The Bioscan
AN INTERNATIONAL QUARTERLY JOURNAL OF LIFE SCIENCES
20(4): 1703-1712, 2025 www.thebioscan.com

A Comprehensive Survey on Data Mining–Driven Approaches for Parkinson's Disease Diagnosis and Analysis

Gayathri C[1], Dr.R.Sankarasubramanian[2]

[1]Ph.D. Research Scholar, Department of Computer Science, Erode Arts and Science College, Erode-638 009, Tamilnadu, India. E-Mail: gayathrimca93@gmail.com
[2]Principal, Erode Arts and Science College, Erode-638 009, Tamilnadu, India.
E-Mail: rsankarprofessor@gmail.com

DOI: 10.63001/tbs.2025.v20.i04.pp1703-1712

KEYWORDS
Parkinson's disease diagnosis, medical data mining techniques, machine learning classification models, feature selection and dimensionality reduction, voice and gait signal analysis, biomedical pattern recognition systems, intelligent healthcare decision support, neurodegenerative disease prediction models.

Received on:

11-10-2025

Accepted on:

18-11-2025

Published on:

26-12-2025

ABSTRACT

Parkinson Disease (PD) is a neurodegenerative progressive disorder that denotes motor deficits like tremor, rigidity, and bradykinesia and non-motor deficits like impaired speech and cognitive deterioration. Data mining approaches have attracted considerable interest in the recent years towards early diagnosis, severity, and progression monitoring of PD in the light of their capability to identify discriminative patterns in high-dimensional and heterogeneous biomedical data. This survey is a systematic review of data mining-based methods used to analyse PD among various data modalities, and these are: clinical records, voice signals, handwriting dynamics, gait parameters, neuroimaging data. SVM methods, random forests, k-nearest neighbors and neural networks are the most commonly used forms of supervised learning, which have been used to classify PD with impressive diagnostic accuracy on benchmark datasets. The methods of clustering and association rule mining that have been applied unsupervised have been used to discover latent symptom patterns, disease subtypes. Dimensionality reduction techniques (e.g. Principal Component Analysis) and feature selection techniques (e.g. mutual information and evolutionary optimization) are important to improve the generalization and interpretation of the models. Recent works are growing to combine deep learning with data mining pipelines to create feature extraction and model intricate temporal dependencies automation. Though such developments are made, there are issues with data imbalance, interpretability, inadequate longitudinal data, and clinical generalizability. This survey identifies the up-to-date trends, comparative, and open research issues, which will serve as a unified source of future research based on data mining on PD.

Introduction

Parkinson disease (PD) is considered to be one of the most common chronic neurodegenerative disorders; the disease is mostly associated with the motor system, as a result of progression in the destruction of dopaminergic neurons in the substantia nigra. PD has motor symptoms of resting tremor, muscular rigidity, postural instability and bradykinesia, but also non-motor symptoms of speech disorder, sleep disturbances, depression and impaired cognitive functions. The diagnosis is also complicated at the initial stages as the early symptoms are not always obvious and are subjective and overlap with other neurological disorders. As a result, the need to have computationally basis diagnostic frameworks, which will guide clinicians in making objective decisions based on facts, is on the increase [1].

Conventional PD diagnosis is based on clinical observation and neurological experience, thus, making it subjective and inter-observer [2]. Furthermore, modern sensing and monitoring technologies create biomedical data that is complex, non-linear and of high dimensions, and thus difficult to manage using conventional statistical techniques [3]. In this respect, data mining has become an influential paradigm to unveil obscure patterns, correlations and predictive information of huge medical data. The data mining will allow predicting Pd on a larger scale, identifying its severity, and forecasting outcomes of the disease with a higher degree of accuracy and scalability by utilizing machine learning, pattern recognition, and knowledge discovery methods.

The presence of different PD-related datasets has greatly boosted the data mining studies in this field. These data are voice records that record dysphonia, handwriting and spiral drawing records that indicate motor control impairments, gait sensor records, which indicate postural instability and neuroimaging records that indicate structural and functional brain abnormalities. Data mining methods especially are very successful in combing these heterogeneous data sources to create strong predictive models. The main part is played by feature engineering in which statistical, spectral, temporal, and nonlinear features are obtained to describe disease-specific biomarkers.

