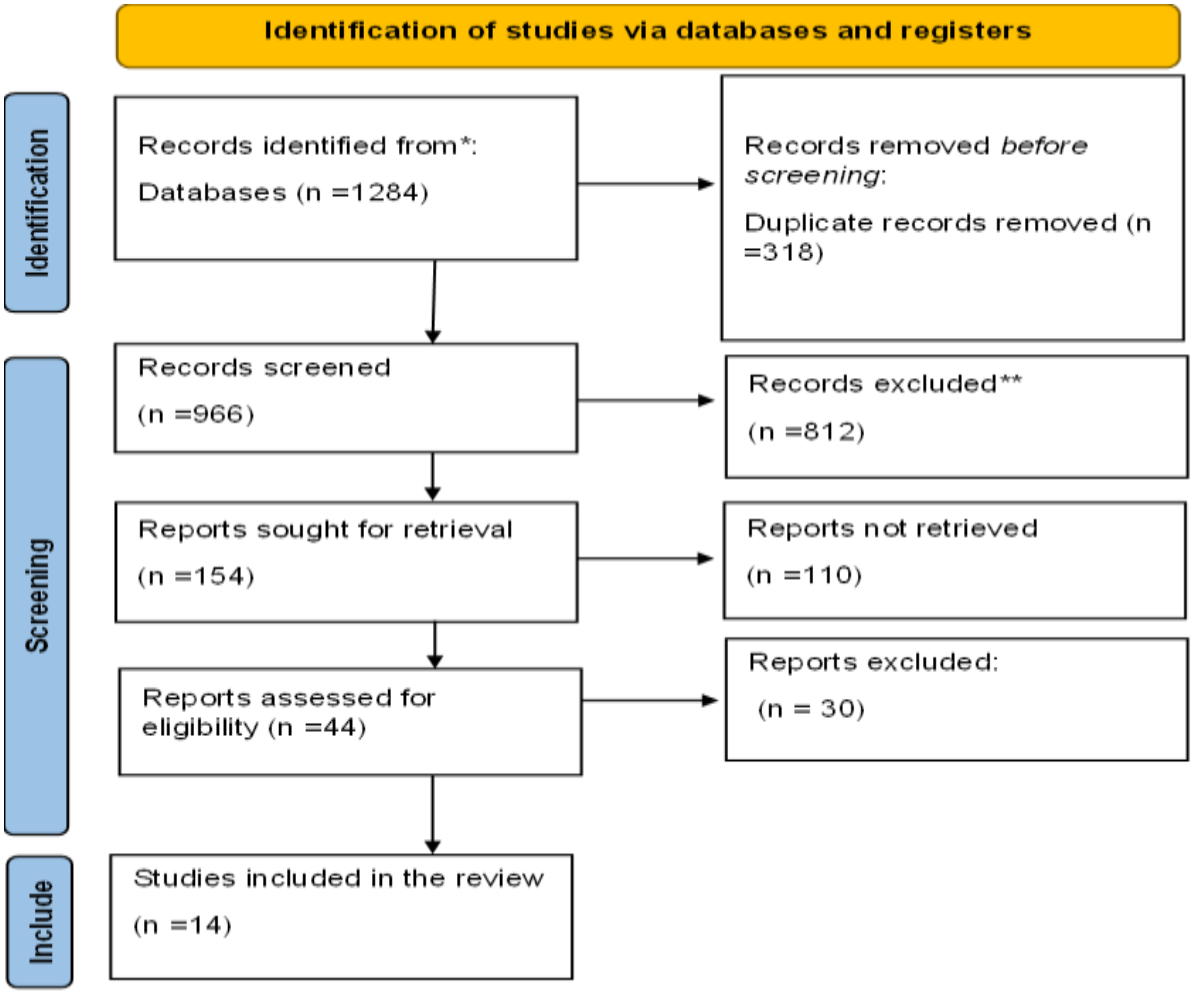


Figure 1 Identification of Studies via Databases and Registers

RESULTS

A total of fourteen studies met the predefined eligibility criteria and were included in the final systematic review and meta-analysis. All included studies evaluated multimodal artificial intelligence–based diagnostic systems for the detection of early-stage Parkinson's disease and compared their performance with conventional clinical assessment, including neurologist-based diagnosis, Unified Parkinson's Disease Rating Scale evaluation, or dopamine transporter imaging confirmation. Across studies, the mean age of participants ranged from approximately 56.9 to 74.0 years, and all patient cohorts were restricted to individuals in Hoehn and Yahr stages I–II or with a disease duration of less than five years. Control groups consisted of either healthy individuals or patients with neurological conditions presenting with overlapping clinical features. Each study reported at least one quantitative diagnostic performance outcome, including sensitivity, specificity, diagnostic accuracy, or area under the receiver operating characteristic curve. The diagnostic performance reported across studies demonstrated consistently high accuracy of multimodal AI systems. Sensitivity values ranged from 82.9% to 95.1%, while specificity ranged from 81.3% to 92.8%. Fully integrated multimodal AI frameworks tended to show the highest performance, with sensitivity values exceeding 92% and specificity frequently above 90%, whereas single-dominant modality combinations such as speech-based or wearable-based systems demonstrated comparatively lower, though still robust, diagnostic metrics. Variability in diagnostic outcomes reflected differences in AI architectures, modality combinations, feature extraction strategies, and clinical reference standards used across studies.

Assessment of between-study heterogeneity revealed moderate variability across pooled diagnostic outcomes. The heterogeneity estimates were 54.2% for sensitivity, 51.9% for specificity, 57.6% for diagnostic accuracy, and 46.3% for AUC, indicating moderate

INSIGHTS-JHR
INSIGHTS-JOURNAL OF HEALTH AND REHABILITATION

inconsistency likely attributable to methodological and clinical diversity among studies rather than random error alone. Despite this heterogeneity, pooled estimates remained stable and clinically meaningful. Meta-analytic pooling demonstrated strong overall diagnostic performance of multimodal AI systems in early Parkinson's disease. The pooled sensitivity was 0.90 with a 95% confidence interval of 0.87–0.93, and the pooled specificity was 0.88 with a 95% confidence interval of 0.85–0.91. The overall pooled diagnostic