npj | parkinson's disease

Review article

Published in partnership with the Parkinson's Foundation

https://doi.org/10.1038/s41531-025-01025-9

Machine learning for Parkinson's disease: a comprehensive review of datasets, algorithms, and challenges

Sahar Shokrpour[1], AmirMehdi MoghadamFarid[2], Sepideh Bazzaz Abkenar[3], Mostafa Haghi Kashani[4] ✉, Mohammad Akbari[5] & Mostafa Sarvizadeh[6]

Parkinson's disease (PD) is a devastating neurological ailment affecting both mobility and cognitive function, posing considerable problems to the health of the elderly across the world. The absence of a conclusive treatment underscores the requirement to investigate cutting-edge diagnostic techniques to improve patient outcomes. Machine learning (ML) has the potential to revolutionize PD detection by applying large repositories of structured data to enhance diagnostic accuracy. 133 papers published between 2021 and April 2024 were reviewed using a systematic literature review (SLR) methodology, and subsequently classified into five categories: acoustic data, biomarkers, medical imaging, movement data, and multimodal datasets. This comprehensive analysis offers valuable insights into the applications of ML in PD diagnosis. Our SLR identifies the datasets and ML algorithms used for PD diagnosis, as well as their merits, limitations, and evaluation factors. We also discuss challenges, future directions, and outstanding issues.

Parkinson's disease (PD) is a neurodegenerative brain disorder caused by the death or impairment of specific midbrain neurons, mainly categorized as a "Movement Disorder"[1]. This condition is primarily caused by the degeneration of dopamine-producing neurons in the substantia nigra (SN) of the brain. This impairment causes a significant reduction in dopamine levels, which disrupts motor control and leads to symptoms such as tremors, muscular rigidity, bradykinesia, postural instability, and difficulty walking[2]. Dopamine is a chemical messenger that helps produce smooth and coordinated muscle and movement functions by transferring signals between the SN and another area of the brain called the corpus striatum[3]. Regarding motor control, dopamine enhances communication between the SN and the striatum, leading to the nigrostriatal pathway. This pathway regulates motor circuit activity, allowing for the smooth initiation of voluntary motions. Cognitive and behavioral problems, including dementia and depression, may develop as the disease progresses[2].

Beyond motor symptoms, PD can also manifest with various non-motor issues. Rigidity, speech disorder, rest tremors, constipation, instability in gait, rapid eye movement (REM), postural instability, and bradykinesia are examples of motor symptoms. In contrast, non-motor symptoms include neurobehavioral, mood disorders, sleep disturbances, cognitive decline, and sensory issues[4]. These symptoms are primarily due to a lack of dopamine, a neurotransmitter in the brain that regulates coordination and motion. Understanding the significance of dopamine insufficiency is critical for creating successful therapies for PD. A vital treatment is dopamine replacement therapy, which functions as a precursor to dopamine and helps refill its generalizability might represent an ongoing, unresolved levels in the brain, alleviating motor symptoms. However, a cure for the condition is still not available[5]. Scientists are conducting research to improve patient care, develop earlier diagnosis methods, and create more effective treatments for the illness[6,7].

Detecting the appropriate treatments to halt the progression of the disease can be facilitated by the early diagnosis of PD, ongoing monitoring of the condition's severity (i.e., preserving the health of the brain's neurons), and monitoring the progress of those treatments over time. However, diagnosing PD can be challenging due to its complexity and the need for clinical expertise, which may result in misdiagnosis[8]. Therefore, novel, less expensive, simpler, and more reliable approaches to Parkinson's diagnosis and treatment should be developed[9,10]. Researchers employed several machine learning (ML) techniques to improve intelligent systems that can accurately diagnose PD across various datasets. ML approaches are the process of automatically applying algorithms to datasets to retrieve valuable patterns. Typically, these approaches are used to train a computer-aided

[1]Department of Computer Engineering, ST.C., Islamic Azad University, Tehran, Iran. [2]Department of Computer Science, Michigan Technological University, Houghton, MI, USA. [3]Department of Computer Engineering, SR.C., Islamic Azad University, Tehran, Iran. [4]Department of Computer Engineering, ShQ.C., Islamic Azad University, Shahr-e Qods, Iran. [5]Department of Computer Science, Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran. [6]The Advocate Center for Clinical Research, Ayatollah Yasrebi Hospital, Kashan, Iran. ✉e-mail: mh.kashani@iau.ac.ir

diagnosis system to make decisions about classifying previously unidentified data instances. ML approaches are categorized into supervised learning (e.g., classification and regression), unsupervised learning (e.g., clustering), and reinforcement learning (e.g., Q-learning)[11]. In the context of PD diagnosis, ML approaches may assess patient data such as motor symptoms, imaging scans, and genetic data to recognize patterns over time and estimate disease progression[7,12].

Furthermore, some researchers have used a group of ML methods named deep learning (DL) to address the challenges of ML by automating feature extraction. DL uses multilayer ANN. DL excels at extracting beneficial features from diverse data such as neuroimaging data, motor symptoms, and time series[13,14]. This feature is beneficial during PD diagnosis, where slight alterations in motor functions or brain activity might be challenging to recognize manually. Popular DL models, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), demonstrated potential in processing multiple types of patient data for PD diagnosis[15–17]. By comparing the efficacy of several ML and DL algorithms in diagnosing PD, we intend to emphasize their potential to enhance early diagnosis, track disease progression, and assess the treatment process.

Conducting a thoroughly systematic literature review (SLR)[18,19] in diagnosing PD based on ML with an emphasis on improvements, research questions, future works, and PD detection directions is rare. This paper not only reviews the application of well-known ML algorithms in detecting Parkinson's patients but also studies popular datasets and approaches and compares their limitations and specifications. Increasing usage of artificial intelligence (AI) in medicine and disease detection has shaped the scheme of our research on diagnosing PD. Finally, our research may assist in developing more precise and reliable diagnostic methods for PD, improving patients' quality of life as well as health conditions. We have addressed the following research questions to synthesize the relevant information and efficiently convey knowledge to the research associations:

- What types of data sets are used to diagnose PD?
- Which category of data sets is used the most to diagnose PD?
- What tools are used the most in assessing ML approaches in diagnosing PD?
- What evaluation metrics are commonly used to assess the ML techniques in diagnosing PD?
- What ML algorithms have been considered the most in diagnosing PD?
- What validation methods are used in studies diagnosing PD with ML?
- What are the major challenges, future trends, and open issues in diagnosing PD with ML?

Our investigation highlights the growing interest and rapid expansion of studies incorporating ML in diagnosing PD. The utilization of ML technology within healthcare systems, particularly in the context of neurodegenerative diseases such as PD, is becoming increasingly necessary as it continues to improve. In particular, this literature review offers valuable information for:

- Healthcare and ML researchers: We offer a thorough review of the current state of ML approaches and methods for PD diagnosis, providing insights for those interested in further exploration in this field.
- Neurologists and medical professionals: Understanding ML-based PD diagnosis methods, strategies, and tools could prove helpful for the diagnosis of neurodegenerative diseases.

The structure of the remaining sections of this SLR is outlined in Fig. 1 and the key abbreviations are defined in Table 1. The Related Work section discusses some related works and motivations. The research method, including the selection procedure and research questions, is covered in the Research Methodology section. The Section on Classification provides a comprehensive analysis and classification of the selected papers, emphasizing their primary advantages, disadvantages, and evaluation factors. An examination of the findings, potential trends, and outstanding issues is given in the Analysis of Results and Open Issues, Challenges, and Future Trends, respectively. Threats and limitations of validity are covered in the Threats to Validity and Limitations section. The conclusion is finally given in the final section.

Related work

This part of the paper investigates important reviews of the existing datasets and ML approaches for diagnosing PD. The mentioned studies are evaluated in terms of main idea, taxonomy, paper selection procedure, database investigations, open issues, and covered years, and their summary is also provided. The outcomes are listed in Table 2.

Khachnaoui et al.[20] assessed ML and DL-based computer-aided diagnosis methods for PD, along with introducing single photon emission computed tomography (SPECT) and positron emission tomography (PET) for detection. They deliberated the pros and cons of hand-crafted ML techniques, concluding that DL approaches offered the most robust and dependable solution for feature extraction in the diagnostic domain. Additionally, Salari et al.[21] proposed an SLR to evaluate the effectiveness of ML methods in detecting PD cases until the end of 2020. This review involved seven distinct phases, wherein they categorized various approaches and databases, providing statistical insights. Ultimately, their findings suggested that ML approaches offered practical utility in diagnosing PD.

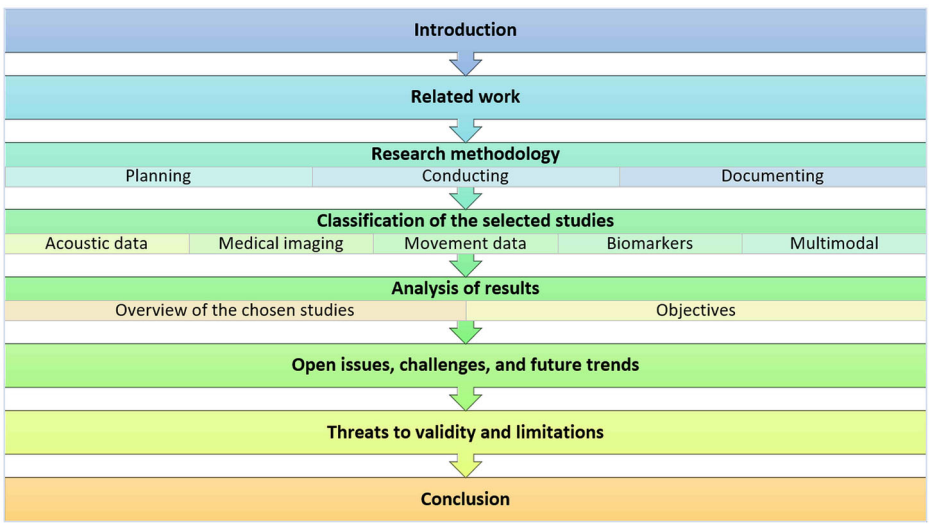


Fig. 1 | Structural map of this SLR.

Table 1 | List of abbreviations

Abbreviation	Definition	Abbreviation	Definition
AI	Artificial Intelligence	LASSO	Least absolute shrinkage and selection operator
AD	Alzheimer's disease	ML	Machine learning
ANN	Artificial neural networks	mRMR	Minimum redundancy maximum relevance
ART	Artifact detection tools	MLP	Multi-layer perceptron
ANT	Advanced normalization tools	MRI	Magnetic resonance imaging
ARR	Analysis of variance with recursive reduction	MCOA	Modified crayfish optimization algorithm
AdaBoost	Adaptive Boosting	MDS	Multidimensional scaling
BGRU	Bidirectional gated recurrent unit	MFDFA	Multifractal detrended fluctuation analysis
BLSTM	Bidirectional long short-term memory	MS-ResNet	Multi-scale residual neural network
BOSS	Bag of symbolic Fourier approximation symbols	MLA	Modified local accuracy
BRF	Bagged random forests	MLP_BPC	Multilayer perceptron back propagation classifier
BNB	Bernoulli naive Bayes	MetDNA	Metabolite identification and dysregulated network analysis software
ChOA	Chimp optimization algorithm	NB	Naive Bayes
CRNN	Convolutional recurrent neural networks	N3	Non-parametric non-uniform intensity normalization algorithm
CatBoost	Categorical Boosting	NPNN	New probabilistic neural network classifier
CART	Classification and regression tree	PCA	Principal component analysis
CUDA	Compute unified device architecture	PNN	Probabilistic neural networks
DL	Deep learning	PET	Positron emission tomography
DNN	Deep neural network	PD	Parkinson's disease
DSL	Deep sample learning	PAC	Passive aggressive classifier
DCNN	Deep convolutional neural network	RIPPER	Repeated incremental pruning to produce error reduction
DT	Decision tree	PLS-DA	Partial least-squares discriminant analysis
DCT	Discrete cosine transforms	QDA	Quadratic discriminant analysis
DAT	Dopamine transporter	QSM	Quantitative susceptibility mapping
DWT	Discrete wavelet transforms	QC-RLSC	Quality control robust loess signal correction
DRSN	Deep residual shrinkage network	QReLU	Quantum rectified linear unit
ELM	Extreme learning machine	REM	Rapid eye movement
EM	Expectation-maximization	RNN	Recurrent neural network
ELA	Ensemble learning based AdaBoost	ReLU	Rectified linear unit
ECG	Electrocardiogram	RF	Random forest
ET	Extra trees classifier	RQ	Research question
ERT	Extremely randomized trees	ResNeXt	Residual neural network
ECOCMC	Error correcting output codes model classifier	RMSPROP	Root mean square propagation
FOG	Freezing of gait	RMA	Robust multi-array average
FT	First step transfer	RDFSA	Regularized discriminative feature selection algorithm
FFT	Fast Fourier transform	SLR	Systematic literature review
FSL	FMRIB software library	SPECT	Single-photon emission computed tomography
FNN	Feedforward neural networks	SVM	Support vector machine
FSKL-LLC	Feature selection and kernel learning for local learning-based clustering	SPM	Statistical parametric mapping
GMM	Gaussian mixture models	SGD	Stochastic gradient descent
GB	Gradient boosting	SMOTE-ENN	Synthetic minority oversampling technique – edited nearest neighbors
GDABC	Gbest dimension artificial bee colony	SHAP	Shapley additive explanation
GAN	Generative adversarial network	SN	Substantia nigra
GA	Genetic Algorithm	SNPC	Substantia nigra pars compacta
Grad-CAM	Gradient-weighted class activation mapping	SVM-RFE	SVM recursive feature elimination
GNB	Gaussian naive Bayes	SDTW	Sequential dynamic time warping
GBDT	Gradient boosting decision trees	SMOTE	Synthetic minority oversampling technique
HC	Healthy control	STMIM	Spatiotemporal microstate identification model
HGSA	Hybrid grid search algorithm	SVR	Support vector regression

Table 1 (continued) | List of abbreviations

Abbreviation	Definition	Abbreviation	Definition
HRQoL	Health-related quality of life	SF-PC	Sort features based on pairwise correlations
IoT	Internet of things	SSAE	Stacked sparse autoencoder
ICA	Independent Component Analysis	TL	Transfer learning
IMC	Iterative means clustering	TCS	Transcranial Sonography
IO-HMM	Input-Output Hidden Markov Model	TQWT	Tunable Q-factor wavelet transform
IPA	Ingenuity pathway analysis	UFS-MCC	Unsupervised feature selection for multi-class cluster
JCR	Journal citation reports	UMAP	Uniform manifold approximation and projection
KNN	k-nearest neighbors	UFS-OL	Unsupervised feature selection algorithm with ordinal locality
LightGBM	Light gradient boosting machine	VGG	Visual geometry group
LR	Logistic regression	WFM-DSS	Weighted fusion mechanism based on deep sample space
LDA	Linear discriminant analysis	WPT	Wavelet packet transform
L1R&FS	L1 regularization feature selection	XGBoost	eXtreme gradient boosting

In another study, Tanveer et al.[22] offered a review of papers from 2013 to 2021 that focused on using deep neural network (DNN) and ANN to detect PD. The authors analyzed various data modalities and model performances, finding that convolutional recurrent neural networks (CRNN) models excelled in time-series accuracy. They also concluded that the performance of ANNs improved when clinical features were incorporated. Lastly, they outlined the pros and cons of these models and suggested topics for future research. Further, Sigcha et al.[23] assessed 69 papers on using ML and DL to analyze motor and non-motor data from wearables. The selected papers were from 2012 to 2022, focusing on monitoring and diagnosing PD. Also, they provided some future trends and challenges.