The use of supervised classification algorithms prevails in the field of PD diagnostics as labeled datasets are present. Support Vector Machines, Decision Trees, Random Forests, Naive Bayes and Artificial Neural Networks are some of the techniques that have been shown to perform well in distinguishing PD patients and healthy controls. These techniques are normally used along with feature selection techniques to minimize redundancy, overfitting, and enhance computation efficiency [4]. Filter-based, wrapper-based and hybrid feature selection methods are used commonly, often with evolutionary and swarm intelligence methods to optimize them.

Uncontrolled forms of data mining can also be used in PD research through facilitation of exploratory analysis in the absence of initial classes [5]. The clustering methods are used to determine disease subtypes, stages in disease progression, and groups of patients which are useful in planning individualized treatment. The association rule mining is applied to reveal co-occurring pattern of symptoms and associations between clinical characteristics. This knowledge discovery means that the interpretability of such discoveries is augmented and that the ultimate results are not only classification results [6].

Over the past few years, there is a trend of using deep learning in conjunction with data mining pipelines in the analysis of PD. Deep architectures, such as Convolutional Neural Networks and Recurrent Neural Networks, enable automatic learning of features via raw signal inputs, e.g. speech waveforms, sensor time-series data, etc. Deep learning models enhance the power of representation and maintain the flexibility of analysis when applied together with traditional data mining methods [7]. These models, however, have associated issues that are connected with data demands, the cost of computation, and explainability that are important factors in health care.

Although great progress has been made, there are still a number of challenges that are present in data mining-based PD research [8]. The factors that influence the model generalization include limited sample sizes, imbalance in classes, noise, and variability in datasets. Comparative analysis is further complicated by the fact that there are no standardized evaluation procedures and clinically verified benchmarks. Besides, most of the studies are done on binary classification with insufficient consideration of disease staging or longitudinal disease progression modeling. The interpretability and reliability of the data mining models have been noted to be necessary to make them clinical, which requires transparent and explainable frameworks [9].

To conclude, data mining has been embraced as a pillar of the computational research on the study of Parkinsonism, and it provides appropriate tools to aid in diagnosis, prognosis, and knowledge discovery. Data mining methods can be used to find solutions to complex biomedical challenges with scalability, objectivity, and reproducibility, adding to clinical expertise. This introduction provides motivation, scope and importance of the data mining techniques in PD analysis and provides the basis of more sophisticated methodological development and translational research in intelligent healthcare systems [10].

Related Works

Parkinson's disease (PD) is a progressive neurodegenerative condition that entails motor deficits including tremor, rigidity, bradykinesia and postural instability as well as non-motor symptoms, with a broad spectrum, comprising speech impairment, cognitive, sleep disorders, and affective disorders. The high rate of PD cases across the world has grown over time owing to aging of population, thus the need to ensure that PD is detected in early stages and followed throughout is essential to prevent clinical complications. The traditional methods of diagnosis however are mostly based on clinical observation and rating scales including the Unified Parkinson’s Disease Rating Scale (UPDRS), which are subjective and tend to miss some subtle disease patterns at earlier stages. This shortcoming has inspired the incorporation of computational intelligence into the PD diagnosis and analysis.

Over the past few years, data mining has become a central computational paradigm to infer clinically relevant patterns of large-scale and heterogeneous biomedical data related to Parkinson disease. The innovative capabilities of sensing technologies and digital health platforms have allowed the generation of a wide range of modalities of PD, such as voice recordings, gait and foot-pressure, handwriting, wearable sensor streams, neuroimaging, and electronic health records. The data mining methods namely classification, clustering, feature selection, and pattern discovery are especially useful in processing high-dimensional, noisy, and non-linear data, thus, automated detection of diseases, estimation of their severity and motor subtyping. It has been shown that machine learning and hybrid data mining models can attain high diagnostic accuracy by being able to identify discriminative biomarkers that cannot be easily identified using conventional statistical methods.

