International Journal of Industrial Engineering
2024;8(2):48-56.
ISSN: 2456-8449

Cloud-Based Early Detection of Parkinson's Disease Using AI Tools in Amazon Web Services SageMaker

M. Muthulekshmi[1], S. Sujatha[1*]

[1]*Department of Biomedical Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences (SIMATS), Chennai, Tamil Nadu, India.*

[*]*Corresponding author: sujathasmvr@gmail.com*

Abstract. Parkinson's disease is a progressive neurological disorder that remains difficult to diagnose at early stages due to subtle symptom onset and reliance on subjective clinical assessments. Early identification is crucial to improving patient outcomes, reducing long-term healthcare costs, and enabling timely therapeutic interventions. This study proposes a cloud-based diagnostic framework leveraging Amazon Web Services (AWS) SageMaker to integrate diverse patient datasets including clinical records, speech samples, handwriting analysis, and wearable sensor data—for predictive modeling. The architecture incorporates an Extract, Transform, and Load (ETL) module for preprocessing, supervised machine learning models for training and validation, and a feedback loop for continuous refinement. Through interactive dashboards, healthcare professionals can access predictions and risk assessments in real time, supporting informed decision-making. Security and compliance are ensured through HIPAA-compliant encryption and role-based access control, while scalability is achieved via SageMaker's managed infrastructure. Quantitative evaluation confirms high predictive sensitivity and accuracy, with heatmap and bar chart analyses identifying high-risk individuals and time-series tracking demonstrating effective progression monitoring, establishing the framework's reliability in early Parkinson's detection.

Keywords: Early Detection, Parkinson's Disease, AI Tools, Amazon Web Services SageMaker, Cloud Computing

INTRODUCTION

Parkinson's disease, a progressive neurodegenerative disorder affecting millions worldwide, is often diagnosed at advanced stages, creating major challenges for healthcare delivery. Early identification is essential, as it improves patient outcomes, enables timely therapeutic interventions, and reduces long-term healthcare costs. Traditional diagnostic methods, which rely heavily on subjective assessments and visible symptoms, often delay detection and limit opportunities for early treatment. Cloud-based AI platforms, such as AWS SageMaker, provide scalable and accurate solutions for early diagnosis by analysing large-scale patient data to identify subtle disease markers and emerging trends. Leveraging advanced machine learning capabilities, AWS SageMaker enables the development of comprehensive diagnostic frameworks that integrate supervised learning, feature extraction, and real-time data processing. These AI-driven methods enhance diagnostic precision, improve scalability, and support healthcare professionals in delivering timely, effective, and personalised care for Parkinson's disease.

Guidelines for Parkinson's disease diagnosis and treatment in German-speaking countries are detailed in [1], developed collaboratively by German, Austrian, and Swiss Neurological Societies. These S2k guideline updates stress early detection, standardised diagnostics, and both pharmacological and invasive therapies, aiming to harmonise clinical practice and improve patient outcomes. Parkinson's disease as an environmentally driven disorder is discussed in [2], where factors such as pesticide exposure (paraquat, rotenone), industrial solvents (trichloroethylene), and air pollution are identified as major contributors, with genetic factors playing a comparatively modest role. Sleep disorders are recognised as early manifestations of Parkinson's disease in [3], with insomnia, REM sleep behaviour disorder, excessive daytime sleepiness, and restless legs syndrome linked to illness duration and autonomic dysfunction, pointing to biological rather than psychological origins.

Genetic determinants of Parkinson's disease are highlighted in [4], noting that pathogenic mutations in SNCA, LRRK2, PRKN, PINK1, DJ-1, and GBA1 play significant roles in onset and variability, particularly in familial cases. Broader perspectives on molecular, clinical, and therapeutic aspects are provided in [5], which outlines traditional dopaminergic therapies alongside advances in antidiabetic drugs, intranasal insulin, GLP-1 receptor agonists, stem cell therapies, and gene therapy. Refractory cases benefit from deep brain stimulation, lesioning,

Received: 10.04.2024 Revised: 20.05.2024 Accepted: 05.06.2024

and targeted ultrasound, while lifestyle interventions further support personalised care. Dysregulation of brain renin–angiotensin signalling is described in [6], where overactivation of the AngII/AT1 receptor axis drives oxidative stress, neuroinflammation, and α-synuclein aggregation, suggesting therapeutic potential for AT1 blockers and RAS enhancers.