Additionally, Skaramagkas et al.[24] presented a systematic review leveraging DL approaches to discriminate PD symptoms effectively with motor symptoms. They investigated 87 papers from 2016 to 2023 over DL, considering speech, upper limb, facial expression, and gait data. Moreover, Amato et al.[25] reviewed acoustic features and ML methods for detecting PD. They analyzed 102 papers published from 2017 to 2022. Their focus was on statistically evaluating the applied techniques and algorithms. Khanna et al.[26] investigated the application of neuroimaging and ML methodologies in the diagnosis of different disorders, such as PD, Alzheimer's disease (AD), and Schizophrenia. The distinctive aspect of this study resided in its incorporation of the most recently published scholarly publications.

Keserwani et al.[27] examined various ML, meta-heuristic, and DL models for diagnosing PD. Additionally, they improved the accuracy of existing models by utilizing speech signal datasets. Furthermore, potential future directions were explored. Also, Islam et al.[28] analyzed ML and DL methods for diagnosing PD, specifically looking at handwriting and wave databases. They explored different algorithms and the nuances of biomarkers to improve diagnosis. Findings suggested these techniques in diagnosing patients, though the study also outlined certain constraints and potential directions for future research. In another review, Sabherwal and Kaur[29] assessed the effectiveness of ML and DL algorithms regarding PD detection. Various methods were analyzed. Also, their limitations and future trends were discussed. Papers were selected from 2013 to 2023 over the mentioned area. Giannakopoulou et al.[30] also reviewed ML algorithms trained on data collected from wearable sensors and IoT devices for predicting PD. The authors surveyed 112 papers in an SLR manner, highlighting the best methods and tools. Moreover, some open issues and future challenges were discussed.

Furthermore, Rana et al.[31] offered a review to identify the common and accessible AI algorithms and ML techniques for detecting PD patients. They reviewed 112 papers and examined them in terms of data and methods. The study results indicated that ML and biomarker data had the highest efficiency in PD detection. They also discussed some future trends and challenges. Zhang[32] categorized ML-based techniques of PD diagnosis into 3 groups by assessing 51 papers from 2006 to 2019, namely discrimination between PD and healthy control (HC), differential diagnosis, and early PD detection. Also, results demonstrate that the use case of ML improves the accuracy of PD identification. Moreover, some challenges and future solutions were introduced. Also, Chandrabhatla et al.[33] investigated the common ML models and computational techniques used for detecting PD from 1970 to 2020. This review was carried out by using the US National Library of Medicine PubMed database, showing the significant advances in the mentioned area.

Table 2 shows that while the covered time by previous papers is broad, they evaluated a limited number of studies. The works of Salari et al.[21], Sigcha et al.[23], Giannakopoulou et al.[30], and Rana et al.[31] cover fewer studies over more extended period. Moreover, only three papers (Rana et al.[31], Keserwani et al.[27], and Islam et al.[28]) offer classifications. Also, it should be noted that several studies did not examine datasets at all, including those by Rana et al.[31], Sabherwal and Kaur[29], Khachnaoui et al.[20], Khanna et al.[26], Giannakopoulou et al.[30], Zhang[32], and Chandrabhatla et al.[33]. However, out of these studies, the work of Skaramagkas et al.[24] is similar to our investigations. In this SLR, 80 publications between 2016 and 2023 have been reviewed, but its authors did not present any taxonomy or study applied tools. Islam et al.[28] conducted another close effort; however, it introduced a taxonomy that focused solely on handwriting and voice data, examining fewer experiments than ours from 2000 to March 2023 and did not present the employed tools. The closest in breadth, Rana et al.[31], assessed a small number of studies, did not evaluate datasets, and did not comprehensively address open issues. In addition, it was not written in a systematic way.

Furthermore, the reviews by Amato et al.[25], Chandrabhatla et al.[33], and Khachnaoui et al.[20] did not discuss open issues, did not present any taxonomy or classification, and were not written systematically. Given these gaps, a comprehensive SLR is needed that not only addresses these limitations but also clearly identifies open issues and future research directions. On the other hand, the statistical correctness of our work is enhanced by its careful evaluation of 133 publications published between 2021 and April 2024. The majority of prior evaluations were not structural and were missing some crucial pieces of information. They often failed to consider tools, present any taxonomy or classification, or discuss the merits, downsides, and unanswered concerns. As of right now, a comprehensive evaluation of the various datasets utilized for PD diagnosis with ML has not been published.

Research methodology

Researchers have introduced various techniques for diagnosing PD. This section describes the systematic phases of examining various ML approaches for diagnosing PD. Figure 2 illustrates that, in contrast to the non-structured review process, the SLR process reduces bias while identifying research directions and open issues by following the correct phases in a methodical sequence for investigating the literature[19,34,35]. SLR methods are

Table 2 | Summary of related works

Review type	Article	Main topic	Publication year	Paper selection process	Taxonomy	Open issue	Covered year	Dataset investigation	Tools
Survey	20	DL techniques in PD feature extraction	2020	No	No	No	2013–October 2020	No	No
	22	Review of DNN/ANN in PD detection	2022	Yes	No	Yes	2013–-2021	Yes	No
	25	Voice analysis survey for PD diagnosis	2023	Yes	No	No	2017–June 2022	Yes	Yes
	26	ML approaches overview in PD diagnosis	2023	Yes	No	Yes	2017–February 2023	No	No
	27	DL, ML, and meta-heuristics approaches in PD detection	2024	No	Yes	Yes	2012–January 2023	Yes	No
	29	ML/DL methods for PD identification	2024	Yes	No	Yes	2013–April 2023	No	No
	31	ML's role in PD detection	2022	Yes	Yes	Yes	1996–2022	No	Yes
	32	ML approaches with imaging and clinical data for PD	2022	No	No	Yes	2006–2019	No	No
	33	ML and digital tech evolution in PD detection	2022	No	No	No	1970–2020	No	No
SLR	21	ML approaches performance in PD diagnosis	2022	Yes	No	No	2011–November 2020	Yes	No
	23	Wearable devices for PD monitoring	2023	Yes	No	Yes	2012–April 2022	Yes	No
	24	ML for PD detection via motor symptoms	2023	Yes	No	Yes	2016–January 2023	Yes	No
	28	PD diagnosis via voice and handwriting	2024	Yes	Yes	Yes	2000–March 2023	Yes	No
	30	IoT monitors and ML for PD diagnosis	2022	Yes	No	Yes	2012–August 2021	No	No
	This paper	Applied ML datasets for PD diagnosis	–	Yes	Yes	Yes	2021–April 2024	Yes	Yes

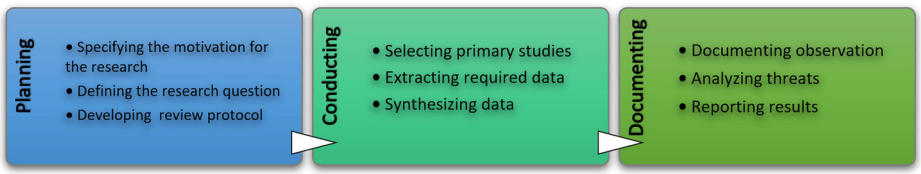


Fig. 2 | Overview of the research methodology.

based on well-defined instructions to identify, provide the desired results, analyze, and answer the defined questions. Therefore, this paper utilizes SLR guidelines to design a three-stage review procedure: planning, conducting, and documenting[18].

Planning

The planning phase originates by defining the reason for the research and culminates in establishing the methodology for the review protocol in the following manner:

Stage 1: research motivation. An SLR is needed to classify and compare published studies on diagnosing PD by utilizing a characterization framework. While the causes of PD are uncertain, medication and surgery can relieve its symptoms. However, early diagnosis of the syndrome and its progression rates may be crucial in determining suitable treatments. Therefore, scientists have developed various techniques to address this issue and improve PD's diagnosis accuracy. This has raised the motivation to conduct a systematic review on the application of different ML approaches for diagnosing PD, categorizing them taxonomically, and presenting a concise comparative analysis of the applied datasets and techniques, as well as their potential limitations and challenges.

According to our findings, few papers give an in-depth review of this subject. Since the evaluations of previous studies had limitations, provided no analysis on applied tools, and evaluated insufficient datasets, presenting an SLR is essential. By contrast, our SLR carefully assesses 133 current papers to offer a systematic categorization of datasets utilized in ML for PD diagnosis, addressing significant gaps in prior research. This detailed assessment helps to improve diagnostic methods, achieve our following goals, and guide future advancements in the field.

- To include papers that have recently been published
- To outline any potential future works
- To investigate evaluation factors and applied tools
- To present a comprehensive taxonomy, concise statistical information, and simulation tools and precisely define the paper selection procedure

Stage 2: research questions. We have developed research questions based on our motivation to conduct an impartial and scientific examination of PD datasets and the ML algorithms utilized in detecting the disease at its early stages. Answering the defined questions identifies available issues on this topic, which could also assist in brainstorming fresh ideas throughout this documentation phase. Table 3 highlights these research questions (RQs):

Stage 3: review protocols. We have presented and implemented a review protocol, which includes a set of questions, the paper selection method, and data extraction in Stages 1, 2, and 3 of the conducting phases. Following the guidelines in[18] and[36], we sought feedback from a third-party expert with experience in conducting SLRs on ML and

healthcare systems to evaluate the protocol before its implementation. We incorporated his recommendations to modify the protocol. In addition, we conducted a pilot study (including about 20% of the papers) to eliminate researcher bias and boost the extraction of data. We have improved the inclusion/exclusion criteria, search strategies, and the scope of the review during this stage. In addition, a protocol was designed based on[37] and our prior experience in conducting systematic reviews[38–51].

Conducting

The conducting phase of the research process begins with choosing the papers and ends with the extraction of data. This section attempts to clarify the process of seeking and selecting articles throughout the second phase of the SLR. This section consists of two parts. First, study selection explains the process of selecting papers, and second, data extraction and synthesis, expressing how we have accomplished the review.

Stage 1: study selection process. This section describes the process of identifying and choosing articles during the second phase of SLR. As demonstrated in Fig. 3, we choose papers in this timeframe employing a two-step procedure. One of the most crucial requirements of developing a research plan is finding all publications without bias. Standard strings were found and added to the search phrase to achieve this objective. A review methodology was also designed to help identify relevant and unbiased research. During the first selection procedure, all search terms associated with Parkinson and ML algorithms were chosen to ensure that no papers were missed. Throughout this stage, the following search terms were employed in the abstracts, titles, and keywords of six academic online databases:

- *Initial selection:* we conducted a thorough search throughout many reputable databases, namely Elsevier, IEEE, Taylor & Francis, Springer, ACM, Nature, and Wiley, using the search phrase supplied below, which included keywords, titles, and abstracts. Our search, specifically utilizing the journal citation reports (JCR) as a source, spans 2021 through April 2024. This timeframe was selected due to the significant variety and volume of publications in PD detection employing ML methods. 729 papers related to the collaborative field of ML and PD detection were found as a result of this search. These papers were published in various sources, including journals, conferences, and book chapters.

Parkinson [AND]
(Supervised <OR > SVM < OR> "support vector machine" <OR> "linear discriminant analysis" <OR> "naive Bayes" <OR> "Machine Learning" <OR > LDA < OR> "Deep Learning" <OR > KNN < OR> "k-nearest neighbors" <OR> "Neural Network" <OR> "Neural Networks" <OR> "Decision Trees" <OR> "Random Forests" <OR> Gaussian <OR> "Latent Variable" <OR> Unsupervised <OR> "Data Clustering" <OR> "Dimension Reduction" <OR> Ensemble)

- *Final selection:* the papers extracted in the previous step were examined, and the inclusion/exclusion criteria (Table 4) were applied. Survey papers, theses, non-English papers, books, non-peer-reviewed papers, conference papers, short papers, and book chapters were excluded. The selected papers were thoroughly read, and quality assessment was used to include only those papers that presented evaluating details and approaches. At the end of this stage, 133 relevant studies were chosen for qualitative evaluation.

Stages 2 and 3: data extraction and synthesis. We acquired data from a list of online academic databases. We provided it in a systematic format based on characterization aspects employing the guidelines above. A structured analysis was created by exploring the restrictions and potentials of the studied papers, as well as the perspectives on future studies, providing us with an exploration of the collective study findings. We extracted data from the examined papers, concentrating on dataset types, applied tools/algorithms, advantages, disadvantages, paper main ideas, and evaluation metrics in Section "Classification of the selected studies". Then, based on the data extracted in Section "Classification of the selected studies", the results of this study were analyzed, and the findings were discussed and addressed alongside the research questions in Section "Analysis of results". Furthermore, the review of the papers in Section "Classification of the selected studies" and the data extracted from the studied works in Section "Analysis of results", highlight key challenges, open issues, and future trends in applying ML for PD diagnosis, which are comprehensively detailed in Section "Open issues, challenges, and future trends".

Table 3 | Research questions

RQ1: What types of data sets are used to diagnose PD?
RQ2: Which category of data sets is used the most to diagnose PD?
RQ3: What tools are used the most in assessing ML approaches in diagnosing PD?
RQ4: What metrics are significantly used to assess the ML techniques in diagnosing PD?
RQ5: What ML algorithms have been considered the most in diagnosing PD?
RQ6: What validation methods are used in studies diagnosing PD with ML?
RQ7: What are the major challenges, future trends, and open issues in diagnosing PD with ML?