Furthermore, feature selection, ensemble learning, and optimization-based data mining approaches have been combined to enhance the robustness, interpretability and cross dataset generalization of models. Recent developments also point to the increased use of deep learning-supported data mining pipelines in order to model more intricate temporal and spatial relationships in PD data. These developments notwithstanding, issues like imbalance in the data, shortage of longitudinal datasets, cross cohort variability and clinical interpretability prevail. The systematic analysis of data mining-based strategies is thus necessary to organize the current knowledge base, recognize gaps in methods and directions of future research to develop credible, scalable and clinically implementable systems to diagnose Parkinson diseases.

Table 1. Summary of Data Mining–Based Parkinson's Disease Literature

Ref.	Authors (Year)	Data Modality	Data Mining / ML Technique	Key Contribution	Limitations
[11]	Jiao et al. (2025)	Dynamic foot pressure	Clustering + Feature Selection + Classifiers	Improved PD classification using pressure distribution patterns	Limited generalization across cohorts
[12]	da Silva et al. (2025)	Multi-dataset clinical data	ML models with EDA	Web-based exploratory and classification framework	Lacks advanced optimization
[13]	Hussain et al. (2023)	Voice signals	Ensemble classifiers + Data augmentation	Enhanced robustness against data scarcity	Increased computational overhead
[14]	Chandru et al. (2025)	Multi-modal biomedical data	Artificial Intelligence survey	Comprehensive overview of AI in PD and AD	No quantitative comparison
[15]	Hema et al. (2023)	UCI PD dataset	Multiple feature selection + classifiers	Demonstrated impact of feature selection on accuracy	Binary classification focus
[16]	Petit et al. (2025)	Administrative health records	Data mining risk analysis	Large-scale population-level PD risk assessment	Limited clinical biomarkers
[17]	Haroon & Padma (2024)	Behavioral and clinical data	Ensemble learning + PCA	Improved diagnostic accuracy via dimensionality reduction	PCA reduces interpretability
[18]	Raza et al. (2024)	Behavioral data	Rough set–based hybrid mining	Effective handling of uncertainty and incompleteness	Sensitive to parameter tuning
[19]	Martinez-Eguiluz et al. (2023)	Non-motor symptom data	ML classifiers	Highlighted importance of non-motor features	Excludes motor signals
[20]	Yadav et al. (2023)	Clinical attributes	Supervised ML algorithms	Comparative analysis of classifiers	Small dataset size
[21]	Srinivasan et al. (2024)	Multi-class clinical data	Multi-class ML framework	Disease stage–aware classification	Limited longitudinal modeling

[22]	Hosseini et al. (2025)	Neuroimaging (Radiomics)	Relationship-based feature mining	Motor subtype identification	High computational complexity
[23]	Dawood (2023)	Brain disease datasets	ANN-based data mining	Generalized framework for neurological disorders	PD-specific tuning absent
[24]	Eliwa & Abd El-Hafeez (2025)	Clinical PD data	PSO-optimized classifiers	Improved prediction through optimization	Convergence instability
[25]	Allayith et al. (2024)	Multi-source PD data	Data mining survey	Consolidated PD mining techniques	Lacks future modeling insights
[26]	Majed et al. (2023)	Voice and clinical data	Comparative DM models	Performance benchmarking	No hybrid framework
[27]	Velu & Jaisankar (2025)	Clinical and signal data	Early prediction ML model	Focus on early-stage PD detection	Dataset imbalance
[28]	Peng et al. (2025)	Short-term motor tasks	Multi-scale feature mining	Captured fine-grained motor variations	Requires controlled data acquisition
[29]	Ambhika et al. (2024)	High-dimensional data	Hybrid feature selection	Reduced dimensionality with accuracy gain	Not PD-specific
[30]	Zhang et al. (2025)	Sensor time-series	Multi-task deep mining (TCN)	Joint task learning for PD recognition	Limited explainability
[31]	Huang et al. (2025)	Incomplete multi-view data	Adaptive embedding mining	Robust PD diagnosis under missing data	High training complexity