Global genetic research is reviewed in [7], which identifies key Parkinson's disease genes and over 200 risk loci through genome-wide association studies, with GBA1 emerging as a common demographic risk factor. Predictive polygenic risk scores and the Global Parkinson's Genetics Program are advancing precision therapies. Longitudinal monitoring of disease progression is examined in [8] using data from the Parkinson's Progression Markers Initiative (PPMI), where milestone-based tracking identified cognitive decline, functional dependence, and autonomic dysfunction as early predictors. Mitochondrial dysfunction, a central hallmark of Parkinson's disease, is highlighted in [9], linking genetic mutations, environmental toxins, and α-synuclein aggregation to neuronal loss. Therapeutic strategies include antioxidants, antidiabetic drugs, and gene-based therapies, though clinical translation remains limited. Established and emerging therapeutic strategies are further reviewed in [10], emphasising the limitations of current symptom-focused treatments and the promise of disease-modifying interventions.

Neuroinflammatory biomarkers in blood and cerebrospinal fluid are systematically analysed in [11], where elevated IL-6, TNF-α, IL-1β, CRP, and CCL2, along with reduced IFN-γ and IL-4, highlight inflammation's role in disease progression and diagnostic potential. AI-driven diagnostic automation is presented in [12], employing multimodal data from speech, handwriting, imaging, and gait signals to enable accurate and flexible disease classification. Gut microbiome alterations are reported in [13], with large-scale sequencing identifying dysbiosis-associated microbes and functional pathways linking microbial imbalance to disease mechanisms. Genetic insights and novel therapeutic strategies are reviewed in [14], discussing neuronal replacement, α-synuclein-targeted interventions, and apomorphine for disease modification.

Multimodal machine learning approaches are explored in [15], where clinico-demographic, genomic, and transcriptomic data are integrated using GenoML to generate predictive models. Classifiers such as logistic regression, SVM, random forests, gradient boosting, and multilayer perceptrons are compared, with feature importance explained via Shapley Additive Explanations. Cholinergic circuit alterations are discussed in [16], with deficits across cortical, striatal, cerebellar, and autonomic networks contributing to motor and non-motor symptoms. Targeted muscarinic/nicotinic therapies and neuromodulation are emerging to address these deficits. A hybrid deep learning framework for severity assessment is introduced in [17], combining CNNs and Locally Weighted Random Forests to analyse gait data from wearable sensors, enabling non-invasive, cost-effective monitoring.

Advances in digital biomarkers for Parkinson's disease monitoring are discussed in [18], showing how wearable and digital devices provide objective indicators of disease progression and therapeutic response. Voice-based classification frameworks are introduced in [19], where acoustic features and spectrograms are analysed using Inception V3 CNNs, logistic regression, and random forest classifiers, enabling accurate distinction between Parkinson's patients and healthy controls. Together, these advances highlight the integration of clinical guidelines, genetics, environmental research, molecular insights, and AI-driven approaches in shaping the future of Parkinson's disease diagnosis and treatment.

PROPOSED METHODOLOGY

The proposed system leverages AWS SageMaker AI technologies to deliver a cloud-based platform for early Parkinson's disease detection. Parkinson's disease, a progressive neurodegenerative disorder, affects both motor and cognitive functions, making timely diagnosis critical. Figure 1 presents a block diagram of the system architecture, illustrating how AWS SageMaker integrates diverse data sources such as clinical records, speech patterns, handwriting samples, and motor movement assessments. Once uploaded to AWS cloud storage, these datasets undergo cleaning and organization through an Extract, Transform, and Load (ETL) module to ensure data quality and consistency. The processed and integrated data is then stored centrally, providing a structured foundation for machine learning algorithms to analyze and generate predictive insights for early disease identification.

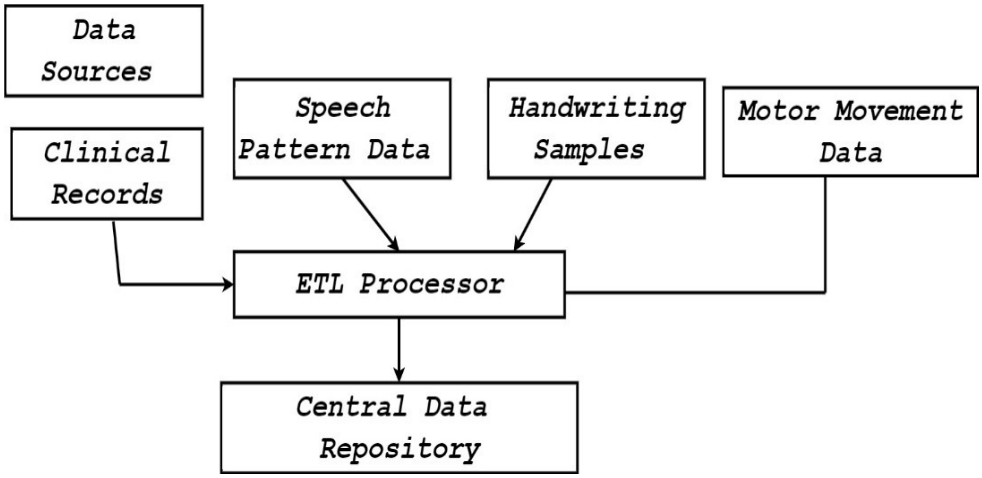

FIGURE 1. Block Diagram: Data Collection and Integration for Parkinson's Detection