Table 4 | Inclusion/exclusion criteria

Criteria	
Inclusion	• The papers focusing on PD diagnosis based on ML approaches • The papers published from 2021 to April 2024
Exclusion	• Surveys, books, review papers, and book chapters • Short papers with fewer than six pages • Non-English papers • Papers not directly related to PD and ML approaches • Conference papers

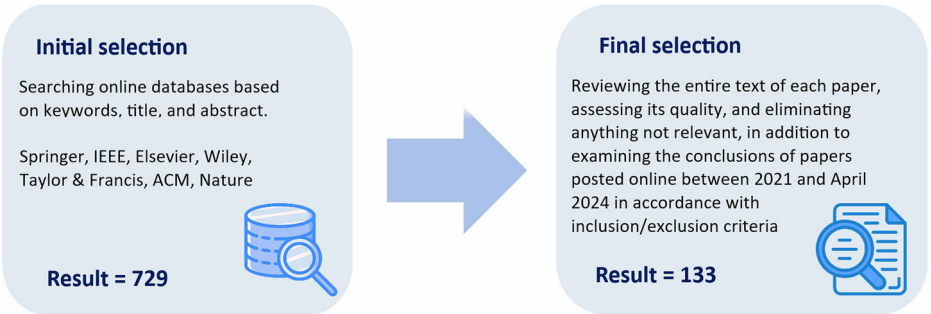
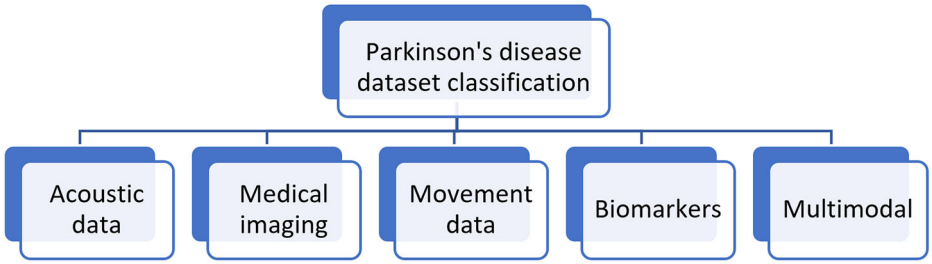


Fig. 3 | Paper selection process.

Fig. 4 | The classification of PD analysis.



Documenting

In the documenting phase (see Fig. 2), observations are recorded, and the discussion of threats to validity and limitations is provided in Section "Threats to validity and limitations". The results are evaluated, visualized, and presented in Section "Analysis of results".

Classification of the selected studies

This paper presents a classification designed to achieve two major objectives. First, it facilitates an SLR of the extensive body of research on PD diagnosis using ML approaches. Second, organizing the studies into five data-driven categories—illustrated in Fig. 4—enhances readers' understanding of the main research areas in the field. The proposed classification highlights the important role of datasets in diagnostic results. For instance, the adoption of multimodal datasets demonstrates growing interest in innovative approaches that leverage diverse data sources to improve the accuracy of PD diagnosis. Although alternative classification exists, our classification, to the best of our knowledge, has not been adopted in earlier reviews. By concentrating on datasets, this work offers a unique perspective that emphasizes both the pivotal role of data in ML performance and the existing gaps in data diversity and availability.

This section also investigates 133 papers to determine the PD dataset, diagnosis approaches, objectives, and applied approaches. In addition, we provide the advantages and disadvantages of these approaches by using assessment metrics defined in Table 5. Moreover, Fig. 4 illustrates the proposed classification for PD; while presenting a classification for PD is not a trivial task, it allows readers to easily access dataset resources and the related papers, providing a holistic understanding of the field and helping to identify research gaps. However, each researcher may adopt a different approach to categorization due to their unique perspective.

Furthermore, the selected paper applies ML approaches to speed up the diagnosis of PD using proper datasets. These datasets are classified into five groups: acoustic data/features, medical imaging, movement data, biomarkers, and multimodal datasets.

Figure 4 presents the categorization used in this review, which is based on the datasets utilized for diagnosing PD. Five major categories were identified, each reflecting the unique characteristics and data extraction methods of their respective datasets.

Acoustic data/features

Acoustic data is crucial for diagnosing PD, as audio symptoms can be detected early, often before noticeable motor symptoms appear[52]. Therefore, diagnosing PD through this information can significantly enhance the quality of life for the affected individuals[2]. Symptoms of PD can be classified into two categories: dysphonia and dysarthria[53]. In the early stages of diagnosis, the identification of dysphonic indicators can delay the severity of the disease. Furthermore, recording voices is a relatively low-cost approach, and physical examination can be time-consuming[54]. Therefore, having sufficient data for training ML models could significantly expedite PD prediction and improve patient well-being. In Table 6, we present data from databases related to acoustic-based papers, including the number of patients and healthy subjects, the gender of patients, and details such as extracted features, audio file format, and the number of samples. Table 7 compares the reviewed studies, highlighting their advantages, disadvantages, main ideas, applied tools, and algorithms. The evaluation parameters employed in acoustic-based papers are analyzed in Table 8. In Section "Review of acoustic-based approaches", acoustic-based approaches are reviewed. Ultimately, Section "Qualitative analysis of acoustic-based approaches" presents a qualitative analysis of acoustic-based approaches, examining their strengths, weaknesses, opportunities, and threats.

Review of acoustic-based approaches. Yao et al.[55] proposed a model based on deep convolutional neural networks (DCNN) along with an IP-based whale optimization algorithm (WOA). This study aimed to diagnose pathological speech in PD and cleft lip and palate patients. According to the results, the proposed model provided high precision and accuracy. Also, Khaskhoussy and Ayed[56] proposed a method to detect PD by analyzing speech data, and they also used SVM and convolutional neural network (CNN) to classify speech data. The results showed that this method has high accuracy and specificity. Meanwhile, Celik et al.[57] developed the SkipConNet + RF diagnostic model for PD detection via speech signals. The model integrated CNN and RF algorithms. SkipConNet extracted vital speech signal features, and the RF predicted these features. According to the study, the presented model surpassed other DL and ML methods in accuracy.

Following a similar investigative field, Ali et al.[58] introduced an ensemble method named ensemble model with optimal features and sample-dependent base classifiers (EOFSC) for identifying individuals with PD. The technique addressed issues of generalization and low accuracy by analyzing voices. Also, Masud et al.[59] developed the crow search and deep learning (CROWD) model for diagnosing and classifying PD. The model employed the adaptive crow search algorithm (ACSA) to select the compressed feature vector and a DL-based autoencoder to generate the compressed feature vector. The outcomes also indicated that the proposed model achieved high sensitivity, specificity, and accuracy. Yücelbaş[60] introduced a model to diagnose PD in the early stages using acoustic signals. Based on the results, the information gain algorithm-based k-nearest neighbors (IGKNN) model had high accuracy and recall.

Regarding the detection of PD with ML approaches in a quicker manner, Wang et al.[61] introduced a system for diagnosing PD using speech data and DL methods. Also, results showed that the mentioned system was highly accurate. Moreover, Dhar[54] designed a two-phase reduction system with the objective of diagnosing PD in its early stages. The first reduction steps comprised ML-based feature selection, whereas the other phase employed unsupervised techniques. High performance, as indicated by accuracy and AUC, was demonstrated through the comparisons and assessment findings. Li et al.[62] introduced a method called two-step transfer learning (TSTL) to identify Parkinson's patients from healthy individuals based on speech patterns. Also, their method was evaluated with a new and publicly available dataset. In addition, the results showed that the presented method had high accuracy, sensitivity, and specificity.

Moreover, Vital et al.[63] presented a model based on neural networks to detect individuals with PD by analyzing their voices. They compared the method with various ML algorithms, including naive Bayes (NB), random forest (RF(, and AdaBoost. Based on the results, the accuracy and efficiency of the model were confirmed. In a related study, Jyotiyana et al.[64] presented an approach for PD diagnosis using DL techniques. In this study, audio data

Table 5 | Definition of evaluation metrics

Metrics name	Description
Condition positive (P)	The count of positive samples in data.
Condition negative (N)	The count of negative samples in data.
True positive (TP)	The count of correctly positive predictions.
True negative (TN)	The count of correctly negative predictions.
False positive (FP)	The count of wrongly positive predictions.
False negative (FN)	The count of wrongly negative predictions.
Accuracy	It is the proportion of correctly labeled samples to the total observations[55]: $ACC = \frac{TP+TN}{P+N} = \frac{TP+TN}{TP+TN+FP+FN}$
Sensitivity, recall, or true positive rate (TPR)	It is a metric that measures the ratio of correctly positive predictions out of the total number of positive samples[56]: $TPR = \frac{TP}{P} = \frac{TP}{TP+FN}$
Specificity, selectivity, or true negative rate (TNR)	It represents the ratio of true negatives to the total number of actual negatives in the data[145]: $TNR = \frac{TN}{N} = \frac{TN}{TN+FP}$
Kappa coefficient	It is used to calculate the level of agreement between two raters[114]: $\kappa = \frac{p_0 - p_e}{1-p_e}$
Precision or positive predictive value (PPV)	The proportion of correctly classified positive samples (TP) to all classified positive samples is known as precision[146]: $PPV = \frac{TP}{TP+FP}$
F1-score or F-measure	F1-score offers a method for combining recall and precision into a single measure that accounts for both parameters. It represents the harmonic mean (average) of recall and precision[147]: $F1 - score = 2 \times \frac{PPV \times TPR}{PPV+TPR} = \frac{2TP}{2TP+FP+FN}$
False positive rate (FPR)	This metric represents the ratio of negative samples that are incorrectly classified as positive by a model[57]: $FPR = \frac{FP}{FP+TN}$
Receiver Operating Characteristic (ROC) or Area Under the Curve (AUC)	It is plotted with TPR against the FPR, with TPR on the y-axis and FPR on the x-axis. A higher AUC indicates higher performance[146].
Matthew's correlation coefficient (MCC)	It calculates the correlation of predicted and actual binary outcomes, considering all four factors of a confusion matrix. In other words, it measures the performance of binary classification[58]: $MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}$
Root mean square error (RMSE)	It is a statistical metric that's used to assess model performance and gauge how far the actual value distance is from the predicted value[59]: $RMSE = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \hat{x}_i)^2}{N}}$
Precision-recall curves (PRC)	PR curves are binary classification evaluation tools used in ML for imbalanced data sets. They visualize performance at different thresholds, allowing comparison of models and selection of optimal thresholds[60].
Negative predictive value (NPV)	NPV shows the proportion of correctly negative results over all the negative predictions. It expresses the likelihood that a predicted negative is a true negative[85]: $NPV = \frac{TN}{TN+FN}$
Standard deviation	Standard deviation is a statistical value for representing the variation or dispersion of a dataset from its mean value. Also, it is calculated as the square root of the variance[86]: $\sigma = \sqrt{\frac{\sum (x_i - \mu)^2}{N}}$
Intraclass correlation coefficient (ICC)	The ICC statistic measures the likelihood of outcomes being comparable within each cluster or different between clusters compared to other clusters[87].
Mean square error (MSE)	It is the mean of the square of the difference between observed values and estimated data[88]: $MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y})^2$
Pearson's correlation coefficients	It represents a statistical parameter that evaluates the strength and direction of the linear correlation between two continuous variables[87]: $r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}}$
Euclidean distance	Euclidean distance measures the straight-line distance between two points in space. It's commonly used in cartesian coordinate systems[89]: $d(p, q) = \sqrt{\sum_{i=1}^{n}(q_i - p_i)^2}$
Dice similarity coefficient (DSC)	DSC is a statistical measure for evaluating the similarity between two sets. It is commonly applied in image processing and bioinformatics, as well as other fields that require binary dataset comparison[89].
Mean absolute error (MAE)	It measures the average magnitude of errors between actual values and predicted values. Additionally, it is used to evaluate the accuracy of regression models[115]: $MAE = \frac{1}{n}\sum_{i=1}^{n}\lvert y_i - \hat{y}_i \rvert$
Correlation coefficient	Correlation coefficient represents the direction and strength of the relationship between two variables through a number between -1 and 1[170]: $r = \frac{\sum(x - \bar{x})(y - \bar{y})}{\sqrt{\sum(x - \bar{x})^2 (y - \bar{y})^2}}$

Table 5 (continued) | Definition of evaluation metrics

Metrics name	Description
Signal-to-noise ratio (SNR)	In signal processing, SNR is a metric used to determine the relative magnitude of a desired signal compared to the nearby noise. A higher SNR number conveys a better signal quality[116].
Mean integrated squared error (MISE)	When estimating nonparametric regression functions, the MISE is a useful measurement to evaluate an estimator's overall accuracy. It displays the integrated squared error's predicted value throughout the data set[8]: $MISE(\hat{f}) = E[\int(\hat{f}(x) - f(x))^2 dx]$
G-mean	In evaluating classification models, particularly in imbalanced datasets, the geometric mean is a criterion. It also aims to balance the performance of the model between different classes[117]: $G - mean = \sqrt{TPR \times TNR}$
Equal error rate (EER)	It is the error observed at the ROC curve point where FPR equals FNR[118].
F1-micro	F1-micro is a measure of the harmonic mean of recall and precision, calculated by summing all TP, FP, and FN across all classes[119]: $F1 - micro = \frac{TP}{TP + \frac{1}{2}(FP+FN)}$
PR (AUC)	For imbalanced datasets, in particular, the precision-recall area under the curve—a binary classification—is utilized to assess the efficacy of the model[119].
Relative absolute error (RAE)	RAE is a ratio used to assess the performance of a forecasting model and compare the average error (residual) to the errors generated by a naive model[115]: $RAE = \frac{\sum \lvert y_i - \hat{y}_i \rvert}{\sum \lvert y_i - \bar{y} \rvert}$
Model likelihood	This value provides a quantitative assessment of a model's data-fitting performance. It is commonly applied when comparing the efficacy of multiple models.[171].
Balanced accuracy	Balanced accuracy, defined as the average recall in each group, is used in group classification problems to deal with imbalanced data sets[174]. $Balanced\ accuracy = \frac{sensitivity + specificity}{2}$