Although there has been massive advancement in the application of data mining techniques in the diagnosis and analysis of Parkinson disease (PD), a number of research gaps have been identified by the available literature. The primary weakness is a great focus on binary classification between PD patient and healthy control groups whereas such clinically important processes as disease staging, motor subtyping, and longitudinal progression modeling get relatively little attention. The majority of existing research is based on cross-sectional datasets, which makes it harder to get the dynamics of disease progression and response to treatment, and thus limits the extrapolation of their data mining models to clinical applications across the board.

One more interesting gap is in the small generalizability and strength of suggested models. Most studies have found high classification accuracy on benchmarks but these have not transferred to other cohorts because of differences in data acquisition protocols, demographics of the population and sensor configurations. This is further worsened by the lack of standardized assessment models and multi-center validation on the data. Therefore, the current implementation of the data mining-based PD systems in the actual clinical settings is limited.

Methodologically, feature selection and optimization methods have been adopted in large numbers, but little has been done with model interpretability. Many studies focus on predictive performance, instead of explainability, especially in data mining systems that are supported by deep learning. This non-transparency undermines clinical trust and reduces the embracement by the healthcare professionals. Additionally, the vast majority of feature engineering pipelines are also modality specific and the extensive scope of existing research has not explored unified multi-modal data mining approaches capable of effectively combining motor, non-motor and neuroimaging features.

Lastly, the research on the lack of data imbalance, absence of data, and uncertainty-aware mining frameworks in the analysis of PD is scarce. Although current research has only started to investigate incomplete multi-view learning and ensemble approaches, there are no in-depth solutions that place the uncertainty modeling, adaptive learning, and personalized prediction together. These gaps are critical to filling in order to take the research of data mining-based Parkinson’s disease towards scalable, interpretable, and clinically reliable decision-support systems.

Challenges and Limitations in Data Mining–Based Parkinson’s Disease Research

Although significant progress has been made with regard to data mining-based methods of diagnosing and analyzing Parkinson’s disease (PD), a number of underlying issues still restrict their effectiveness, scalability and clinical usefulness. Among the most noticeable challenges, one may identify data heterogeneity, caused by the heterogeneity of the sources of PD-related data. These involve voice and speech records, gait and wearable sensor records, handwriting dynamics, neuroimaging records and standardized clinical records. The different modalities have different statistical properties, dimensions, noise, and frequency of sampling, and therefore, it is not an easy task to have a unified data representation and integration. The available literature usually dwells on a single modality thus limiting the comprehensive perspective of manifestations of the diseases.

The other severe constraint is the lack of big and well-labeled datasets, especially to use early-stage PD and longitudinal disease surveillance. The majority of publicly accessible datasets are small, imbalanced, and gathered in a controlled, experimental setting, therefore, restricting the extrapolation of trained models to a real-world clinical environment. Inequality between classes, i.e. having a PD sample massively outnumbering healthy controls or vice versa, is another bias in terms of classification results and inflated reported performance measures. Although resampling and data augmentation are used, the approaches can cause synthetic bias and lower the clinical realism.

Another methodological issue is model overfitting and absence of cross-cohort validation. Most data mining projects test model performance on one dataset where the model is cross-validated on different dataset without external validation. This practice impedes a dependable evaluation of the model robustness to population, device and protocol variations. In addition, due to lack of standard benchmarking systems, it is not easy to have objective comparison of the studies which results in the field being fragmented.

Computationally, the growing use of complex ensemble and deep learning-assisted data mining models brings with it the issues of computational cost, energy use, and feasibility of deployment. These types of models can be very demanding on hyperparameter optimization and computation, potentially unfeasible in real-time or resource-bounded healthcare. Moreover, the vast majority of deep models act as black-boxes that can be used to provide a limited understanding of the decision-making process.