An effective early detection system relies on the availability of high-quality and diverse data. The process begins with the collection of multiple datasets that capture Parkinson's disease symptoms and progression patterns. Examples include patient health records, voice recordings, wearable sensor outputs, and gait analysis video data. Figure 2 illustrates the AWS SageMaker model training and evaluation workflow. Once the data is integrated, it is stored in a central repository and used to train machine learning algorithms within SageMaker. Through this process, the model learns to identify subtle indicators of early Parkinson's disease. Following training, the model undergoes rigorous evaluation to measure accuracy and sensitivity. The validated model is then deployed to generate real-time predictions, providing healthcare practitioners with early-warning insights for timely intervention. Moreover, the system supports continuous improvement, as new data is incorporated to retrain and refine the model, thereby enhancing predictive performance over time.

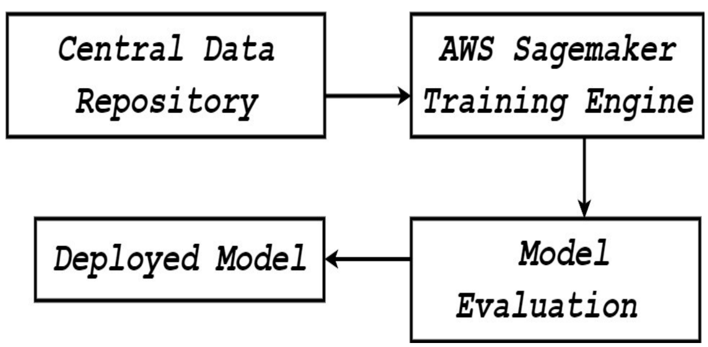

FIGURE 2. Block Diagram: Model Training and Evaluation in AWS SageMaker

AWS SageMaker's AI tools enable the development of machine learning models from prepared healthcare data, offering a wide range of algorithms and frameworks suitable for predictive modeling tasks. Supervised learning models such as decision trees, support vector machines (SVMs), and random forests can be applied to identify early signs of Parkinson's disease with high reliability. Figure 3 presents a block diagram illustrating how healthcare practitioners interact with the early detection system. Through an interactive dashboard, providers can review analytic results, evaluate prediction accuracy, and gain actionable insights into patient conditions. Importantly, the system incorporates feedback from practitioners to refine and improve the model, ensuring it remains accurate and aligned with clinical findings. Integration with AWS SageMaker enables real-time feedback loops and adaptive learning, allowing the system to evolve continuously. This iterative process enhances prediction accuracy and sensitivity over time, making the framework robust and clinically relevant.

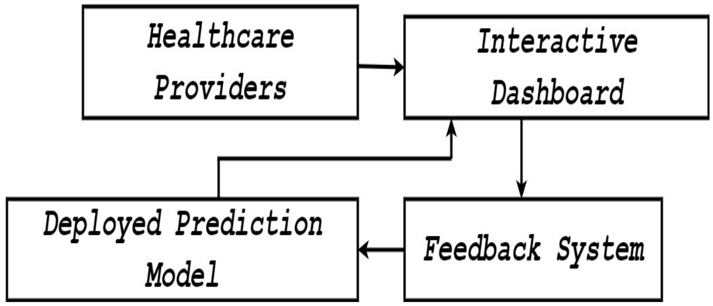

FIGURE 3. Block Diagram: User Interaction and Feedback System

After selecting the most suitable machine learning models, training is carried out using the processed datasets. AWS SageMaker provides a fully managed training environment with scalable infrastructure, enabling efficient model training without the complexity of manual setup or administration. Figure 4 illustrates the end-to-end workflow of the AWS SageMaker–based Parkinson's disease detection system.