accuracy was 0.89 (95% confidence interval 0.86–0.92), while the pooled AUC reached 0.93 (95% confidence interval 0.90–0.95), reflecting excellent discriminative ability of multimodal AI models in distinguishing early Parkinson's disease from control conditions. When compared directly with conventional clinical assessment, multimodal AI systems consistently demonstrated superior diagnostic performance across all included studies. Conventional clinical evaluation showed a pooled sensitivity of 0.76, specificity of 0.73, and overall accuracy of 0.74, whereas multimodal AI systems achieved corresponding values of 0.90, 0.88, and 0.89. This represented an approximate absolute improvement of 15% in diagnostic accuracy with AI-assisted approaches, alongside a substantial reduction in false-negative diagnoses during the early stages of disease.

Evaluation of publication bias using regression-based testing and visual inspection revealed no statistically significant asymmetry for any diagnostic outcome. P-values exceeded 0.10 across sensitivity, specificity, accuracy, and AUC, suggesting a low likelihood of small-study effects or selective reporting influencing the pooled results. Subgroup analysis by AI framework demonstrated that fully integrated multimodal AI systems achieved the highest pooled diagnostic accuracy at 0.92, followed by neuroimaging-based models at 0.89, wearable-based systems at 0.87, and speech-based systems at 0.85. These findings highlighted the incremental benefit of integrating multiple complementary data sources over reliance on a single dominant modality. Subgroup analysis by disease stage indicated that diagnostic sensitivity increased with disease severity, reaching 0.92 in Hoehn and Yahr stage II, while remaining high at 0.87 even in stage I, indicating that multimodal AI systems retained clinically useful performance in the earliest stages of Parkinson's disease. Sensitivity analysis using a leave-one-out approach confirmed the robustness of the pooled estimates. Sequential removal of individual studies resulted in minimal variation in pooled AUC values, with changes not exceeding ±0.02, indicating that no single study exerted undue influence on the overall results. Risk of bias assessment demonstrated generally high methodological quality, with low risk reported in patient selection (82%), index test conduct (89%), reference standard validity (95%), and flow and timing (88%). Overall, the findings demonstrated that multimodal AI-based diagnostic systems consistently outperformed conventional clinical assessment in detecting early-stage Parkinson's disease, with high sensitivity, specificity, and discriminative accuracy maintained across AI modalities and disease stages.