Lack of interpretability and explainability is perhaps the most significant weakness which prevents clinical adoption. Clinicians need to have transparency in their reasoning, relevance of features, and estimation of uncertainty to trust automated diagnostic systems. Nevertheless, most available data mining systems can put more emphasis on accuracy than in interpretability, which gives them minimal explanation of how certain features influence diagnostic results. This loophole inhibits regulatory consent and inclusion in clinical practices.

Future Research Directions and Open Issues

To surmount the current drawbacks, further studies will have to shift the paradigm of individual diagnostic assignments to all-encompassing, progression-sensitive data mining models. As an improvement, the advanced models need to include disease staging, motor subtyping, and progression prediction, in addition to binary PD detection. The longitudinal data mining algorithms, such as temporal pattern mining, sequential learning, and survival analysis are necessary to understand disease progression and treatment reaction over time.

Multi-modal and multi-view data mining is another potential line of study, as the analysis of heterogeneous data sources is performed together with the use of highly sophisticated fusion techniques. It should be systematically investigated whether there are complementary types of information that can be utilized by feature-level, decision-level, and representations level fusion techniques between motor, non-motor, imaging and sensor-based modalities. Scalable and modality-agnostic fusion frameworks are an open issue to develop.

Elucible and plausible information mining is a paramount open study problem. The inclusion of explainable artificial intelligence (XAI) methods, including rule extraction, feature attribution, attention visualization, and counterfactual analysis can augment model transparency to a large degree. The systems of the future must be able to make clinically interpretable output, such as scores of symptom relevance, confidence intervals, predictions that are not supportable by uncertainty, so that health care professionals can make informed decisions.

Another necessary research direction is to solve real-world data problems. PD datasets are frequently incomplete with missing data, artificially noisy signals and erratic sampling as a result of sensor failure or patient non-cooperation. Real-time monitoring and home-based assessment need robust data mining systems that can deal with incomplete, uncertain, and streaming data. The methods of transfer learning and domain adaptation are also promising to enhance cross-dataset generalization.

Moreover, the necessity of standardized evaluation procedures and multi-centre data of large scale is increasing. Reproducibility and equitable comparison of algorithms can be encouraged with help of collaborative data sharing efforts and open benchmarking platforms. In the future, the system design should also consider ethical aspects, data privacy and security especially where sensitive medical information is involved.

Conclusion

Data mining has become a paradigm shift in computational research in the field of Parkinson disease and provides potent tools of automated diagnosis, pattern discovery, and predictive modeling of complex biomedical data. The reviewed literature evidence that machine learning, feature selection, optimization and hybrid data mining frameworks are applicable in extracting discriminative biomarkers, using various modalities of PD-related data. These methods are much superior to the conventional statistical methods, in working with high-dimensional, nonlinear and heterogeneous data.

Nonetheless, in spite of these developments, the existing data mining-based studies of PD are limited by unavailable datasets, inability to generalize, inadequate explanations, and little attention to longitudinal disease modeling. Most of the available systems have controlled experimental assumptions that are not fully applicable in real life clinical conditions. To close the divide between algorithmic performance and a clinical usability paradigm, a paradigm shift to strong, interpretable, and progression-aware data mining systems is needed.

Future studies are needed to focus on the multi-modes of integration, explainability, uncertainty modeling, and validation on the large population. These obstacles can be overcome to develop data mining-based methods into high-quality clinical decision-support systems not limited to experimental prototypes. However, in the long run, these developments can considerably improve the early detection and individualized therapy, as well as the extended care of the Parkinson disease, which will become an important part of the intelligent and data-driven health systems.