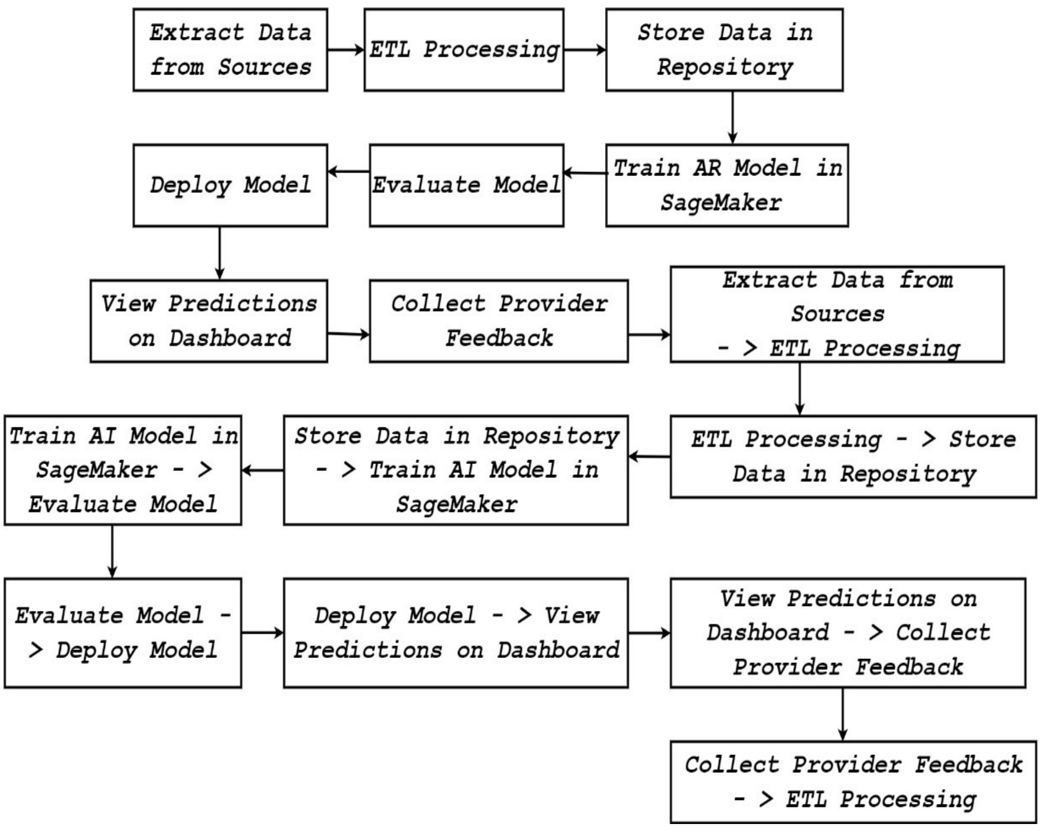

FIGURE 4. Data Flow Diagram: End-to-End Detection Data Flow

The process begins with data ingestion from multiple sources, followed by ETL processing to ensure clean and structured datasets. Machine learning algorithms are then applied during the training phase to identify patterns associated with early Parkinson's disease. Once trained, the model undergoes evaluation and deployment, after which healthcare providers can access predictions and insights through an interactive dashboard. Finally, practitioner feedback is incorporated to refine the model continuously, ensuring ongoing improvement in accuracy and clinical relevance. Following training and validation, machine learning models are deployed for real-time applications. AWS SageMaker simplifies this process by providing scalable endpoints for seamless model deployment. Figure 5 illustrates the complete architecture of the AWS SageMaker–based Parkinson's disease

International Journal of Industrial Engineering
2024;8(2):48-56.
ISSN: 2456-8449

detection system. In this framework, a central repository stores data collected from multiple sources, which is then processed through an ETL layer to ensure consistency and reliability. The processed data is used both for model training and for generating predictions, which are presented to healthcare providers through an interactive dashboard. A continuous feedback loop with practitioners enables model refinement, ensuring that the system remains accurate, responsive, and clinically relevant. This architecture not only supports early disease detection but also ensures scalability and adaptability, making it suitable for diverse healthcare contexts.