Table 6 | Acoustic datasets

Article	Repository/Source	#PD	#HC	#PDF	#PDM	#HCF	#HCM	Information
55	CIEMPIESS[189]	–	–	–	–	–	–	Mexican Spanish podcast, 6717 samples (45 F, 96 M)
	190	–	–	–	–	–	–	135 children (ages 5–15)
	PC-GITA corpus[191]	50	50	–	–	–	–	100 Spanish speakers
56	Sarkar dataset[192]	20	20	6	14	10	10	1208 audio recordings (.wav), Test: 168 samples from 28 participants (14 HC and 14 PD)
57	Kay Elemetrics Disordered Voice Database (Max Little)[193]	23	8	–	–	–	–	195 speech samples, 23 acoustic features from audio signals
	PD Speech (PDS)[194]	188	64	81	107	41	23	Age range: 33–87 years
58	Sarkar dataset[192]	20	20	6	14	10	10	68 participants: 40 for training and 28 for testing, 26 extracted acoustic features
	UCI Parkinson's Telemonitoring Voice Dataset[195].	60	100	19	41	21	79	480 voice samples (180 PD, 300 HC), Age range: 43–88 years (age-matched)
59	PD Speech (PDS)[194]	188	64	81	107	41	23	Age range: 33–87 years
60	PD Speech (PDS)[194]	188	64	81	107	41	23	Age range: 33–87 years
61	MDVR-KCL[196]	16	21	–	–	–	–	Data collected in 2017 at KCL Hospital via smartphone
54	PD Speech (PDS)[194]	188	64	81	107	41	23	Age range: 33–87 years
	Kay Elemetrics Disordered Voice Database (Max Little)[193]	23	8	–	–	–	–	197 samples from 31 subjects 22 audio features
	Parkinson Speech Dataset with Multiple Types of Sound Recordings[197].	40	40	–	–	–	–	Sample: 120 patients and 120 healthy individuals, 45 audio features
	Sarkar dataset[192]	20	20	6	14	10	10	26 speech features: vowels, numbers, and short phrases
62	TIMIT[198]	–	–	–	–	–	–	Dataset with 6300 sentences, 240 samples, 80 individuals (40 F and 40 M)
	Sarkar dataset[192]	20	20	6	14	10	10	26 speech features (vowels, numbers, short phrases)
		23	8	–	–	–	–	26 audio features (time and frequency information)

Table 6 (continued) | Acoustic datasets

Article	Repository/Source	#PD	#HC	#PDF	#PDM	#HCF	#HCM	Information
	Kay Elemetrics Disordered Voice Database (Max Little)[193]							
	DNSH, Clinical	90	-	43	47	–	–	Dataset with untreated and 54 treated participants
63	Clinical	–	–	–	–	–	–	1200. wav samples (500 HC, 700 PD), collected 2016–2019 in Andhra Pradesh, India
64	199	42	–	–	–	–	–	42 PD patients monitored for 6 months via telemonitoring
65	200,201	23	8	–	–	–	–	22 features: MDVP (Shimmer, F0, Jitter), HNR, nonlinear (DFA, Spread1, Spread2)
	Sarkar dataset[192]	20	20	6	14	10	10	26 features (parameters related to time and frequency)
	202	30	–	–	–	–	–	50 voice samples, 71 features extracted using 3 recorder types
66	Clinical	40	40	–	–	–	–	Spanish speaker
67	PC-GITA corpus[191]	50	50	–	–	–	–	Spanish speaker
	Saarbrücken Voice Database (SVD)[203]	668	687	–	–	–	–	1355 patients with 71 diseases
	Dataset of Vowels[204]	–	–	–	–	–	–	Audio in wav format, vowel sounds included
68	LSVT_voice_rehabilitation data set[205]	14	–	6	8	–	–	Participant age range: 51–69 years, 9 speech samples per participant, 5 pre-treatment, 4 post-treatment
	Sarkar dataset[192]	20	20	6	14	10	10	PD ranges aged from 43 to 77 years old, HC ranges in age range from 45 to 83 years old, 26 speech features (vowels, numbers, and short phrases)
69	GYENNO SCIENCE Parkinson, Clinical	30	15	–	–	–	–	25 F and 5 M, Age range: 37–75 years
	PC-GITA corpus[191]	50	50	–	–	–	–	Spanish speaker
70	CIEMPIESS[189]	–	–	–	–	–	–	Mexican Spanish podcast, 16717 samples (45 F, 96 M)
	190	–	–	–	–	–	–	135 children (5–15 years)
	PC-GITA corpus[191]	50	50	–	–	–	–	Spanish speaker
71	206	20	20	6	14	10	10	1208 samples, Audio in wav format
		–	–	–	–	–	–	Total of 28 individuals
72	Kay Elemetrics Disordered Voice Database (Max Little)[193]	23	8	–	–	–	–	22 features: MDVP (Shimmer, F0, Jitter), HNR, nonlinear (DFA, Spread1, Spread2)
73	Sarkar dataset[192]	20	20	6	14	10	10	1208 audio samples, 688 PD and 520 HC, (“a”, “o”, “u”) vowels, numbers (1-10), words, and short phrases
74	Parkinson’s Disease Diagnosis Dataset[207]	188	64	81	107	–	–	–
	201	23	8	–	–	–	–	–
75	Kay Elemetrics Disordered Voice Database (Max Little)[193]	23	8	–	–	–	–	22 features: MDVP (Fundamental Frequency, Shimmer, F0, Jitter), HNR, nonlinear (DFA, Spread1, Spread2)
76	Kay Elemetrics Disordered Voice Database (Max Little)[193]	23	8	–	–	–	–	22 features: MDVP (Shimmer, Fundamental Frequency, Jitter), HNR, nonlinear (DFA, Spread1, Spread2)
	PD Speech (PDS)[194]	188	64	81	107	41	23	Age range: 33–87 years
77	PD Speech (PDS)[194]	188	64	81	107	41	23	Age range: 33–87 years
78	Giuliano Parkinson’s Voice Dataset[208]	55	64	24	31	–	–	–
79	209	16	16	–	16	–	16	PD: average age (61 ± 12) HC: average age (62.6 ± 13.4)
	PC-GITA corpus[191]	50	50	25	25	25	25	PD: average age (62.2 ± 11.2) HC: average age (61.2 ± 11.3)
	Italian Parkinson’s Voice and Speech dataset[210]	28	22	9	19	12	10	PD: average age (67.2 ± 8.7) HC: average age (67.1 ± 5.2)
	211	28	13	20	8	4	9	The group mean age is (67.1 ± 6.3) years

Table 6 (continued) | Acoustic datasets

Article	Repository/Source	#PD	#HC	#PDF	#PDM	#HCF	#HCM	Information
53	PC-GITA[191]	50	50	25	25	25	25	PD: average age (71.6 ± 9.44) HC: average age (60.98 ± 9.46) 100 Colombian-Spanish speakers, Sustained vowel recordings
	Saarbrücken Voice Database (SVD)[212], Kiel Corpus[213]	–	42	–	–	25	17	Dysphonia: 44 individuals (22 F, 22 M), Laryngitis: 44 individuals (16 F, 28 M), HC: average age (61.74 ± 6.63) >2,000 voice recordings from 71 pathological and HC
80	Clinical	251	–	152	99	–	–	101 SCD patients (41 M and 60 F), PD: average age (71.6 ± 7.4) SCD: average age (64.3 ± 11.1) Native Japanese speaker
2	214	188	64	81	107	41	23	Features: 754 total (MFCCs, wavelet, time-frequency)
52	Voice Samples for Patients with Parkinson's Disease and Healthy Controls[215]	40	41	19	21	24	16	Vowel /a/ samples collected via telephone
81	200,201	23	8	–	–	–	–	The Max Little dataset included 195 audio samples
	Sarkar dataset[192]	20	20	–	–	–	–	Sarkar dataset: 1040 training samples, 168 testing samples

In this table, we detailed all the datasets in papers and compared participant demographics (number, gender, and health status: #PD => Parkinson's Disease, #HC => Number of Healthy Control Participants, #PDF => Number of Parkinson's disease Female Participants, #PDM => Number of Parkinson's disease Male Participants, #HCF => Number of Healthy Control Female Participants, #HCM => Number of Healthy Control Male Participants).

were used to evaluate the proposed approach, and high accuracy was achieved. Masood et al.[65] developed a two-level ensemble-based characteristic selection model aimed at early detection of PD using voice data analysis. In addition, this model was highly accurate. Similarly, García et al.[66] offered a high-accuracy two-level ensemble-based attribute selection model aimed at early detection of PD using voice data analysis. In a similar vein, Hireš et al.[67] suggested a method based on a CNN ensemble for diagnosing PD using voice data. Additionally, the results showed that this method provides high ROC (AUC), specificity, and accuracy.

Ma et al.[68] presented a deep dual-side learning ensemble model based on the deep sample learning (DSL) algorithm, combined with deep feature learning, for recognizing PD via speech. An embedded deep-stacked group sparse autoencoder (EGSAE) extracted high-quality deep features. The results demonstrated the method's high accuracy. Nevertheless, this approach required further validation and rebuilding of deep samples using a deeper neural network. Quan et al.[69] developed a model to diagnose PD via speech signals and an end-to-end learning architecture. The proposed method employed a combination of time-distributed two-dimensional CNN (2D-CNN) and a one-dimensional CNN (1D-CNN) to extract dynamic features from speech signal time series and remove interdependence. Moreover, the method demonstrated high accuracy based on the obtained results.

In a similar vein, Chen et al.[70] suggested an IP-based chimpanzee optimization algorithm (IPChOA) to enhance DCNN. Furthermore, the effectiveness of this method was assessed using speech signals from both PD and cleft lip and palate cases. The results showed that this method reached a high level of accuracy. Meanwhile, Khaskhoussy and Ayed[71] presented an approach for diagnosing PD using ML techniques and speech processing, while also identifying early changes associated with the condition. They also used features like the mel-frequency brain coefficients-gaussian mixture model (MFCC-GMM). This study demonstrated high accuracy. In addition, Biswas et al.[72] introduced an approach for early diagnosis of PD using an ensemble-based ML model called the ensembled expert system for diagnosis of PD (EESDPD). The suggested method had high accuracy, recall, and F1-score.

Wei Liu et al.[73] developed a model based on speech features and an ANN to diagnose PD. The results showed that the proposed model has high accuracy, sensitivity, and AUC. Also, Yuan et al.[74] investigated the detection of PD utilizing speech signals and ML approaches. Using the ReLU activation function, the authors developed a DNN architecture with multiple concealed layers. In addition, the minimum redundancy-maximum correlation (mRMR) method was used to identify key features. The outcomes showed that the suggested method obtained a high F1-score, accuracy, and MCC. Moreover, Kamalakannan et al.[75] evaluated different classification methods for diagnosing PD. J48, SVM, and multilayer perceptron neural network (MPNN) methods achieved the highest accuracy.

Saleh et al.[76] presented a method for identifying PD based on audio data as well as ML and ANN approaches. Additionally, this method achieved a high level of accuracy. Also, Devarajan et al.[77] examined the application of ML methods to enhance the quality of diagnoses in the healthcare field, with a specific focus on PD. In addition, they have devised ML ensembles to identify PD, emphasizing the utilization of nonclinical patient data for early detection.

Guatelli et al.[78] offered a technique for diagnosing PD based on neural networks utilizing the extreme learning machine (ELM) and acoustic signal spectrograms. The results demonstrated the practical application potential of this method, particularly considering the lower cost and shorter time required for training compared to traditional CNN-based methods. Additionally, the suggested method had a high level of accuracy. Also, Hireš et al.[79] measured the efficiency of two varied ML models, namely CNN and XGboost, to detect computerized PD. The Authors evaluated their model regarding accuracy, specificity, sensitivity, and AUC via four datasets. The result revealed that even if the result in a database is acceptable, it could not be the same as others.

Pah et al.[53] evaluated the efficacy of ML approaches in PD identification. They also investigated other voice disorders, such as laryngitis and dysphonia. Moreover, Eguchi et al.[80] proposed a methodology that employed a transformer model to distinguish between patients afflicted with PD and those with spinal cerebellar degeneration using speech data. Also, accuracy and AUC metrics for the proposed model were comparatively high. In addition, Dhanalakshmi et al.[2] provided a technique that uses speech features and ML techniques to detect early-stage PD. In addition, they focused on optimizing classification performance and addressing

Table 7 | A comparison of acoustic data/features papers

Article	Main idea	Tools	Applied algorithms	Advantages	Disadvantages
55	DCNN-based model for pathological speech detection	• MATLAB	• DCNN • SpecAugment • WOA	• High precision • High sensitivity • High accuracy • High specificity • High F1-score	• Not evaluating a large hybrid dataset • Lack of multilingual data • Not investigating other classifiers or forms of DCNN
56	ML-based approach using speech signal analysis	• Not mentioned	• SVM • CNN • GMM • MLP • Adam optimizer • Expectation Maximization	• High F1-score • High accuracy • High sensitivity • High specificity • High precision	• Low scalability
57	DL and ML-based PD diagnosis via speech signals	• Not mentioned	• CNN • RF	• High accuracy • High precision • High recall • High AUC • High F1-score • High specificity	• Low scalability
58	Ensemble-based model for PD detection using voice signals	• Not mentioned	• DNN • EOFSC	• High accuracy	• Not analyzing parameters such as F1-score and precision
59	PD detection using DL and ACSA algorithm	• Python (TensorFlow)	• ACSA • AutoEncoder	• High accuracy • High sensitivity • High specificity	• Not investigating unsupervised classification methods • Lack of identification of scrunched feature vectors
60	Acoustic-based PD detection using IGKNN algorithm	• Not mentioned	• KNN	• High accuracy • High recall	• High cost • Poor handling of multi-dimensional features
61	Optimized ResNet50 model for PD diagnosis	• Not mentioned	• ResNet50 • GDABC	• High accuracy	• High time consumption
54	Two-phase ML-based system for early PD diagnosis	• Not mentioned	• LightGBM	• High accuracy • High AUC • High precision • High F1-score	• Not optimizing the hyperparameters
62	TL-based approaches for PD detection via speech	• MATLAB	• KNN • SVM • FT & KNN • FT&SVM • TSTL	• High sensitivity • High accuracy • High specificity	• Lack of multimodal data
63	Speech-based PD diagnosis using a neural network	• Not mentioned	• PNN	• High accuracy	• Lack of mobile apps for PD detection
64	PD detection using DNN on voice data	• Not mentioned	• DNN	• High accuracy	• Low scalability
65	Feature correlation–driven ensemble model for early PD diagnosis	• Python	• MLP • NB • KNN • SVM • DT	• High accuracy • High AUC	• lack of clinical validation
66	ML-based feature selection for PD diagnosis	• Not mentioned	• SVM	• High accuracy	• Low scalability • Low quality of the recording device
67	CNN on speech recordings for PD diagnosis	• Not mentioned	• ResNet50 • Xception • SGD	• High accuracy • High specificity • High sensitivity • Language independence • High ROC (AUC)	• Low scalability • Not considering the drug effects
68	DL-based PD diagnosis via feature extraction	• MATLAB	• DSL • L1R&FS • SVM • EGSAE • IMC • WFM-DSS	• High accuracy	• Not exploring methods to rebuild deep samples • Not applying more DNN architectures
69	End-to-end DL model for PD diagnosis	• Python (Scikit-learn, librosa) • NeuroSpeech	• 2D-CNN • 1D-CNN	• High accuracy • High specificity	• Low scalability
70	Optimized DCNN for PD & cleft lip and palate	• MATLAB	• DCNN • ChOA • IPChOA • SpecAugment	• High specificity • High accuracy • High F1-score • High sensitivity • High precision	• Lack of other languages in the dataset • Not investigating other CNNs or GAN