Reference

1. Elshewey, A. M., Shams, M. Y., El-Rashidy, N., Elhady, A. M., Shohieb, S. M., & Tarek, Z. (2023). Bayesian optimization with support vector machine model for parkinson disease classification. *Sensors*, *23*(4), 2085.
2. Nijhawan, R., Kumar, M., Arya, S., Mendirtta, N., Kumar, S., Towfek, S. K., ... & Abdelhamid, A. A. (2023). A novel artificial-intelligence-based approach for classification of Parkinson's disease using complex and large vocal features. *Biomimetics*, *8*(4), 351.
3. Shaban, M. (2023). Deep learning for Parkinson's disease diagnosis: a short survey. *Computers*, *12*(3), 58.
4. Hussain, S. S., Degang, X., Shah, P. M., Islam, S. U., Alam, M., Khan, I. A., ... & Ismail, E. A. (2023). Classification of Parkinson's disease in patch-based MRI of substantia nigra. *Diagnostics*, *13*(17), 2827.
5. Franco, A., Russo, M., Amboni, M., Ponsiglione, A. M., Di Filippo, F., Romano, M., ... & Ricciardi, C. (2024). The Role of Deep Learning and Gait Analysis in Parkinson's Disease: A Systematic Review. *Sensors*, *24*(18), 5957.
6. Jeevika Tharini, V., Ravi Kumar, B., Sahaya Suganya Princes, P., Sreekanth, K., Kumar, B. R., & Sengan, S. (2024, January). Business Decision-Making Using Hybrid LSTM for Enhanced Operational Efficiency. In *International Conference on Multi-Strategy Learning Environment* (pp. 155-166). Singapore: Springer Nature Singapore.
7. Belyaev, M., Murugappan, M., Velichko, A., & Korzun, D. (2023). Entropy-based machine learning model for fast diagnosis and monitoring of Parkinson's disease. *Sensors*, *23*(20), 8609.
8. Alrawis, M., Al-Ahmadi, S., & Mohammad, F. (2024). Bridging Modalities: A Multimodal Machine Learning Approach for Parkinson's Disease Diagnosis Using EEG and MRI Data. *Applied Sciences*, *14*(9), 3883.
9. Cesarini, V., Saggio, G., Suppa, A., Asci, F., Pisani, A., Calculli, A., ... & Costantini, G. (2023). Voice disorder multi-class classification for the distinction of parkinson's disease and adductor spasmodic dysphonia. *Applied Sciences*, *13*(15), 8562.