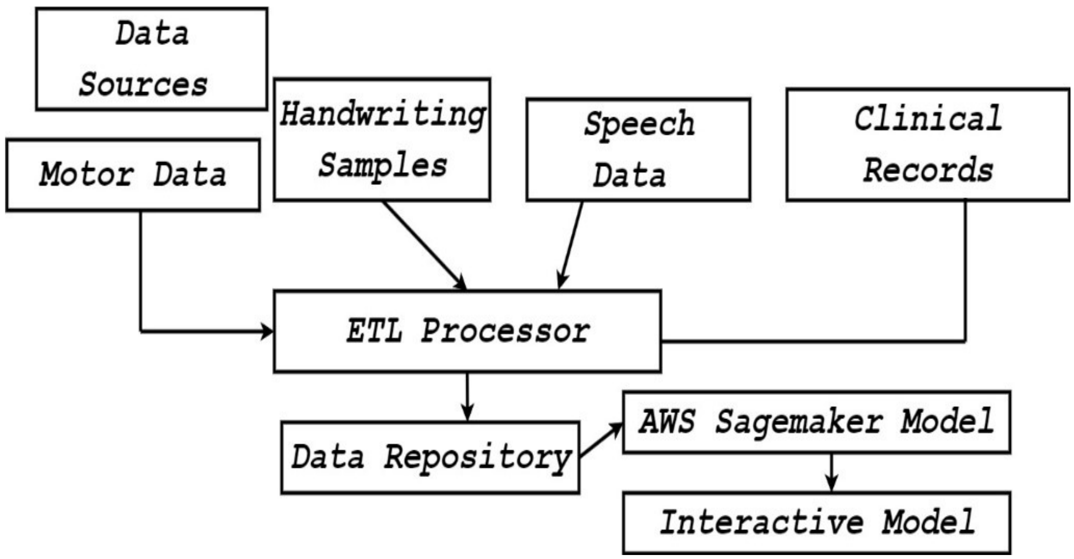

FIGURE 5. Overview Diagram: System Architecture Overview

Continuous monitoring and improvement are fundamental to the proposed system. Incremental learning enables the machine learning model to be updated whenever new patient data becomes available. AWS SageMaker facilitates this process by allowing models to be retrained with fresh data without the need to restart training from the beginning, ensuring that they remain aligned with the latest clinical insights. As data volume and variability increase, the feedback loop progressively enhances system accuracy. In addition, real-time monitoring of model performance helps detect any decline in predictive quality, automatically triggering a retraining cycle. This adaptive approach ensures the long-term effectiveness of the system and strengthens its capability for early detection of Parkinson's disease.

The system should incorporate an intuitive and user-friendly interface that enables healthcare professionals to effectively apply AI techniques and interpret predictive results. This interface will display forecast outcomes along with confidence ratings, supporting clinicians in making informed decisions at various diagnostic stages. Through an interactive online dashboard, users will be able to input new patient data, review historical prediction trends, and explore correlations between biomarkers and disease progression. Beyond visualization, the system's reporting functionality will generate actionable insights derived from AI-driven forecasts. These reports will include the likelihood of Parkinson's disease, associated confidence levels, symptom analyses, and suggested clinical interventions, thereby enhancing decision-making and supporting personalized patient care.

RESULTS AND DISCUSSIONS

Given the sensitivity of medical data, cloud-based AI systems must prioritize privacy and security. AWS ensures data protection through robust measures such as encryption, access controls, and audit logging, all in compliance with HIPAA standards. Role-Based Access Control (RBAC) further restricts access to sensitive data, ensuring that only authorized users can interact with protected information. Data transmitted to and from the cloud is encrypted both in transit and at rest, significantly reducing the risk of breaches. Beyond security, clinical validation is essential before large-scale deployment to confirm the system's accuracy, usability, and applicability in real-world settings. This requires close collaboration with healthcare institutions for pilot testing, where the system's performance on actual patient data can be evaluated and refinements suggested. Neurologists and Parkinson's disease specialists play a key role in improving predictive capabilities and integrating the system into clinical workflows. Clinical validation thus ensures both the efficacy and safety of the early AI-driven detection system.

The process begins with the aggregation of patient health records from multiple sources, ensuring a comprehensive dataset as input. Data preparation then involves cleaning and normalizing the datasets to enhance consistency and enable the model to identify meaningful patterns. Feature extraction focuses on isolating critical indicators—such as motor impairments and speech abnormalities—that are essential for accurate Parkinson's diagnosis. Using AWS SageMaker's scalable, cloud-based infrastructure, deep learning models are trained on these large and complex datasets to detect subtle signs of disease onset. The final predictive model provides high-sensitivity early symptom detection alerts, enabling timely medical intervention. By integrating these components into a single framework, the system delivers a comprehensive, scalable, and cost-effective solution for early Parkinson's disease detection, ultimately supporting prompt treatment and improved patient outcomes.

For long-term success, the proposed system must remain adaptive to advancements in artificial intelligence, healthcare practices, and patient care methodologies. Regular updates will be required to integrate the latest scientific research, emerging machine learning models, and evolving healthcare technologies. Collaboration with academic and clinical institutions can further enhance innovation and improve predictive accuracy. Such integration will also enable physicians to explore correlations between symptom patterns and long-term disease progression, facilitating more personalized treatment strategies. Figure 6 presents a bar chart of patients' speech, handwriting, and balance symptom ratings, illustrating variations in severity that help identify high-risk individuals. For example, Patient P04 demonstrates elevated motor and handwriting scores, signaling an increased risk level. These symptom profiles can be used to train AWS SageMaker models for predicting Parkinson's onset, while the visualization itself supports clinicians in monitoring disease progression and detecting emerging complications.