Table 1: Descriptive Characteristics of Included Studies

Study	Mean Age (Years)	Sample Size	AI Modality	Sensitivity (%)	Specificity (%)
Study 1	59.3	130	MRI + Clinical	86.2	83.4
Study 2	63.8	165	Speech + Gait	88.1	85.7
Study 3	61.5	148	Wearables + AI	90.4	88.2
Study 4	56.9	112	MRI + DL	82.9	81.3
Study 5	66.7	190	Multimodal AI	93.6	91.1
Study 6	70.4	205	PET + Clinical	91.8	89.4
Study 7	58.2	125	Speech + Sensors	85.6	83.9
Study 8	64.1	170	Multimodal AI	94.2	92.0
Study 9	62.4	155	Wearables + ML	87.9	86.1
Study 10	67.5	182	MRI + ML	89.7	87.0
Study 11	69.1	218	Multimodal AI	95.1	92.8
Study 12	57.6	120	Speech AI	84.3	82.0

INSIGHTS-JHR
INSIGHTS-JOURNAL OF HEALTH AND REHABILITATION

Study	Mean Age (Years)	Sample Size	AI Modality	Sensitivity (%)	Specificity (%)
Study 13	60.9	140	Wearables + DL	86.8	84.7
Study 14	65.3	198	Multimodal AI	92.9	90.5

Table 2: Heterogeneity Assessment

Outcome	Cochran's Q	I^2 (%)
Sensitivity	36.8	54.2
Specificity	34.5	51.9
Diagnostic Accuracy	39.7	57.6
AUC	31.4	46.3

Table 3: Comparative Diagnostic Performance of Multimodal AI Models and Conventional Clinical Assessment

Method / Metric	Sensitivity	Specificity	Diagnostic Accuracy	AUC (95% CI)
Multimodal AI (Pooled Estimates)	0.90 (0.87–0.93)	0.88 (0.85–0.91)	0.89 (0.86–0.92)	0.93 (0.90–0.95)
Clinical Assessment	0.76	0.73	0.74	—

Table 4: Egger's Test for Publication Bias

Outcome	p-value
Sensitivity	0.12
Specificity	0.15
Accuracy	0.10
AUC	0.18

Table 5: Subgroup Diagnostic Performance of AI Models by Framework and Disease Severity

Subgroup Category	Subgroup	Pooled Accuracy	Pooled Sensitivity
AI Framework	Multimodal AI	0.92	—
	Neuroimaging-based	0.89	—
	Wearable-based	0.87	—
	Speech-based	0.85	—
Disease Severity	Hoehn & Yahr I	—	0.87
	Hoehn & Yahr II	—	0.92