10. Tharini, V. J., & Shivakumar, B. L. (2024). A Canonical Particle Swarm Optimization (C-PSO) Approach to Identify High Utility Itemset. *Journal of Computational Analysis & Applications*, *33*(5).
11. Jiao, S., Huo, H., Liu, W., Zhao, C., Ma, L., Wang, J., & Li, D. (2025). Efficient Parkinson's disease classification from dynamic foot pressure data: A combined approach of clustering and feature selection. *Biomedical Signal Processing and Control*, *105*, 107654.
12. da Silva, D. H., da Silva Souza, L. R., Ribeiro, C. T., da Silva Brasileiro, S. H., Nardo, J. R. M., Pereira, A. A., & de Oliveira Andrade, A. (2025). A Web Application for exploratory data analysis and classification of Parkinson's Disease patients using machine learning models on different datasets. *Software Impacts*, *23*, 100737.
13. Hussain, M. M., Weslin, D., Kumari, S., Umamaheswari, S., & Kamalakannan, K. (2023). Enhancing Parkinson's disease identification using ensemble classifier and data augmentation techniques in machine learning. *Clinical eHealth*, *6*, 150-158.
14. Chandru, M., Abinesh, M., Ananth, M. S., Vashist, A., & Manickam, P. (2025). Artificial Intelligence in Neurodegenerative Disease Diagnosis: Advancing Alzheimer's and Parkinson's Diseases. *Current Opinion in Biomedical Engineering*, 100638.
15. Hema, M. S., Maheshprabhu, R., Reddy, K. S., Guptha, M. N., & Pandimurugan, V. (2023). Prediction analysis for Parkinson disease using multiple feature selection & classification methods. *Multimedia Tools and Applications*, *82*(27), 42995-43012.
16. Petit, P., Berger, F., Bonneterre, V., & Vuillerme, N. (2025). Investigating Parkinson's disease risk across farming activities using data mining and large-scale administrative health data. *npj Parkinson's Disease*, *11*(1), 13.
17. Haroon, A. S., & Padma, T. (2024). An ensemble classification and binomial cumulative based PCA for diagnosis of Parkinson's disease and autism spectrum disorder. *International Journal of System Assurance Engineering and Management*, *15*(1), 216-231.
18. Raza, I., Jamal, M. H., Qureshi, R., Shahid, A. K., Vistorte, A. O. R., Samad, M. A., & Ashraf, I. (2024). Adaptive neighborhood rough set model for hybrid data processing: a case study on Parkinson's disease behavioral analysis. *Scientific Reports*, *14*(1), 7635.
19. Martinez-Eguiluz, M., Arbelaitz, O., Gurrutxaga, I., Muguerza, J., Perona, I., Murueta-Goyena, A., ... & Gabilondo, I. (2023). Diagnostic classification of Parkinson's disease based on non-motor manifestations and machine learning strategies. *Neural Computing and Applications*, *35*(8), 5603-5617.
20. Yadav, S., Singh, M. K., & Pal, S. (2023). Artificial intelligence model for parkinson disease detection using machine learning algorithms. *Biomedical Materials & Devices*, *1*(2), 899-911.
21. Srinivasan, S., Ramadass, P., Mathivanan, S. K., Panneer Selvam, K., Shivahare, B. D., & Shah, M. A. (2024). Detection of Parkinson disease using multiclass machine learning approach. *Scientific Reports*, *14*(1), 13813.
22. Hosseini, M. S., Aghamiri, S. M. R., & Panahi, M. (2025). Cross-regional radiomics: a novel framework for relationship-based feature extraction with validation in Parkinson's disease motor subtyping. *BioData Mining*, *18*(1), 67.
23. Dawood, A. S. (2023). Machine learning and artificial neural network for data mining classification and prediction of brain diseases. *International Journal of Reasoning-based Intelligent Systems*, *15*(3-4), 313-322.
24. Eliwa, E. H. I., & Abd El-Hafeez, T. (2025). Particle swarm optimization framework for Parkinson's disease prediction. *PeerJ Computer Science*, *11*, e3135.
25. Allayith, S. A., Hassan, I. A., Shakir, K. H., Ali, M. A., Shnain, A. H., Jassim, M. M., & Amir, N. A. (2024, November). Data mining applications in parkinson's disease: a comprehensive review. In *2024 International Conference on IoT, Communication and Automation Technology (ICICAT)* (pp. 706-712). IEEE.
26. Majed, R. J., Al-Heddi, R. M., & Zeki, A. M. (2023, October). Parkinson's Disease Detection Using Data Mining Models: A Comparative Study. In *2023 4th International Conference on Data Analytics for Business and Industry (ICDABI)* (pp. 290-294). IEEE.
27. Velu, K., & Jaisankar, N. (2025). Design of an Early Prediction Model for Parkinson's Disease Using Machine Learning. *IEEE Access*.

28. Peng, X., Zhao, Y., Li, Z., Wang, X., Nan, F., Zhao, Z., ... & Yang, P. (2025). Multi-Scale and Multi-Level Feature Assessment Framework for Classification of Parkinson's Disease State From Short-Term Motor Tasks. *IEEE Transactions on Biomedical Engineering*.
29. Ambhika, C., Gayathri, S., & Sheena, B. G. (2024, August). Enhancing predictive modeling in high dimensional data using hybrid feature selection. In *2024 5th International Conference on Electronics and Sustainable Communication Systems (ICESC)* (pp. 873-879). IEEE.
30. Zhang, Y., Meng, L., Chen, C., & Chen, W. (2025). SASE-TCN: A Multi-Task Learning Network for MDS-UPDRS III Task Classification and Parkinson's Disease Recognition. *IEEE Sensors Journal*.
31. Huang, Z., Wang, K., Chen, C., Chen, J., Wan, J., Yang, Z., ... & Gan, H. (2025). Incomplete Multi-view Data Learning via Adaptive Embedding and Partial l 2, 1 Norm Constraints for Parkinson's Disease Diagnosis. *IEEE Journal of Biomedical and Health Informatics*.