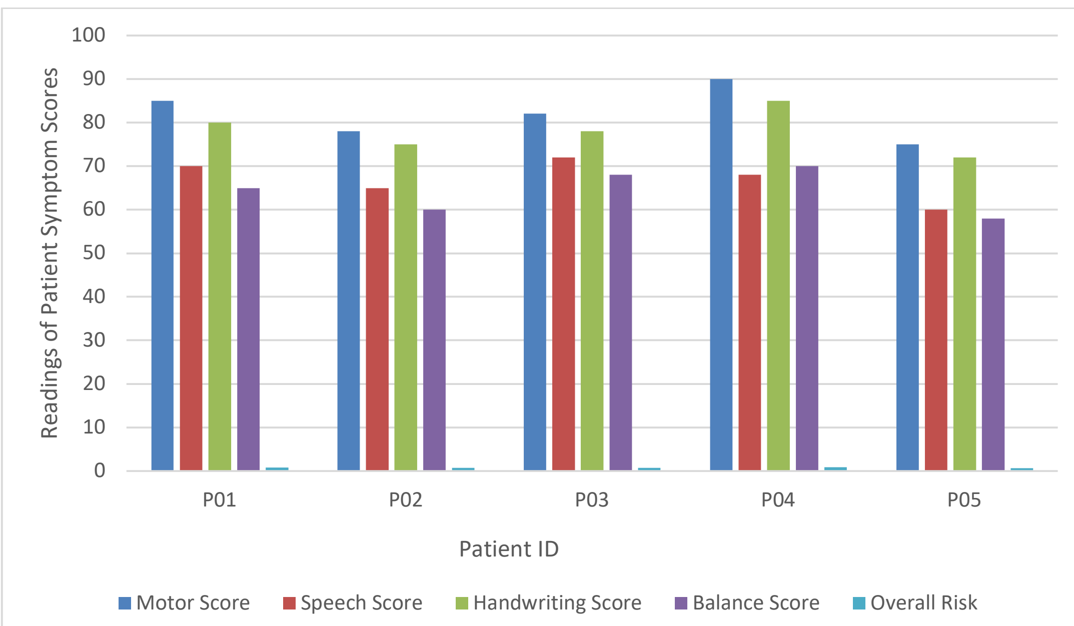

FIGURE 6. Bar Chart of Patient Symptom Scores

The key challenges include unbalanced datasets, limited availability of labeled medical data, and concerns regarding patient privacy. Despite these obstacles, the system demonstrates significant potential by enabling disease progression tracking, supporting remote monitoring, and assisting clinical decision-making. However, scalability may be constrained by the large datasets and high computational requirements involved. Looking ahead, future developments aim to address these limitations through the integration of IoT devices for real-time data collection, the design of cost-efficient AI models, and the implementation of secure data encryption strategies. Collectively, these advancements will strengthen the system's diagnostic capabilities and management effectiveness, ultimately support personalized treatment strategies and improve diagnostic accuracy for Parkinson's disease.

The successful implementation of the proposed system requires strong collaboration across the healthcare ecosystem. Integration with electronic health record (EHR) systems, telemedicine platforms, and patient monitoring applications can support a holistic strategy for managing Parkinson's disease. Embedding the AI-powered early detection framework into clinical workflows ensures a seamless experience for healthcare providers, reducing the resistance often associated with adopting new technologies. Partnerships with research institutions and Parkinson's disease advocacy groups can further enhance the system's therapeutic value by contributing to more diverse datasets, enabling the model to learn from a wider range of patient cases and improving predictive accuracy. Collaboration with clinicians, neurologists, and researchers ensures that the system's output remains practical, clinically relevant, and aligned with real-world treatment needs. Figure 7 illustrates a line chart depicting each patient's Parkinson's disease risk progression over time.