Table 7 (continued) | A comparison of acoustic data/features papers

Article	Main idea	Tools	Applied algorithms	Advantages	Disadvantages
					• Not evaluating other non-pathology-related cues
71	Speech-based PD diagnosis using SVM	• Not mentioned	• SVM • FFT • DCT • AutoEncoder	• High accuracy • High recall • High F1-score	• Low scalability • Lack of phonation, prosody, and speech rhythm evaluation
72	Stacking ensemble model for PD diagnosis	• Python	• XGBoost • RF • LR • Stacking ensemble technique	• High accuracy • High F1-score • High precision • High recall	• Low scalability • Lack of comprehensive clinical data
73	PD diagnosis using ANN and speech features	• Statistical software	• ANN • Levenberg-Marquardt • Back-propagation algorithm	• High accuracy • High sensitivity • High AUC • High specificity	• Not evaluating additional languages • Low scalability
74	PD diagnosis via ML and speech signals	• Not mentioned	• mRMR • RF • LR • KNN • DNN	• High F1-score • High accuracy • High MCC	• Unclear data accuracy and privacy • Lack of patient usability and convenience evaluation
75	Evaluation of ML classifiers for the detection of PD	• WEKA	• J48 • NB-tree • MPNN • SVM	• High accuracy	• Low scalability • Not evaluating other advanced ML algorithms • Not evaluating feature selection techniques
76	Ensemble voting for PD detection via acoustic data	• Python	• LR • Ridge classifier • SGD • PAC • KNN • Extra tree • DT • SVC • Gaussian NB • AdaBoost • Bagging classifier • RF • Gaussian process classifier • GB • LDA • QDA • XGBoost • MLP	• High accuracy • High precision • High recall • High AUC • High F1-score • High specificity	• Low scalability
77	ML approach for early PD detection via healthcare decision-making	• R programming	• NB • ANN • DT • RF	• High accuracy	• Lack of applying DL methods • Motor symptoms were not considered
78	PD classification based on audio data using ELM	• MATLAB	• Short-time Fourier transform • ELM • CNN	• Low training time • Low cost • High accuracy	• Low scalability
79	ML-based voice analysis for PD detection	• Not mentioned	• XGBoost • CNN	• High accuracy • High specificity • High sensitivity • High AUC (ROC)	• Only acoustic data intended • Only one of the datasets determines gender
53	Assessing ML reliability for PD voice-based diagnosis	• MATLAB	• SVM • Relief-F algorithm	• High recall • High accuracy	• Low scalability
80	DNN-based classification of PD and SCD using speech data	• Not mentioned	• DNN • Patchout faSt spectrogram transformer	• High AUC (ROC) • High accuracy • High specificity • High sensitivity • Non-invasive screening method	• Low scalability
2	PD diagnosis using speech features	• Python (Scikit-learn)	• SVM • XGBoost • RF • KNN • DT • LR	• High accuracy • High F1-score • High AUC (ROC)	• Low scalability

Table 7 (continued) | A comparison of acoustic data/features papers

Article	Main idea	Tools	Applied algorithms	Advantages	Disadvantages
			• PCA • SMOTE-ENN		
52	ML-based acoustic analysis for PD detection	• R programming • Python	• CNN • RF • LR	• High AUC (ROC)	• Not checking the health status of HC • Low scalability
81	Audio-based PD detection via ML and DL	• Python • Praat	• SVM • DNN • Adam optimizer • HGSA	• High specificity • High sensitivity • High accuracy • High MCC	• Not examining the severity of PD in patients • Low scalability

This table includes key concepts (main ideas), the utilized tools, the applied algorithms, advantages, and disadvantages.

Table 8 | Acoustic evaluation metrics

Article	Accuracy	Sensitivity	Specificity	F1-score	Precision	AUC (ROC)	FPR	MCC	RMSE	Kappa coefficient	PRC
55	+	+	+	+	+	+					
56	+	+	+	+	+	+					
57	+	+		+	+	+	+				
58	+	+	+			+		+			
59	+	+	+	+					+		
60	+	+		+	+			+		+	+
61		+	+	+	+						
54	+	+		+	+	+					
62	+	+	+								
63	+	+	+		+						
64	+	+		+	+						
65	+	+		+	+	+		+		+	
66	+	+	+			+					
67	+	+	+			+					
68	+	+	+								
69	+	+	+	+				+	+		
70	+	+	+	+	+	+					
71	+	+		+	+						
72	+	+		+	+						
73	+	+	+			+					
74	+	+		+	+			+			
75	+										
76	+	+	+	+	+			+			
77	+	+	+		+				+		
78	+	+	+								
79	+	+	+			+					
53	+	+	+	+							
80	+	+	+			+	+				
2	+	+		+	+	+	+				
52	+					+					
81	+	+	+					+			

Plus (+) means the metric that has been evaluated in the paper, and blank cells mean the metric has not been evaluated in the paper.

imbalanced datasets. The results of the proposed method indicated that it has high accuracy and AUC (ROC).

Iyer et al.[52] proposed a method leveraging data obtained from recorded sounds and applying ML approaches to distinguish Parkinson's patients from healthy individuals. Also, the proposed method achieved a high AUC (ROC). Ali et al.[81] developed a two-stage diagnostic system based on ML and DL to improve PD diagnosis using speech data. In addition, the proposed system provided acceptable accuracy.

Qualitative analysis of acoustic-based approaches. An analytical review of acoustic-based approaches has been conducted, highlighting

Table 9 | Medical imaging datasets

Article	Repository/Source	#PD	#HC	#PDF	#PDM	#HCF	#HCM	Information
85	Clinical	305	227	132	173	123	104	From Chang Gung Memorial Hospital, Linkou, Collected from 2008 to December 2017
86	Clinical	120	–	58	62	–	–	External validation includes 43 PD patients (24 M and 19 F)
87	Clinical	144	55	–	–	–	–	Collected from May 2015 – Mar 2020, Age: 18–75 years, Disease duration: <4 years
88	Parkinson's Disease Society Brain Bank[216]	247	125	–	–	–	–	Collected from 2014 to March 2018, 23 incomplete clinical data
89	217	22	–	12	10	–	–	Age range: 48–77 years
	218,219			–	–	–	–	In total, 82 individuals participated
90	PPMI[220]	252	–	84	168	–	–	Age: 35–86 (average 62.4), DAT SPECT images: Year 0 & Year 1
91	Clinical	267	160	–	–	–	–	The data includes DAT SPECT images from years 0 and 1
92	PPMI[220,221]	92	22	–	–	–	–	–
	222	14	14	–	–	–	–	–
		15	15	–	–	–	–	–
93	219,223	860	350	–	–	–	–	350 patients from ADNI, 350 patients from PPMI, 160 patients from NIFD, 350 healthy people
94	PPMI[220] CCNA[224] BioCog[225] PD-MCI Calgary[226] PD-MCI Montreal[227] C-BIG3[228] NEUROCON dataset[229] Tao Wu dataset[229] OpenNeuro Olfactory dysfunction[230,231], Hamburg dataset[232] UK Biobank, OASIS3[233] SALD[234]	1024	1017	–	–	–	–	Data collected from 13 separate studies
95	PPMI[220]	236	82	–	–	–	–	–
96	Clinical	115	115	–	–	–	–	–
97	Clinical	–	–	–	–	–	–	In total, 210 individuals participated
98	Clinical	43	55	–	–	–	–	–
99	Clinical	103	255	–	–	–	–	–
	Clinical	22	26	–	–	–	–	–
84	Clinical	–	–	–	–	–	–	239 individuals, 137 from Municipal Medical Center, 102 from Kanazawa University Hospital
100	Clinical	95	–	–	–	–	–	–
101	NewHandPD[235]	31	35	–	–	–	–	66 individuals
102	Clinical	96	–	–	–	–	–	From September 2016 and December 2020
103	236,237	–	–	–	–	–	–	582 MRI images, 249 MRI for PD
104	PPMI[220]	255	249	–	–	–	–	–
105	Ruijin Hospital, Clinical	92	287	–	–	–	–	–
	First Affiliated Hospital of Zhengzhou University, Clinical	83	72	–	–	–	–	–
106	PPMI[220]	357	210	125	232	74	136	1213 images, PD: Average age (61.83 ± 9.89) HC: Average age (60.77 ± 11.29)
107	PPMI[220,221]	170	170	–	–	–	–	2720 SPECT DaTSCAN images (1360 PD and 1360 HC)
108	PPMI[220,221]	366	163	–	–	–	–	SWEED Group: 63 individuals, Prodromal Group: 48 individuals

Table 9 (continued) | Medical imaging datasets

Article	Repository/Source	#PD	#HC	#PDF	#PDM	#HCF	#HCM	Information
83	Clinical	854	775	–	–	–	–	1758 TCS images, Accessed from Sep 2019 to May 2022
109	PPMI[220,221]	213	213	88	125	80	133	HC: average age (59.44 ± 11.33) PD: average age (61.27 ± 9.71) 213 age-matched T1-weighted MRI scans, Randomly selected
	NEUROCRON[229]	27	16	10	17	12	4	HC: average age (67.63 ± 11.89) PD: average age (68.13 ± 12.86)
	Tao Wu dataset[229]	18	18	8	10	7	11	HC: average age (67.63 ± 11.89) PD: average age (65.61 ± 4.45)
110	UK Biobank[238]	84	84	–	–	–	–	Two groups: Prevalent and Incident datasets, Prevalent Dataset: 154 fundus images (77 PD, 77 HC), Incident dataset: 92 images (46 PD, 46 HC)

In this table, we detailed all the datasets in papers and compared participant demographics (number, gender, and health status: #PD => Parkinson's Disease, #HC => Number of Healthy Control Participants, #PDF => Number of Parkinson's disease Female Participants, #PDM => Number of Parkinson's disease Male Participants, #HCF => Number of Healthy Control Female Participants, #HCM => Number of Healthy Control Male Participants).

the key strengths, weaknesses, opportunities, and threats, which are discussed in detail below:

- Strengths: Our analysis of acoustic-based approaches reveals three considerable advantages. Firstly, it is low-cost. Secondly, it is non-invasive, so it does not provide any thread for the patients. Lastly, the solution they provide is scalable and can be utilized in real-world settings. Moreover, the very first important feature is that vocal symptoms may appear years before other symptoms, and as a result, it makes early detection possible. The accessibility is outstanding since it can be used as a mobile application, making it widely available and suitable for real-world usage.
- Weaknesses: On the other hand, it needs to be mentioned that the diagnosis models may not work effectively in noisy environments, leading to misdiagnosis. The lack of diverse language and culturally rich datasets may produce bias in results. On top of this, models in this approach did not suggest a combination of vocal data with other forms of data, such as movement and medical imaging.
- Opportunities: Addressing these weaknesses can improve model reliability. Employing noise reduction and advanced techniques in speech analysis is essential. Moreover, using diverse databases that reflect a variety of ethnicities and languages may improve the model's capabilities. The combination of acoustic data and other forms of data, widely utilized to detect PD, may lead to a better resolution.
- Threats: Several vital concerns should be addressed in future work. As the volume of data increases, maintaining standardization becomes more challenging, which may directly impact the real-world applicability of the models. Additionally, reliance solely on acoustic data could lead to misclassification.

Medical imaging

Neurological disorders directly affect the brain; therefore, a visual brain image can be suitable for detecting a disease[82]. PET and SPECT are two approaches that can detect PD; however, they are less commonly utilized due to their cost and invasiveness[83]. Nevertheless, alternative methods, such as dopamine transporters (DAT) scanned applying 123I-ioflupane and magnetic resonance imaging (MRI), can be employed alongside ML for detecting PD[84]. Table 9 presents the databases utilized in medical imaging, including such details as the name of the data source, type of images, number of samples, and additional information like the participants' age and data format. Table 10 compares the papers based on their main ideas, employed tools, and algorithms, as well as their advantages and disadvantages. Additionally, Table 11 provides the critical parameters used in the evaluation process. Section "Review of medical imaging-based approaches" reviews medical imaging-based approaches. In the end, a qualitative analysis of approaches, including strengths, weaknesses, opportunities, and threats, is performed in Section "Qualitative analysis of medical imaging-based approaches".

Review of medical imaging-based approaches. Zhao et al.[85] presented a CNN-based method utilizing diffusion tensor imaging to assess the diagnostic performance of a hybrid architecture for detecting PD across multiple brain regions. The study also employed a greedy algorithm to combine various areas for final prediction. The proposed method demonstrated a high AUC according to the results. Likewise, Shibata et al.[86] suggested a model to identify moderate cognitive impairment (MCI) in PD patients using quantitative susceptibility mapping (QSM) images. The study utilized ML algorithms, including light GB, extreme GB, and RF. The results indicated that RF achieved a satisfactory level of performance and accuracy.

Gaurav et al.[87] introduced a CNN-based framework for automatically assessing and segmenting neuromelanin in the SN of early-stage Parkinson's patients. According to the results, the proposed framework achieved high accuracy and repeatability. Furthermore, the method could process large datasets in a significantly shorter time. Also, Nakano et al.[88] investigated the effects of motor and non-motor symptoms on health-related quality of life (HRQoL) and identified brain networks associated with using MRI in Parkinson's patients. Next, Dünnwald et al.[89] suggested a CNN model for automating the extraction of biomarker data from multi-rater segmentation and multi-scale localization to detect individuals with PD. They validated their approach by evaluating the model using metrics such as the DSC and Euclidean distance.