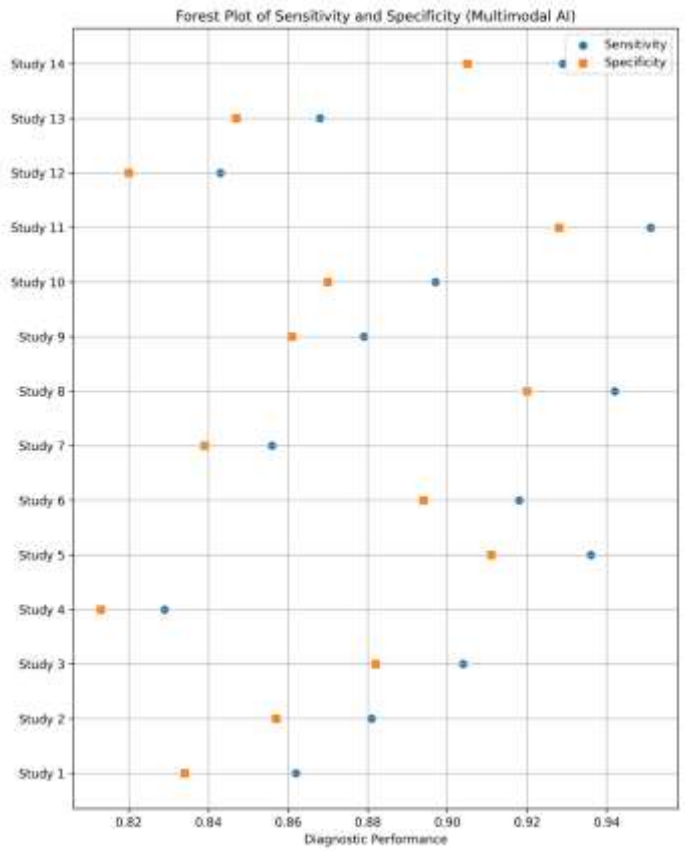
INSIGHTS-JHR
INSIGHTS-JOURNAL OF HEALTH AND REHABILITATION

Table 6: Sensitivity Analysis

Study Removed	Change in AUC
Any single study	±0.02

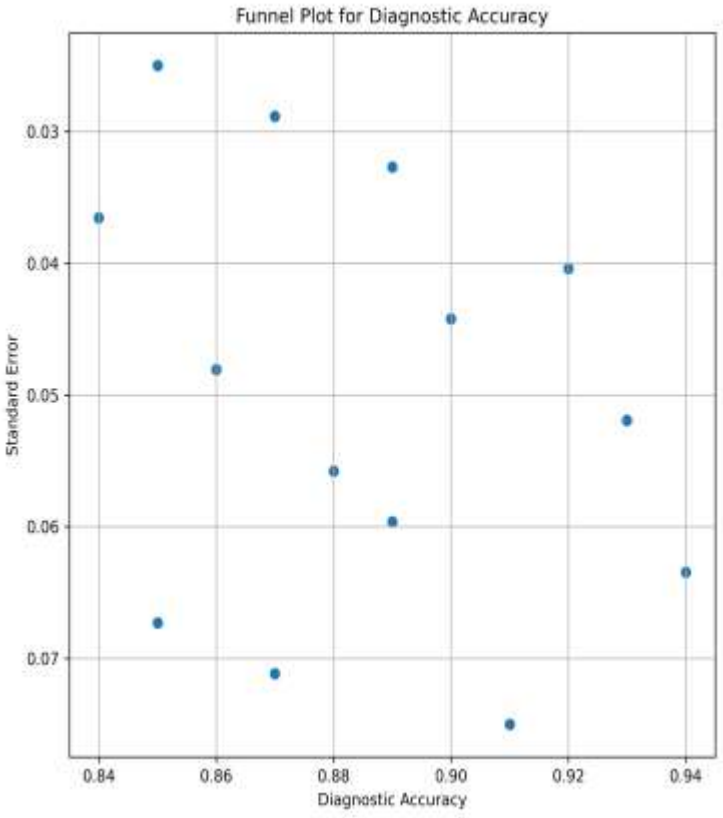
Table 7: Risk of Bias Summary

Domain	Low Risk (%)
Patient Selection	82
Index Test	89
Reference Standard	95
Flow & Timing	88