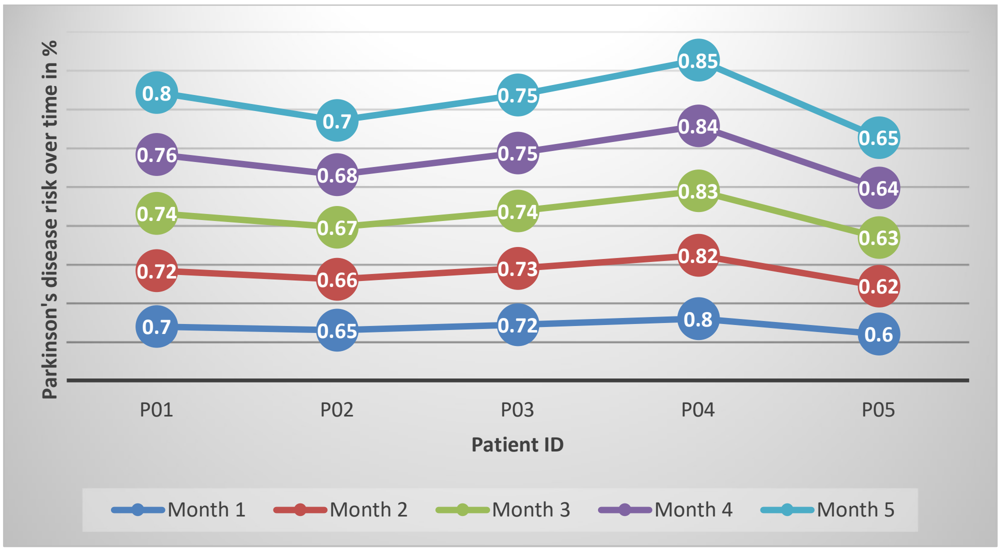

FIGURE 7. Line Chart of Patient Risk Over Time

Tracking individual risk trajectories highlights symptom development and overall disease progression. For instance, Patient P04 demonstrates a steady increase in risk levels, indicating worsening symptoms. Using AWS SageMaker's time-series analysis, the system can forecast future risk trends, enabling clinicians to closely monitor patients and intervene promptly. This visualization supports the identification of progression patterns and facilitates timely adjustments to treatment strategies.

Early identification of Parkinson's disease through this cloud-based approach has the potential to significantly improve public health outcomes. Detecting the condition in its initial stages enables timely interventions, which can delay the onset of severe symptoms and enhance patients' quality of life. Proactive management may also reduce hospitalizations, emergency care requirements, and prolonged treatment costs, ultimately lowering long-term healthcare expenditures. Furthermore, the system offers a cost-effective and scalable solution that can democratize early detection and intervention, particularly in underserved or resource-limited regions. By extending accessibility to both urban and rural populations, the approach has the capacity to reduce healthcare inequities. The integration of multimodal data—such as voice recordings, wearable sensor outputs, medical imaging, and cognitive assessments—further strengthens the system's accuracy and resilience. This holistic perspective enables a deeper understanding of Parkinson's disease at both individual and population levels, fostering more precise diagnostics and more effective public health strategies.

CONCLUSION

The proposed AWS SageMaker–based system demonstrates the viability of cloud-enabled artificial intelligence for early Parkinson's disease detection and monitoring. By integrating multimodal datasets, applying supervised learning models, and employing continuous feedback loops, the framework enhances predictive accuracy, scalability, and clinical relevance. Experimental outcomes, including heatmap-based symptom profiling and time-series progression tracking, validate its ability to identify high-risk patients and forecast disease development. Despite its promise, several challenges remain: unbalanced datasets, high computational demands, and data privacy concerns limit scalability and clinical adoption. Addressing these issues will require expanded datasets, incorporation of IoT-driven real-time monitoring, and deployment of lightweight, explainable AI models that maintain transparency for clinical decision-making. Future work should also emphasize large-scale clinical validation through collaborations with healthcare institutions to ensure robustness in real-world environments. By overcoming these barriers, the proposed system has the potential to democratize early Parkinson's detection, reduce diagnostic inequities, and enable more personalized, timely, and cost-effective care strategies worldwide.

REFERENCES

[1]. G. Höglinger, and C. Trenkwalder, 2024, "Diagnosis and treatment of Parkinson´ s disease (guideline of the German Society for Neurology)," *Neurological Research and Practice*, 6(1), Article. 30.

[2]. E. R. Dorsey, and B. R. Bloem, 2024, "Parkinson's disease is predominantly an environmental disease," *Journal of Parkinson's Disease*, 14(3), pp. 451-465.

[3]. P. Dodet, M. Houot, S. Leu-Semenescu, J. C. Corvol, S. Lehéricy, G. Mangone, M. Vidailhet, E. Roze, and I. Arnulf, 2024, "Sleep disorders in Parkinson's disease, an early and multiple problem," *NPJ Parkinson's Disease*, 10(1), Article. 40.

[4]. S. Y. Lim, and C. Klein, 2024, "Parkinson's disease is predominantly a genetic disease," *Journal of Parkinson's Disease*, 14(3), pp. 467-482.