Adams et al.[90] presented a technique to predict the motor performance of individuals with PD. They employed CNN to interpret data from a combination of DAT SPECT imaging and clinical measures, including motor segment assessment. The results demonstrated that this DL-based combination improved the prediction of movement performance in Parkinson's patients. Also, Shin et al.[91] presented an algorithm that employs a DL-based CNN to interpret nigrosome 1 susceptibility map-weighted imaging (SMwI) to assess nigrostriatal degeneration in idiopathic PD (IPD). The results demonstrated that their proposed method provides rapid and

Table 10 | A comparison of medical imaging papers

Article	Main idea	Tools	Applied algorithms	Advantages	Disadvantages
85	PD diagnosis using diffusion MRI and CNN	•SPSS software	•Greedy algorithm • CNN	• High AUC	• Low scalability • Not utilizing various scanners
86	ML-based model for PD cognitive impairment diagnostics	• FSL • Python	• RF • XGBoost • LightGBM • V-SHARP algorithm • Grid search algorithm	• High accuracy	• Insufficient validation • Low scalability
87	CNN framework for SNPC segmentation	• MATLAB • FSL • SPM • MRtrix3 • R programming • Python (Scikit-learn)	• CNN	• High accuracy • High AUC • Reduced processing time for large datasets • High reproducibility	• No external validation • Not evaluating the effects of medicines on neuromelanin • No SNPC topography analysis • Low scalability
88	PD symptoms impact on HRQoL and neural network mapping	• Python • (Scipy, Python factor_analyzer) • MATLAB • FSL • ART	• Multiple regression • RF	• A thorough analysis of the various factors affecting HRQoL	• Not analyzing parameters such as accuracy and sensitivity • Lack of sleep-related questions • Lack of longitudinal studies about patients
89	DL-based PD diagnosis via MRI and locus coeruleus classification	• Not mentioned	• CNN	• High precision	• Not analyzing parameters such as sensitivity • Considering the limited amount of data
90	DL-based motor performance prediction in PD	• Python	• CNN	• High-performance prediction	• Low scalability • Not evaluating metrics such as sensitivity and specificity
91	Deep complex neural networks for classifying IPD	• MedCalc • R programming	• CNN • YOLOv3 algorithm	• High diagnostic performance • High accuracy	• Not investigating other CNNs • Low scalability
92	Rs-fMRI and topological ML for diagnosing PD	• MATLAB • Python	• UMAP • DNN • SVM • GB	• High accuracy	• Lack of high-quality data • Insufficient analysis of scanning device and embedded parameters • Not evaluating classification performance of different stages of PD
93	DCNN model for dementia diagnosis via FDG-PET	• Python	• DCNN • GAN • Adam optimizer	• High accuracy • High specificity • High sensitivity	• Not investigating other types of dementia • Not assessing non-imaging features
94	Interpretable DL model for PD diagnosis	• ANT	• CNN Jacobians • N3 • SmoothGrad	• High accuracy • High AUC(ROC) • High precision • High sensitivity • High specificity	• Not investigating other DL architectures • Lack of utilization of multimodal MRI data
95	3D and 2D CNN to distinguish PD from healthy	• Python	• 2D CNN • 3D CNN	• High AUC • High accuracy	• Not applying parallelization to increase performance
96	Detection of PD using parameter-weighted matrices	• MATLAB	• CNN	• High AUC (ROC)	• Not able to detect at early PD stages • Lack of varied data • Not verifying the model's effectiveness externally
97	ML-based PD diagnosis via I-123 FP-CIT scans	• MATLAB	• SVM	• High accuracy • High sensitivity • High specificity	• Applying a relatively small number of patients • Not considering additional parameters • Limited evaluation of diverse methods
98	[18 F] DOPA PET/CT and CNN for PD classification	• Python	• CNN	• High accuracy • High specificity • High sensitivity	• Not performing a regional analysis • No neuropathologic confirmation
99	DL model for PD/HC separation via FDG-PET	• MATLAB	• Radiomatic DL	• High accuracy	• Lack of general data such as race, and nationality • Low scalability • Not involving other kinds of data, such as MRI
84	ML-based PD detection using 123I-ioflupane images	• Not mentioned	• Gradient boosted trees • LR • KNN	• High AUC	• Not considering personal information such as gender • Low scalability

Table 10 (continued) | A comparison of medical imaging papers

Article	Main idea	Tools	Applied algorithms	Advantages	Disadvantages
100	DL model for PD detection via facial expression	• Python (PyTorch)	• StarGAN • CNN	• High accuracy	• Limited database diversity
101	DL approach for diagnosing PD from handwriting	• Not mentioned	• ResNet50 • VGG19 • INCEPTION-V3 • KNN	• High accuracy • High precision	• Not analyzing F1-score and specificity
102	Automatic PD subtype diagnosis using SVM	• Python (Scikit-learn)	• SVM • LASSO • SHAP	• High AUC	• Low scalability • No analysis of genetic biomarkers
103	DL-based PD classification using MRI data	• Not mentioned	• ResNeXt	• High accuracy	• Lack of clinical validation
104	DL framework for PD classification via MRI	• Python (PyTorch)	• CNN • Adam optimizer	• High specificity • High accuracy • High sensitivity • High F1-score	• Existence of uncertain details in the model of DL • Low scalability • Not evaluating 3D imaging data
105	PD identification via DL on T1-weighted and QSM scans	• Python (PyTorch) • R programming	• SE-ResNeXt50 • CNN	• High AUC • High accuracy	• Limited number of centers • Low scalability • Reduced performance due to inaccurate segmentation
106	DNN-based PD diagnosis using SPECT images	• Python (Scikit-Learn) • OpenCV	• PARNet	• High specificity • High accuracy • High sensitivity • High F1-score • High precession	• Low scalability • Not examining a large domain of patients
107	PD classification with CNN using DaTSCAN images	• Not mentioned	• MobileNet-V2 • EFFICIENTNET-B0	• High accuracy	• Low scalability • Potential overfitting issues
108	PD diagnosis with GNNs through MRI scans	• Python (PyTorch)	• GNN • Sparsity ATopk model	• High F1-score	• Not evaluating multimodal data
83	DCNN-based PD diagnosis using TCS images	• ImageJ software • Python (PyTorch) • R programing	• DCNN	• High accuracy • High PPV • High F1-score • High sensitivity	• Low scalability • Retrospective and single-center nature • Not evaluating different PD subtypes
109	The impact of CNN design and data leakage on PD diagnosis using MRI	• 3D Slicer • FSL	• CNN • N4 algorithm	• High accuracy	• Unequal dataset conditions • Limited to T1-weighted MRI • Not using multicenter or multimodal approaches
110	DL-based PD detection using retinal fundus images	• Not mentioned	• LR • SVM • Elastic Net • ResNet50 • Inception-V3 • GoogleNet • VGG-16	• High NPV • High sensitivity	• Low scalability

This table includes key concepts (main ideas), the utilized tools, the applied algorithms, advantages, and disadvantages.

precise results for IPD diagnosis. Similarly, Xu et al.[92] presented a technique using a topological ML approach and resting-state functional MRI (rs-fMRI) data to develop a biomarker for early PD detection and treatment evaluation. Based on the findings, the suggested method reached a high degree of accuracy.

Noella and Priyadarshini[93] introduced a DCNN-based model to identify PD and AD using fluorodeoxyglucose PET brain scans. The results showed that the model had high specificity, sensitivity, and accuracy. Furthermore, Camacho et al.[94] developed an explainable model to identify Parkinson's patients based on DL methods utilizing T1-weighted MRI data. The results of the presented model showed high precision and accuracy. In the same vein, Vyas et al.[95] introduced two distinct neural network models, a 2D CNN and a 3D CNN, to detect PD using MRI scans. In addition, both models demonstrated acceptable effectiveness and performance.

Yasaka et al.[96] proposed a CNN model to assess the potential of detecting PD individuals using parameter weighting and the number of streamlines. Results showed moderate performance in terms of AUC. However, the external validation of the presented model's performance had not been conducted. Additionally, Dotinga et al.[97] developed an SVM-based model to distinguish PD patients from healthy people using I-123 FP-CIT images. This model showed high accuracy, sensitivity, and specificity.

Piccardo et al.[98] used a 3D CNN-based method combined with the analysis of brain [18 F] DOPA PET/CT scans to diagnose PD. This method showed acceptable accuracy and robustness. Moreover, Sun et al[99]. introduced a radiomic DL model. This study used [18 F] FDG PET imaging to diagnose PD, and the presented model showed significant accuracy. Nakajima et al.[84] proposed an ML approach to detect dementia and PD based on 123I-ioflupane images. According to the results, the proposed approach had a high AUC (ROC).

Huang et al.[100] proposed a method for improving the diagnosis of PD through emotional facial expressions and DL techniques. The presented model was evaluated using four datasets and achieved high accuracy. Also, Abdullah et al.[101] introduced a framework based on TL to diagnose PD through handwriting analysis. Features gathered from the model were optimized using a GA. The evaluations showed high accuracy and effectiveness of the framework. In the same light, Pang et al.[102] employed ML to

Table 11 | Medical imaging evaluation metrics

Article	Accuracy	Sensitivity	Specificity	F1-score	Precision	AUC(ROC)	NPV	FPR	MCC	Standard deviation	ICC	MSE	Pearson's correlation coefficients	Euclidean distance	DSC	MAE
85	+	+	+	+	+	+	+									
86	+	+	+			+		+		+						
87	+					+					+		+			
88												+				
89					+									+	+	
90																+
91	+	+	+			+										
92	+	+														
93	+	+	+	+												
94	+	+	+		+	+										
95		+		+	+	+		+								
96	+	+	+													
97	+	+	+	+												
98	+	+	+		+		+									
99	+	+	+													
84	+	+		+	+	+										
100	+															
101	+	+			+	+										
102	+	+	+		+	+	+									
103	+	+		+	+											
104	+	+	+	+				+	+							
105	+	+	+	+	+		+									
106	+	+	+	+	+	+										
107	+	+		+	+	+										
108	+	+		+	+											
83	+	+	+	+	+	+	+									
109	+	+	+	+	+											
110	+	+	+	+	+		+									

Plus (+) means the metric that has been evaluated in the paper, and blank cells mean the metric has not been evaluated in the paper.

assess a model utilizing multi-level indicators of resting-state functional magnetic resonance imaging (rsfMRI) for detecting different motor subtypes of PD patients. The findings showed that the proposed model achieved a notable AUC value.

Balnarsaiah et al.[103] demonstrated a method for diagnosing PD based on DL approaches applied to MRI data. The Residual network (ResNeXt) architecture was utilized in this study to classify brain MRI images to identify Parkinson's patients. As a result, the suggested method demonstrated significant accuracy in this research. Also, Xinchun Cui et al.[104] presented a method for classifying PD using MRI T2 slices. By combining DL with a multi-branch feature processing module and multi-scale attention guidance, the authors created a method to extract features to enhance classification performance. Based on the outcomes, the proposed method obtained a high percentage of sensitivity, F1-score, accuracy, and specificity, demonstrating its diagnostic efficacy.

Wang et al.[105] offered an approach based on DL to detect PD with QSM and T1-weighted information automatically. The approach compromised CNN and squeeze and excitation (SE)-ResNeXt50 models that analyze image data. Evaluation results demonstrated the mentioned model's performance in terms of AUC. Also, Keles et al.[106] offered a DNN model for identifying PD patients by the SPECT images. They used 1231 images to validate their models. Outcomes showed the quality of the model concerning specificity, precision, accuracy, F1-score, and sensitivity. Besides, Khachnaoui et al.[107] presented a computer-aided system to diagnose PD via pre-trained CNN models, bilinear pooling, and TL. These models were trained by ImageNet. The results showed the model's performance in terms of accuracy.

Zhang et al.[108] presented a method for predicting PD using graph neural networks (GNNs) applied to MRI data. This study addressed two main issues: the efficiency of constructing graphs from MRI data and the overfitting of small data. Also, Ding et al.[83] developed a modified transcranial sonography (TCS) technique employing the DCNN model to predict PD. Further, this model demonstrated higher accuracy, sensitivity, PPV, and F1-score. Moreover, Veetil et al.[109] proposed a method using CNN models and T1-weighted MRI data to diagnose PD. In addition, simulations were used to investigate the problem of data leakage and high accuracy was also achieved. Tran et al.[110] developed DL models for diagnosing PD based on retinal fundus images. Also, they achieved acceptable NPV and sensitivity.

Qualitative analysis of medical imaging-based approaches. The narrow and critical review of medical-imaging-based approaches papers derived key strengths, weaknesses, opportunities, and threats, detailed below:

- Strengths: Two of the most apparent results of reviewing the medical imaging-based approach are that first, via image analysis, a good understanding of brain visualization and its relation to PD has been

Table 12 | Movement data datasets

Article	Repository/Source	#PD	#HC	#PDF	#PDM	#HCF	#HCM	Information
8	239 PaHaW dataset[240]	29	18	9	20	8	10	PD: average 57.9 HC: average (61.6 ± 8.8)
	241 PaHaW dataset[240]	29	26	13	16	14	12	HC: average age 39.31 PD: average age 66.80
	242 PaHaW dataset[240]	35	29	13	22	11	18	HC: average age (64.5 ± 6.8) PD: average age (67.2 ± 9.1)
4	Oxford Parkinson's Disease Telemonitoring Dataset[243]	–	–	–	–	–	–	16 features, 5875 records, 2 outputs, 28 M and 14 F patients
114	Daphnet Freezing of Gait Dataset[244]	10	–	3	7	–	–	PD: average age 66.40 Walking tasks: Straight hallway, randomly initiating stops and turns, daily activity
115	245	12	–	2	10	–	–	–
116	Clinical	4	4	–	–	–	–	Patients' age 63, 67, 72, and 73
117	239 PaHaW dataset[240]	29	18	9	20	8	10	HC: average age (57.9 ± 6.7) PD: average age (61.6 ± 8.8)
	241 PaHaW dataset[240]	29	26	13	16	14	12	HC: average age 39.31 PD: average age 66.80
	242 PaHaW dataset[240]	35	29	13	22	11	18	HC: average age (64.5 ± 6.8) PD: average age (67.2 ± 9.1)
118	246	21	–	3	18	–	–	PD: average (69.3 ± 9.7)
	247,248	38	21	–	–	–	–	28 M and 10 F, Average age: (70.7 ± 8.2) years
	249	59	–	22	37	–	–	Average age: (69.2 ± 10.2) years
119	Clinical	55	31	–	–	–	–	Total: 113 individuals (56 M / 57 F), 13 with essential tremors, 4 with other diagnoses
120	Clinical	30	30	–	–	–	–	A total of 90 individuals, Thirty Alzheimer's patients
113	Clinical	42	24	–	–	–	–	83 subjects (41 M / 41 F), 13 with essential tremors, 2 with other tremor disorders, 2 undiagnosed, Age range: 22–84 years
121	PaHaW dataset[240]	–	–	–	–	–	–	Handwriting: 75 (37 PD, 38 HC), Spiral drawing: 69 (33 PD, 36 HC)
	NewHandPD[235]	31	35	–	–	–	–	A total of 66 individuals
122	Clinical	18	–	7	11	–	–	The age range is from 60 to 84 years, the wearable biomechatronic laboratory of Western University
123	Clinical	9	14	–	–	–	–	33 participants, 10 with multiple sclerosis, Self-paced treadmill walking tasks performed
124	Clinical	32	16	15	17	10	6	PD: aged from 52 to 84 HC: aged from 56 to 85
125	Clinical	45	45	15	26	18	25	3 PD and 2 control subjects were excluded, PD: average age: (68.0 ± 9.9) HC: average age: (67.0 ± 9.4)
126	Clinical	–	–	–	–	–	–	6 rats with medial forebrain bundle lesion, 6 normal rats as a control group
127	Clinical	6	6	–	–	–	–	A total of 12 individuals, PD: average age 40 HC: average age 52, Sensors to record the upper limb, Goniometer-based upper limb tracking, 4 move tasks
128	Clinical	58	29	20	38	24	5	87 individuals
129	PhysioNet[250,251]	93	73	34	59	33	40	Average age: 66.3 years
130	Clinical	20	73	–	–	30	43	A total of 93 individuals, HC: average age 66.3