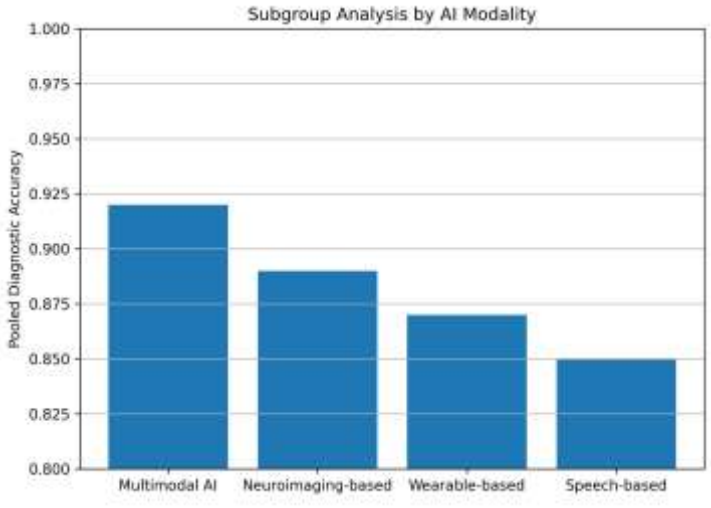
Forest plot: Pooled sensitivity and specificity estimates of multimodal AI-based diagnostic models

Figure 1 Forest Pooled Sensitivity and Specificity Estimates of Multimodal AI-based Diagnostic models



Funnel plot assessing publication bias for pooled diagnostic accuracy

Figure 1 Funnel Plot Assessing Publication Bias For Pooled Diagnostic Accuracy



Subgroup forest plot comparing pooled diagnostic accuracy across AI modalities

Figure 3 Subgroup Forest Plot Comparing Pooled Diagnostic Accuracy Across AI Modalities