[5]. M. Muleiro Alvarez, G. Cano-Herrera, M. F. Osorio Martinez, J. Vega Gonzales-Portillo, G. R. Monroy, R. Murguiondo Perez, J. A. Torres-Ríos, X. A. van Tienhoven, E. M. Garibaldi Bernot, F. Esparza Salazar, and A. Ibarra, 2024, "A comprehensive approach to Parkinson's disease: addressing its molecular, clinical, and therapeutic aspects," *International Journal of Molecular Sciences*, 25(13), Article. 7183.

[6]. J. L. Labandeira-Garcia, C. M. Labandeira, M. J. Guerra, and A. I. Rodriguez-Perez, 2024, "The role of the brain renin-angiotensin system in Parkinson s disease," *Translational Neurodegeneration*, 13, Article. 22.

[7]. M. Funayama, K. Nishioka, Y. Li, and N. Hattori, 2023, "Molecular genetics of Parkinson's disease: Contributions and global trends," *Journal of Human Genetics*, 68(3), pp. 125-130.

[8]. M. C. Brumm, A. Siderowf, T. Simuni, E. Burghardt, S. H. Choi, C. Caspell-Garcia, L. M. Chahine, B. Mollenhauer, T. Foroud, D. Galasko, and K. Merchant, 2023, "Parkinson's progression markers initiative: a milestone-based strategy to monitor Parkinson's disease progression," *Journal of Parkinson's Disease*, 13(6), pp. 899-916.

[9]. M. T. Henrich, W. H. Oertel, D. J. Surmeier, and F. F. Geibl, 2023, "Mitochondrial dysfunction in Parkinson's disease–a key disease hallmark with therapeutic potential," *Molecular Neurodegeneration*, 18(1), Article. 83.

[10]. T. Pardo-Moreno, V. García-Morales, S. Suleiman-Martos, A. Rivas-Domínguez, H. Mohamed-Mohamed, J. J. Ramos-Rodríguez, L. Melguizo-Rodríguez, and A. González-Acedo, 2023, "Current treatments and new, tentative therapies for Parkinson's disease," *Pharmaceutics*, 15(3), Article. 770.

[11]. Y. Qu, J. Li, Q. Qin, D. Wang, J. Zhao, K. An, Z. Mao, Z. Min, Y. Xiong, J. Li, and Z. Xue, 2023, "A systematic review and meta-analysis of inflammatory biomarkers in Parkinson's disease," *NPJ Parkinson's Disease*, 9(1), Article. 18.

[12]. Z. Ayaz, S. Naz, N. H. Khan, I. Razzak, and M. Imran, 2023, "Automated methods for diagnosis of Parkinson's disease and predicting severity level," *Neural Computing and Applications*, 35(20), pp. 14499-14534.

[13]. Z. D. Wallen, A. Demirkan, G. Twa, G. Cohen, M. N. Dean, D. G. Standaert, T. R. Sampson, and H. Payami, 2022, "Metagenomics of Parkinson's disease implicates the gut microbiome in multiple disease mechanisms," *Nature Communications*, 13, Article. 6958.

[14]. O. Uwishema, H. Onyeaka, R. Badri, A. N. Yücel, A. K. Korkusuz, A. O. Ajagbe, A. Abuleil, C. Chaaya, B. H. Alhendawi, and E. Chalhoub, 2022, "The understanding of Parkinson's disease through genetics and new therapies," *Brain and Behavior*, 12(5), Article. e2577.

[15]. M. B. Makarious, H. L. Leonard, D. Vitale, H. Iwaki, L. Sargent, A. Dadu, I. Violich, E. Hutchins, D. Saffo, S. Bandres-Ciga, and J. J. Kim, 2022, "Multi-modality machine learning predicting Parkinson's disease," *NPJ Parkinson's Disease*, 8(1), Article. 35.

[16]. N. I. Bohnen, A. J. Yarnall, R. S. Weil, E. Moro, M. S. Moehle, P. Borghammer, M. A. Bedard, and R. L. Albin, 2022, "Cholinergic system changes in Parkinson's disease: emerging therapeutic approaches,"

The Lancet Neurology, 21(4), pp. 381-392.

[17]. T. Aşuroğlu, and H. Oğul, 2022, "A deep learning approach for Parkinson's disease severity assessment," *Health and Technology*, 12, pp. 943-953.

[18]. Y. M. Sun, Z. Y. Wang, Y. Y. Liang, C. W. Hao, and C. H. Shi, 2024, "Digital biomarkers for precision diagnosis and monitoring in Parkinson's disease," *NPJ Digital Medicine*, 7, Article. 218.

[19]. A. Iyer, A. Kemp, Y. Rahmatallah, L. Pillai, A. Glover, F. Prior, L. Larson-Prior, and T. Virmani, 2023, "A machine learning method to process voice samples for identification of Parkinson's disease," *Scientific Reports*, 13, Article. 20615.