Table 12 (continued) | Movement data datasets

Article	Repository/Source	#PD	#HC	#PDF	#PDM	#HCF	#HCM	Information
131	Clinical	63	63	–	–	–	–	–
132	Clinical	24	–	–	–	–	–	PD: average age (74.1 ± 6.7) HC: average age (74.1 ± 9.1)
	252	37	–	–	–	–	–	PD: average age (69.3 ± 10.9) HC: average age (74.1 ± 9.1)
133	Clinical[253]	–	–	–	–	–	–	Total individuals 39, 25 M, 7 F
134	254	42	43	–	–	–	–	PD: average age of 59.0 HC: average age of 60.1
	255	100	130	–	–	–	–	–
135	239 PaHaW dataset[240]	29	18	9	20	8	10	HC: average age (57.9 ± 6.7) PD: average age (61.6 ± 8.8)
	241 PaHaW dataset[240]	29	26	13	16	14	12	HC: average age 39.31 PD: average age 66.80
	242 PaHaW dataset[240]	35	29	13	22	11	18	HC: average age (64.5 ± 6.8) PD: average age (67.2 ± 9.1)
136	256,257	332	100	–	–	–	–	–
137	258	37	38	–	–	–	–	–
138	NewHandPD[235]	31	35	–	–	–	–	66 individuals
139	Clinical	24	24	–	–	–	–	17 patients with other tremor-related neurological diseases, 44 M and 39 F
140	Clinical	50	50	28	22	28	22	A total of 100 individuals, HC: average age (63.3 ± 8.6) PD: average age (63.6 ± 7.2)
112	Clinical	74	–	32	42	–	–	average age: 64.6 Collected data via 6 sensors, gait (2 min), sway (30 sec)
141	Clinical	276	79	81	195	50	29	Differential diagnoses: 114 individuals (57 M, 57 F), Atypical Parkinsonism: 15 individuals (8 M, 7 F), Essential Tremor: 28 individuals (18 M, 10 F), Multiple Sclerosis: 11 individuals (7 M, 4 F), Other: 60 individuals (24 M, 36 F), Data was collected from 2018 to 2021
142	Neuro[254]	42	43	–	–	–	–	Five datasets combined into two datasets: Comb_Tappy and Comb_Neuro
	PhysioNet[250,251]	159	51	–	–	–	–	
	Timisoara[259,260]	–	80	–	–	–	–	
	BB-MAS[261]	–	117	–	–	–	–	
	Buafo[262]	–	148	–	–	–	–	
143	263	93	73	34	59	33	40	–

In this table, we detailed all the datasets in papers and compared participant demographics (number, gender, and health status: #PD => Parkinson's Disease, #HC => Number of Healthy Control Participants, #PDF => Number of Parkinson's disease Female Participants, #PDM => Number of Parkinson's disease Male Participants, #HCF => Number of Healthy Control Female Participants, #HCM => Number of Healthy Control Male Participants)

attained. Secondly, it has become clear that CNN is one of the most successful methods for PD diagnosis.

- Weaknesses: However, this approach has three major points: it is too expensive due to the need for expertise and clinical tests such as MRI, the techniques employed in this approach are primarily invasive, and it needs high computational resources.
- Opportunities: Resolving these conflicts may increase the reliability of medical-imaging-based approaches. Examining less invasive or non-invasive techniques, making the applications broader. Also, more cost-effective approaches, such as fMRI and DAT scans, can enhance accessibility. Another way could be to combine imaging with genetics or other data to increase the accuracy of diagnosis. Also, using ML as an interpreter for data instead of data analysts may reduce the costs.
- Threats: Nevertheless, bias may rise if we do not consider diversity in gathering imaging data, so researchers should consider global populations in their test samples. Another challenge is logistical and ethical in gathering such information, which also needs to be addressed.

Movement data

Movement datasets are divided into three categories: gait, tremor, and movement[111]. Such information is mainly utilized to determine the severity of the condition. First, monitoring gait data is significantly invaluable because it can be gathered through wearable devices, which are primarily low-cost[112]. Also, continuous data can help to find appropriate therapy for the patient and examine its effectiveness. However, it cannot be a good approach for diagnosing the disease. Second, tremors are uncontrollable body shaking and are mainly misdiagnosed; suitable data by using ML can reduce diagnostic errors[113]. The movement datasets utilized in the examined papers are detailed in Table 12, which includes the names and references of the sources, the categories of movement data, descriptions of the features, and information regarding the age and gender of the participants. We

Table 13 | A comparison of movement data papers

Article	Main idea	Tools	Applied algorithms	Advantages	Disadvantages
8	PD diagnosis and severity evaluation using MCSVM	• MATLAB	• MCSVM • SVM kernel functions (linear, polynomial, cubic, and quadratic)	• High accuracy • High specificity • High sensitivity	• Lack of non-motor symptom evaluation • Not analyzing metrics such as recall and cost
4	DL and neuro-fuzzy model for PD detection	• MATLAB	• DBN • KNN • ANFIS • EM • PCA	• Low time complexity • High accuracy	• Not evaluating parameters such as sensitivity and recall
114	FOG prediction in PD using ResNeXt	• Python (Pytorch)	• ResNeXt • SMOTE • Adam optimizes	• High accuracy • High sensitivity • high specificity	• Low scalability
115	Analyze real-world gait tests in PD patients	• Not mentioned	• SDTW	• High F1-score • High recall • High precision	• Lack of analysis of gait trials
116	Analysis of sEMG signals and hybrid DTL for diagnosing PD	• MATLAB	• CNN • SVM • SGD • Propagation (RMSprop)	• High accuracy • High specificity • High sensitivity	• Low scalability
117	Assessing PD severity via the EnKNN approach	• Python	• EnKNN	• High accuracy	• Low scalability
118	Real-time FOG detection in PD using CNN	• MATLAB • Python • Keras-flops	• CNN	• Low computational complexity • Low processing time • High performance • Reducing memory usage • High AUC • High predicting ability	• No integration with a standalone device for home environment • Utilization of raw input data
119	Early PD detection using wearable sensors and ML	• Not mentioned	• LightGBM • RF	• High precision • High F1-micro • High AUC	• Lack of additional data types such as video/images
120	CNN for PD and AD classification	• R programming • Python	• Multi-layer CNN • LDA • MLP	• High accuracy	• Lack of evaluation on imbalanced datasets • Low scalability
113	PD symptom detection through video analysis	• Python (Scikit-learn) • OpenCV	• LR • XGBoost • RF • SVM • Gaussian process classifier	• High F1-score	• Lack of consideration of datasets with varied disease • Not examining the data with other models, such as CNN
121	Handwriting analysis via CNN for PD diagnosis	• Not mentioned	• CNN	• High accuracy	• Not investigating alternative architectures • Low scalability
122	Movement management in PD patients using DL	• Python	• DNN	• High accuracy	• Lack of minimizing the model effect
123	DL method to distinguish MS from PD via gait	• Python (PyTorch)	• CNN • RNN • MS-ResNet	• High accuracy • High AUC	• Low scalability
124	Neural network–based early PD detection via gait data	• Not mentioned	• Neural network	• High accuracy	• Training model with a limited number of patients • Low scalability
125	Motor symptom–based PD detection using ML	• MATLAB • Python	• Lasso • LR • RF • DT • SVM • KNN • XGBoost • Linear Regression	• High accuracy • High AUC (ROC)	• Low scalability
126	PD severity assessment via DL on movement data	• Python	• CNN-BGRU	• High accuracy	• Not analyzing parameters such as AUC (ROC) and specificity • Lack of clinical validation
127	ML-based detection of PD using upper limb motion	• MATLAB	• DT • RF • KNN • SVM • NB	• High accuracy • High sensitivity • High specificity • High AUC (ROC)	• Low scalability

Table 13 (continued) | A comparison of movement data papers

Article	Main idea	Tools	Applied algorithms	Advantages	Disadvantages
128	Analyzing copied figures with CNN to detect PD	• Python (NumPy, Pandas)	• CNN	• High accuracy • High specificity	• Low scalability
129	Analyzing VGRF gait data via ML to detect PD	• Not mentioned	• SVM • KNN • NB • DT • ELA	• High accuracy	• Low scalability • Neglecting all gait signals but VGRF
130	Learning architecture for PD diagnosis	• Python • CUDA • cuDNN	• CNN • ARR • XGBoost • SMOTE	• High accuracy • Low training time	• Low scalability
131	Gait-based PD detection and stages with ML models	• Python (NumPy, Matplotlib, Scikit-learn, Pandas, Seaborn)	• NB • SVM • DT • MLP • LR • RF • SMOTE	• High accuracy • High precision • High AUC (ROC)	• Low scalability • Not evaluating other motor and non-motor symptoms • Lack of tremor analysis in gait classification
132	Extracting diagnostic features from spiral drawings using ML	• Python (Scikit-learn)	• LR • SVM • KNN • DT • RF • AdaBoost • SVM-RFE	• High predicting ability • High specificity • High accuracy • High sensitivity	• Low scalability • Lack of symptom severity assessment • Not evaluating other tasks related to handwriting and drawing
133	Unsupervised uTUG-based gait assessment for PD using ML	• GroupKFold • GridSearchCV • Python (Scikit-learn)	• NB • SVM • RF	• High accuracy • High recall • High sensitivity • High F1-score • Not requiring manual annotation	• Not evaluating adverse drug reactions • Lack of additional sensor data • Incomplete evaluation of at-home completion time
134	Balanced ensemble learning for PD diagnosis utilizing KD dataset	• R programming	• XGBoost • KNN • NB • LSTM • MLP • SVM	• High sensitivity • High specificity • Ease of integration with conventional desktops • High robustness • High AUC	• Not investigating diseases that affect typing quality • Not assessing the impact of factors such as age, emotional tension, and keyboard layout experience on typing • No evaluation of wearable and mobile sensors for improved data collection • Not evaluating the severe level of the disease
135	Ensemble DT and gait features for PD detection	• Not mentioned	• RF • GB • DT	• High accuracy • High F1-score • High sensitivity • High specificity • High precision	• Low scalability
136	ML-based PD diagnosis using gait and movement data from wearable sensors	• MATLAB	• Random under-sampling boosting • Neighborhood component analysis • mRMR • RF • DT	• High sensitivity • High specificity • High AUC	• Low scalability • Lack of generalization assessment to other motor disorders • Lack of evaluation of motor fluctuations
137	Kinematic handwriting features and ML for PD diagnosis	• Python	• RNN • LSTM • BLSTM • Adaboos • BRF • SVM • LDA • PCA • Bayesian optimization algorithm • Adam optimizes	• High accuracy • High precision • High recall	• Not analyzing diverse handwriting datasets • Low scalability
138	PD detection by using handwriting and neural network	• Python (Tensorflow)	• NB • RF • DT • LR • KNN • GBDT • CNN • BLSTM • LSTM	• High accuracy	• Not expanding image datasets sufficiently

Table 13 (continued) | A comparison of movement data papers

Article	Main idea	Tools	Applied algorithms	Advantages	Disadvantages
139	ML-based PD diagnosis by analyzing exercise effectiveness	• Python	• PCA • ICA • MDS • RF • LR • NB • Boosted trees • KNN • Stacked ensemble model	• High AUC (ROC) • Low hospitalization cost • Reduce diagnosis time	• Small and imbalanced dataset • Sensors with limited battery life • Wireless connection problems • Lack of comfort for people with advanced PD to wear sensors
140	PD detection via CNN based on daily gait patterns	• Python	• CNN	• High accuracy • High AUC	• Not distinguishing the level of PD • Not evaluating other forms of movement except walking
112	Monitoring PD motor symptoms using ML methods	• Python • MobilityLab software	• LR • RF • PCA	• High RMSE	• Low scalability
141	PD detection via ML approaches	• Python	• SVM • FNN • CatBoost • BOSS • XceptionTime	• High accuracy	• Small sample size for validation • One-Time clinical assessment
142	PD diagnosis using keystroke dynamics data	• Not mentioned	• MFDFA • CNN	• High accuracy • High sensitivity • High specificity	• Low scalability
143	LSTM-based classification of PD using walking data	• Not mentioned	• LSTM • MCOA	• High accuracy	• Low scalability

This table includes key concepts (main ideas), the utilized tools, the applied algorithms, advantages, and disadvantages.

tabulate the main ideas, applied algorithms, tools, advantages, and disadvantages in Table 13. Table 14 also includes assessment parameters for the reviewed studies, such as accuracy, sensitivity, specificity, and F1-score. In Section "Review of movement data-based approaches", movement data-based approaches are reviewed. Additionally, Section "Qualitative analysis of movement data-based approaches" presents a qualitative analysis of movement data-based approaches, including their strengths, weaknesses, opportunities, and threats.