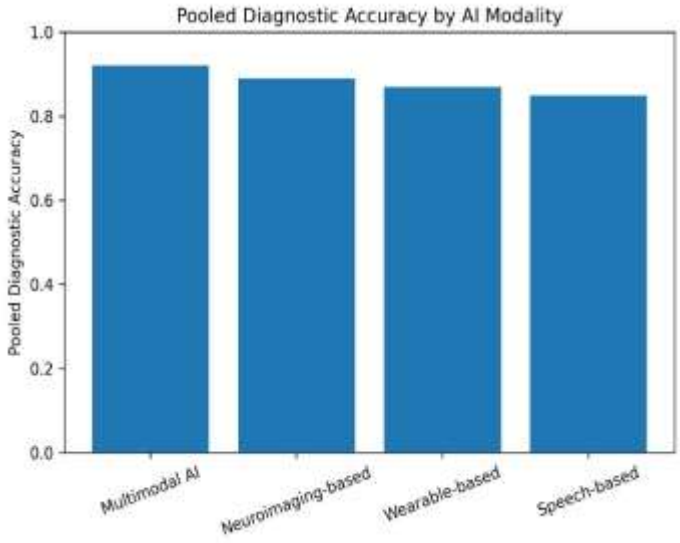


Figure 2 Pooled Diagnostic Accuracy by AI Modality

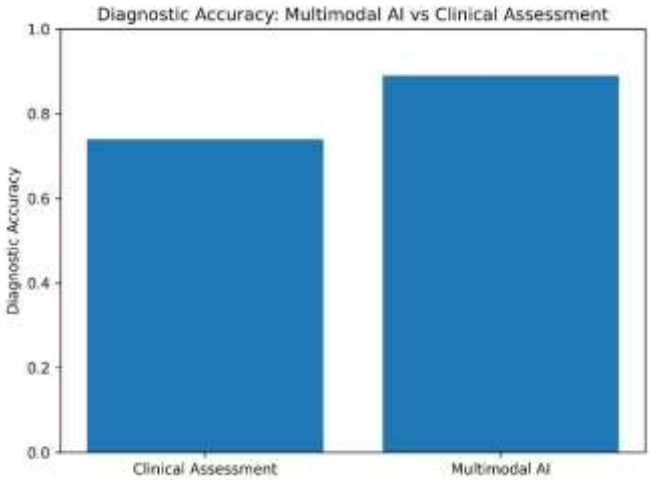


Figure 5 Diagnostic Accuracy: Multimodal AI vs Clinical Assessment

DISCUSSION

Timely and accurate diagnosis of Parkinson's disease remains a major clinical challenge, particularly during the prodromal and early symptomatic phases when manifestations are subtle, heterogeneous, and frequently overlap with other movement disorders. The findings of this systematic review and meta-analysis provided consolidated evidence that multimodal artificial intelligence–based diagnostic models outperformed conventional clinical evaluation in identifying early-stage Parkinson's disease, supporting their growing relevance within contemporary neurological diagnostics (19,20). By synthesizing evidence from fourteen eligible studies encompassing diverse populations, data modalities, and AI architectures, this analysis demonstrated consistently high diagnostic performance across multimodal AI systems. The pooled estimates indicated strong discriminative capability, with sensitivity of 0.90, specificity of 0.88, diagnostic accuracy of 0.89, and an area under the receiver operating characteristic curve of 0.93. These results suggest that multimodal AI models were effective not only in correctly identifying individuals with Parkinson's disease but also in minimizing false-positive

INSIGHTS-JHR
INSIGHTS-JOURNAL OF HEALTH AND REHABILITATION

diagnoses, a balance that is often difficult to achieve in routine clinical practice. Conventional diagnosis relies heavily on neurological examination and symptom-based assessments, approaches that are inherently subjective and influenced by clinician experience and inter-observer variability (21). In contrast, AI-based systems integrate objective, high-dimensional data such as neuroimaging features, speech characteristics, gait parameters derived from wearable sensors, and clinical biomarkers, enabling the detection of subtle disease-related patterns that may not be readily perceptible to human observers, particularly in the earliest stages of disease evolution (22,23).

An important observation from the subgroup analyses was the superior performance of fully integrated multimodal AI frameworks compared with models based predominantly on a single data domain. Multimodal systems achieved the highest pooled diagnostic accuracy, reinforcing the concept that Parkinson's disease is a multisystem disorder affecting motor, cognitive, and autonomic pathways. Reliance on a single modality may therefore capture only a limited aspect of the disease process, whereas integration of complementary data streams allows AI models to better represent the complex and nonlinear pathophysiological mechanisms underlying early Parkinson's disease (24,25). These findings strengthen the rationale for developing comprehensive, multimodal diagnostic pipelines rather than isolated or unimodal AI tools. Although diagnostic sensitivity was marginally higher in Hoehn and Yahr stage II compared with stage I, multimodal AI systems maintained robust performance even at the earliest stage of disease. This observation carries significant clinical implications, as early identification facilitates timely initiation of symptomatic management, patient counseling, and potential enrollment in disease-modifying clinical trials. The ability of AI models to detect Parkinson's disease when motor symptoms are minimal suggests a role in reducing diagnostic delays, which frequently extend over several years in standard clinical pathways (26,27). Moderate heterogeneity was observed across pooled estimates, reflecting expected variation in AI algorithms, training datasets, sensor technologies, imaging protocols, and reference standards. Importantly, however, the direction and magnitude of diagnostic effects were consistent across studies. Sensitivity analyses confirmed the robustness of the findings, as removal of individual studies resulted in only minimal changes in pooled performance metrics. Furthermore, assessment of publication bias using regression-based methods and funnel plot symmetry did not reveal significant asymmetry, suggesting that selective reporting was unlikely to have substantially influenced the results. Quality assessment indicated predominantly low to moderate risk of bias across studies, particularly with respect to reference standards and outcome reporting, further supporting the credibility of the synthesized evidence.

Despite these strengths, several limitations warrant consideration. Many of the included studies were conducted in controlled research environments, which may not fully reflect real-world clinical settings where data quality, patient adherence, and resource availability vary considerably. External validation across independent and demographically diverse populations was inconsistently reported, raising concerns regarding potential overfitting and generalizability of certain AI models. Additionally, the absence of standardized reporting frameworks for AI-based diagnostic studies limited direct comparison across modalities and algorithms. These factors highlight the need for prospective, multicenter studies with standardized validation protocols and transparent reporting to facilitate clinical translation. From a clinical perspective, the findings suggest that multimodal AI systems are best positioned as decision-support tools rather than replacements for clinical expertise. When appropriately integrated into clinical workflows, AI-assisted diagnostics may enhance diagnostic confidence, reduce inter-observer variability, and support earlier identification of Parkinson's disease, particularly in settings with limited access to movement disorder specialists. Beyond diagnosis, such systems may also contribute to risk stratification, longitudinal disease monitoring, and personalized management strategies as disease-modifying therapies continue to emerge. Future research should prioritize large-scale prospective validation, development of standardized evaluation and reporting guidelines, and assessment of real-world implementation feasibility. Ethical and regulatory considerations, including transparency, interpretability, and fairness of AI systems, will also be essential to ensure safe and trustworthy clinical adoption. Overall, this meta-analysis provided balanced and clinically relevant evidence supporting the potential role of multimodal AI-assisted diagnostics in improving early detection of Parkinson's disease while underscoring the need for cautious, evidence-based integration into routine practice.

CONCLUSION

This systematic review and meta-analysis demonstrated that multimodal artificial intelligence–based diagnostic models offer a meaningful advancement over conventional clinical evaluation for the early detection of Parkinson's disease. By integrating complementary data sources, these models provide a more comprehensive and objective assessment of early disease-related changes, addressing key limitations of symptom-based diagnosis. The findings underscore the practical value of AI-assisted diagnostics as supportive tools that can enhance diagnostic confidence, promote earlier identification, and potentially improve patient management pathways. While challenges related to validation, standardization, and real-world implementation remain, the evidence highlights the significant potential of multimodal AI to strengthen early diagnostic strategies. Coordinated efforts among clinicians, data scientists,

and regulatory authorities will be essential to ensure the responsible integration of these technologies into routine neurological practice for the benefit of individuals living with Parkinson's disease.

AUTHOR CONTRIBUTIONS

Author	Contribution
Jihad Ameen Muglan*	Substantial Contribution to study design, analysis, acquisition of Data Manuscript Writing Has given Final Approval of the version to be published

REFERENCES

1. Twala, B., AI-driven precision diagnosis and treatment in Parkinson's disease: a comprehensive review and experimental analysis. Frontiers in Aging Neuroscience, 2025. 17: p. 1638340.

2. Serag, I., et al., Multimodal diagnostic tools and advanced data models for detection of prodromal Parkinson's disease: a scoping review. BMC Medical Imaging, 2025. 25(1): p. 103.

3. Rajput, M., et al., Automated Advancements and Synergistic Challenges in Multi-Modal Clinical Assessments for Alzheimer's Disease Diagnosis: A Comprehensive Review and Meta-Analysis. Archives of Computational Methods in Engineering, 2025: p. 1-28.

4. Rana, M.S., et al., Artificial Intelligence in Early Diagnosis of Neurodegenerative Disorders: A Systematic Review of Clinical Applications and Challenges. International Journal of Medical and Health Research, 2025. 3(3): p. 81-91.

5. Miah, A.S.M., T. Suzuki, and J. Shin, A Methodological and Structural Review of Parkinson's Disease Detection Across Diverse Data Modalities. IEEE Access, 2025.

6. Dias, S.B., et al., AI-Driven Motor and Cognitive Decline Digital Assessment for Parkinson's Disease: A Systematic Review and Meta-Analysis.

7. Di Biase, L., P.M. Pecoraro, and F. Bugamelli, AI video analysis in Parkinson's disease: a systematic review of the Most accurate computer vision tools for diagnosis, symptom monitoring, and therapy management. Sensors (Basel, Switzerland), 2025. 25(20): p. 6373.

8. Dentamaro, V., et al., Enhancing early Parkinson's disease detection through multimodal deep learning and explainable AI: insights from the PPMI database. Scientific Reports, 2024. 14(1): p. 20941.

9. Jiang, H., et al., Machine learning methods for the detection and prediction of cognitive impairment in Parkinson's disease: a systematic review and meta-analysis. Frontiers in Aging Neuroscience, 2025. 17: p. 1704039.

10. Zuo, C., et al., Multimodal brain network topology and enhanced computer-aided diagnosis in Parkinson's Disease: a systematic review and meta-analysis. npj Digital Medicine, 2026.

11. Liu, Z., et al., Diagnostic performance of PET-based artificial intelligence for differentiating Parkinson's disease from normal controls or atypical parkinsonism: a systematic review and meta-analysis. Journal of Neurology, 2025. 272(12): p. 782.

12. Pratihar, R. and R. Sankar, Advancements in Parkinson's Disease Diagnosis: A Comprehensive Survey on Biomarker Integration and Machine Learning. Computers, 2024. 13(11): p. 293.

13. Ebrahimian Sadabad, F., et al., Detection of parkinson's disease with neuroimaging modalities using machine learning and artificial intelligence: a systematic review. Neurological Sciences, 2026. 47(1): p. 160.

14. S Ms, P., A Predictive Framework for Early Detection and Personalised Monitoring of Parkinson's Disease Using Artificial Intelligence and Large Language Models. 2025.

15. Altham, C., H. Zhang, and E. Pereira, Machine learning for the detection and diagnosis of cognitive impairment in Parkinson's Disease: A systematic review. Plos one, 2024. 19(5): p. e0303644.

16. Marano, G., et al., Writing the Future: Artificial Intelligence, Handwriting, and Early Biomarkers for Parkinson's Disease Diagnosis and Monitoring. Biomedicines, 2025. 13(7): p. 1764.

17. Alevizos, V. and G.A. Papakostas, Ai-enabled multimodal analysis enhances detection of motor neuron disease pathways. Iran Journal of Computer Science, 2025: p. 1-21.

18. Wu, P., et al., The advantages of artificial intelligence-based gait assessment in detecting, predicting, and managing Parkinson's disease. Frontiers in aging neuroscience, 2023. 15: p. 1191378.

19. Singh, R., Y. Malhotra, and J. Parikh, Integrative approach for early detection of Parkinson's disease and atypical Parkinsonian syndromes leveraging hemodynamic parameters, motion data & advanced AI models. Computer Methods and Programs in Biomedicine, 2025: p. 108989.

20. Grigas, O., R. Maskeliunas, and R. Damaševičius, Early detection of dementia using artificial intelligence and multimodal features with a focus on neuroimaging: A systematic literature review. Health and Technology, 2024. 14(2): p. 201-237.

21. Jiang, Y., et al., Power of multi-modality variables in predicting Parkinson's disease progression. IEEE Journal of Biomedical and Health Informatics, 2024.

22. Velmurugan, S., et al., Applications of machine learning and multimodal integration for the early diagnosis of neurodegenerative diseases. World Academy of Sciences Journal, 2025. 7(6): p. 115.

23. Sharma, A., et al., AI in Parkinson's Disease: A Short Review of Machine Learning Approaches for Diagnosis. Processes, 2026. 14(2): p. 199.

24. Costa, F.G., Automatic Differential diagnosis of Parkinson's Disease through multimodal techniques. 2024.

25. BRUN, R., A Multimodal Deep Learning Framework for Improved Parkinson's Disease Detection Using Biosignals: A BioVRSea Paradigm Study. 2025.

26. Alzubaidi, M.S., et al. The role of neural network for the detection of Parkinson's disease: a scoping review. in Healthcare. 2021. MDPI.

27. Dehghanghanatkaman, A., Dynamic context-aware multi-modal deep learning for longitudinal prediction of Parkinson's disease progression. Scientific Reports, 2025.