Review of movement data-based approaches. Vidya and P[8] offered a method aimed at detecting and classifying the severity of PD. The proposed approach uses a multi-class support vector machine (MCSVM) along with gait data analysis. This method showed high accuracy and sensitivity. Plus, Nilashi et al.[4] suggested a method that combined the deep belief network (DBN) and adaptive neuro-fuzzy inference system (ANFIS) to improve the accuracy of forecasting the unified PD rating scale (UPDRS) and diagnosing PD. Additionally, the results demonstrated a significant reduction in time complexity while enhancing prediction accuracy in the proposed approach, thereby supporting the early diagnosis of PD through precise and efficient UPDRS prediction. Also, Hua Sun et al.[114] presented an approach for forecasting FOG in PD by integrating deep features acquired via the ResNeXt network with manually selected gait features. The results demonstrated that the approach exhibited high specificity, sensitivity, and accuracy.

Ullrich et al.[115] developed an algorithmic approach utilizing inertial measurement units (IMU) to evaluate and diagnose Parkinson's patients' gait tests. This study aimed to reduce patient interaction with the recording system and reduce the amount of manual data annotation performed by researchers. In addition, this algorithm achieved a high F1-score, recall, and precision. Additionally, Rezaee et al.[116] proposed an ML-based approach, incorporating deep transfer learning (DTL) using electromyographic signals for PD diagnosis. Moreover, the presented method achieved high accuracy and sensitivity.

Zhao et al.[117] introduced a method to detect the severity level of PD through gait data and the ensemble K-nearest neighbor (EnKNN) algorithm. The suggested method effectively handles the imbalanced data distribution from Parkinson's patients. Additionally, the proposed EnKNN exhibited favorable performance and accuracy compared to other methods, as demonstrated by the results. Besides, Borzì et al.[118] used a multi-headed CNN for detecting freezing of gait (FOG) in PD. The method utilized inertial sensor data. Plus, this study fastened processing times, minimized memory usage, and high accuracy. Likewise, Shcherbak et al.[119] introduced an approach for diagnosing early-stage PD (phases 1 and 2) using wearable sensors, movement data, and ML. Overall, distinguishing between healthy individuals and stage 2 patients resulted in improved outcomes, including higher F1-micro scores and precision.

Pedrero-Sánchez et al.[120] suggested a multi-branch CNN-based method to classify PD and Alzheimer's patients from healthy subjects using functional mobility test data. The results showed that the proposed method provided higher accuracy than parameter-based methods. Moreover, Kovalenko et al.[113] presented an approach to detect PD in essential tremors utilizing ML techniques by analyzing video data. Further, this approach obtained a high F1-score.

Gazda et al.[121] presented a model that diagnosed PD by evaluating people's handwriting using CNN. Experimental results confirmed the high accuracy of the mentioned model. However, the authors did not investigate their approach with larger datasets or employ any other network architecture. Similarly, Ibrahim et al.[122] offered a neural network model capable of predicting PD motions across multiple stages and minimizing delay using tremor data. Based on the evaluation results, the model was highly accurate. Similarly, Kaur et al.[123] utilized the data on walking patterns to classify Parkinson's and multiple sclerosis patients. Additionally, they evaluated different DL and ML methods, while CNN showed higher accuracy.

Lin et al.[124] proposed a neural network-based method that utilizes movement data for the early detection of PD and classification of its severity. This model demonstrated high accuracy. Also, Exley et al.[125] presented a method for measuring PD symptoms using ML and quiet standing data. In this method, the AUC (ROC) was high. Furthermore, Li et al.[126] presented a DL-based framework to identify PD in rats with brain abnormalities. The authors collected 3D movement data and employed DL for classification. The evaluated results indicated the high accuracy of the model.

Table 14 | Movement data evaluation metrics

Article	Accuracy	Sensitivity	Specificity	F1-score	Precision	AUC (ROC)	NPV	FPR	MCC	Pearson's correlation coefficients	RMSE	SNR	MISE	Standard deviation	G-mean	EER	F1-micro	PR (AUC)	MAE	RAE	Kappa coefficient
8	+	+	+	+	+								+								
4	+										+										
114	+	+	+											+			+				
115		+		+	+					+									+	+	
116	+	+	+	+	+	+		+				+									
117	+	+		+	+										+						
118	+	+	+	+		+									+	+					
119			+		+	+											+	+			
120	+																				
113	+	+		+	+																
121	+																				
122	+										+										
123	+	+		+	+	+															
124	+	+	+	+	+																
125	+	+	+			+					+										
126	+	+		+	+						+										
127	+	+	+			+															
128	+		+	+																	
129	+	+	+	+	+		+		+												
130	+	+		+	+									+							
131	+	+		+	+	+															
132	+	+	+		+									+							
133		+		+	+									+							
134	+	+	+			+															
135	+	+	+	+	+																
136		+	+			+															
137	+	+			+	+		+													
138	+																				
139	+			+		+								+							
140	+					+															
112											+										
141	+	+		+	+																
142	+	+	+	+											+						
143	+	+		+	+																+

Plus (+) means the metric that has been evaluated in the paper, and blank cells mean the metric has not been evaluated in the paper.

Cesarelli et al.[127] investigated the predictive capability of upper limb features using ML algorithms to differentiate between normal people and those with PD. This study achieved high accuracy. Besides, Alissa et al.[128] presented a CNN-based system for detecting PD based on drawing data. As a result of the proposed method, high accuracy was demonstrated. Additionally, Wang et al.[129] developed an ML-based hybrid signal-processing approach for the detection and severity estimation of PD. Five different ML methods were evaluated in this study. The accuracy evaluation results showed the high performance of the SVM model.

Ma et al.[130] provided an interpretable architecture using DL models and stepping-shoe pressure sensors for the early detection of PD. Results showed that the proposed architecture achieved high identification rates and accuracy. Additionally, Ferreira et al.[131] developed an approach aimed at improving PD diagnosis and stage identification. This study evaluated five ML methods and analyzed the spatial-temporal characteristics of gait. The presented approach showed acceptable accuracy and AUC (ROC).

Valla et al.[132] extracted tremor-related features from the archimedean spiral drawing test to improve PD diagnosis through ML methods. The authors employed filter methods like Fisher's score and wrapper methods like RFE for relevant feature selection. The method demonstrated high accuracy, specificity, and sensitivity in the results. Furthermore, Tavares et al.[133] introduced an unsupervised algorithmic pipeline called uTUG, which employed IMUs for motor assessment and PD diagnosis. Their method significantly enhanced precision, sensitivity, and F1-score in detecting timed up-and-go (TUG) assessments using ML algorithms.

Soumen Roy et al.[134] introduced an ML model that utilized keystroke dynamics (KD) features during text typing on a conventional keyboard. The model employed a bootstrap-based homogeneous ensemble classification architecture along with ML techniques to detect PD in its De-novo and early stages. Among the ML methods examined, XGBOOST exhibited superior performance. Moreover, the proposed model exhibited notable sensitivity and specificity. Also, Huan Zhao et al.[135] proposed a method to diagnose PD by analyzing distinct gait pattern characteristics. They identified features like asymmetry index, mean, and coefficient variance from Parkinson's patients' gait patterns. Also, ensemble DT made an improvement in the diagnostic process and accuracy. In the same light, Mirelman et al.[136] introduced a method that uses the data from wearable sensors and natural language approaches to identify the gait and mobility criteria of different stages of PD. This study aimed to specify optimal sensor locations for each disease stage, and the results showed high sensitivity, specificity, and AUC for the proposed method.

Kumar et al.[137] presented this study to identify the handwriting dataset's most effective task and establish a reliable diagnosis method. The authors employed two variants of RNN, a DL technique based on bi-directional long short-term memory (BLSTM) and LSTM. In addition, the kinematic properties obtained through various ML approaches were examined. The proposed procedure was highly accurate, based on the outcomes. Besides, Zhao et al.[138] introduced a neural network approach to detect PD by classifying individuals' handwriting. The comparison results reveal that the presented model performs better than its counterparts. Nevertheless, the size of the data could be expanded to improve.

Talitckii et al.[139] presented a method for identifying the most effective exercises for diagnosing PD using ML approaches and wearable sensors. Three of the 15 common exercises with the highest discrimination power achieved a high AUC (ROC) score. In addition, this method could improve PD diagnosis, reduce hospitalization costs, and reduce the time required for diagnosis. Moreover, Chen et al.[140] implemented an optimizable model using CNN architecture to detect PD accurately from daily walking and adapt according to the most indicative spatiotemporal motor characteristics. The data was gathered from 100 subjects while walking 10 meters, monitored by five sensors attached to their bodies. Moreover, the results indicated high accuracy and AUC for the model.

Sotirakis et al.[112] offered an approach to monitor the progression of motor symptoms in Parkinson's patients using ML methods. Also, the data used was gathered from wearable sensors. This study obtained an acceptable RMSE. Besides, Varghese et al.[141] investigated different ML methods in order to diagnose PD and differential diagnosis. Furthermore, they presented a movement dataset from wearable technologies to help develop accurate diagnostic tools. This study achieved acceptable accuracy. Also, Yang et al.[142] introduced a method based on multi-level ensemble learning using keystroke dynamics for PD diagnosis. The presented method showed high accuracy. Cuk et al.[143] offered an approach for early detection of PD using LSTM neural networks and gait data. The results showed that this approach was highly accurate.

Qualitative analysis of movement data-based approaches. In this section, we critically analyzed the selected paper on movement-based approaches. Also, the classification of strengths, weaknesses, opportunities, and threats has been detailed:

- Strengths: The qualitative assessment of the selected papers in movement-based approaches led us to conclude that continuous patient monitoring aids in real-time assessment of PD and directly captures hallmark symptoms. The key advantage of movement-based approaches is online data gathering, which is done by applying wearable sensors.
- Weaknesses: On the contrary, although highly accurate sensors may improve understanding of PD, they are expensive and may not be available to everyone. Capturing data in controlled environments may increase bias and cause other symptoms to be overlooked.
- Opportunities: One of the key areas researchers should focus on is developing more affordable and high-tech wearable sensors to ensure their availability for everyone. Moreover, combining these data with vocal and biomarker data will help to get reliable and more accurate results. These real-time datasets should also be used extensively for PD progression monitoring and patient management.
- Threats: However, there are concerns about the privacy of continually gathering PD patients and data sharing. Also, low-quality sensors directly impact the mode performance, and the result may be undermined.

Biomarkers

Parkinson's latest reports suggest that the combination of genetic and abnormal brain activity can contribute to the development of the disease[6]. The analysis of genes and identification of key genetic and brain malfunctions accelerates PD detection, facilitating early diagnosis in potential patients[144]. This approach may enable the development of targeted therapies that address the specific genetic and brain-related factors associated with PD. Table 15 provides a summary of the biomarker datasets utilized in the reviewed articles, detailing the types of biomarkers, extracted features, demographic information, analytical methods, and sampling types. Table 16 presents the main ideas of the papers, applied algorithms, tools, advantages, and disadvantages. Additionally, Table 17 compares several evaluation metrics, such as accuracy, specificity, and sensitivity.

Section "Review of biomarker-based approaches" reviews biomarker-based approaches. In the end, a qualitative analysis of approaches, including strengths, weaknesses, opportunities, and threats, is performed in Section "Qualitative analysis of biomarker-based approaches".

Review of biomarker-based approaches. Arora et al.[1] compared different ML techniques to detect PD through amino acid composition and hydrophobicity. In the proposed approach, the recall and F1-score were high. Likewise, Xie et al.[6] proposed a model that combined ML techniques, such as RF models and ANN, to diagnose PD. They also investigated the role of immune cell infiltration in PD. Moreover, Göker et al.[145] developed a DL-based method for detecting PD in its early stages using EEG signals. In order to create the automatic model, welch spectral analysis was combined with BLSTM. Also, based on the outcomes, this method attained high scores for the evaluation criteria like precision, specificity, MCC, accuracy, sensitivity, and F1-score.

Table 15 | Biomarker datasets

Article	Repository/ Source	#PD	#HC	#PDF	#PDM	#HCF	#HCM	Information
1	GEO databases[264,265]	–	–	–	–	–	–	Protein sequences from NCBI & UniProt (FASTA file format), removed duplicates and incomplete sequences, Dataset: 640 PD-related and 1010 non-PD sequences
6	GEO databases[264]	35	28	–	–	–	–	Training sets: GSE20163, GSE20164, GSE42966, validation: GSE26927
145	266	14	14	–	–	–	–	16 F and 12 M
146	UC San Diego dataset[267,268]	15	16	8	7	9	7	HC: average age (63.5 ± 9.6) PD: average age (63.2 ± 8.2)
147	Clinical	20	21	10	10	10	11	HC: average age (67.5 ± 6.4), PD: average age (67.6 ± 7.0), University of British Columbia
148	269,270	24	24	–	–	–	–	HC: average age (69.33 ± 9.78) PD: average age (69.75 ± 8.91)
149	Clinical	104	11	–	–	–	–	PD: average age (59.43 ± 12.15) HC: average age (57.26 ± 9.15)
144	271–276	–	–	–	–	–	–	Gene expression data from GEO: GSE18838, GSE57475, GSE72267, GSE99039, and GSE6613, 406 PD samples, 336 HC samples
150	Clinical	39	40	17	22	12	28	HC: average age (59.00 ± 4.54) PD: average age (61.31 ± 6.01)
151	277	25	25	9	16	9	16	PD: average age (69.98 ± 8.73) HC: average age (69.32 ± 9.58) Cognitive Rhythms Lab (UNM), collecting data in 2015
	278,279	20	20	11	9	12	8	PD: average age (69.80 ± 7.60) HC: average age (67.80 ± 6.35) Information was gathered at the University of Turku in Finland
152	280	24	24	–	–	–	–	–
153	Clinical	19	–	–	–	–	–	–
154	PPMI[220]	490	197	–	–	–	–	HC: average age 61.3 PD: average age 62
	Clinical	59	31	21	38	17	20	From 2015 to 2018
155	Clinical	187	125	76	111	67	58	–
156	Clinical	31	13	16	15	6	7	–
157	SEED-IV[281]	–	–	–	–	–	–	SEED-IV dataset with 15 subjects, Evaluated using film clip stimuli, Emotions: happy, neutral, fearful and sad
	AMIGOS[282]	–	–	–	–	–	–	AMIGOS dataset: 33 subjects, auditory and visual stimuli Two trial types: short and long videos
	283	20	20	11	9	10	10	Multimodal stimuli: images, audio, video Mean age: 58.7
158	UC San Diego dataset[267,284]	16	10	8	8	9	1	PD: average age 58.7 HC: average age 63.5 ± 9.6
159	Clinical	29	–	20	9	–	–	Female average age: 62 Male average age: 63
160	285	23	26	7	16	13	13	10 patients with ICD (8 M and 2 F)
161	GEO databases[264]	20	20	10	10	12	8	GEO datasets: GSE8397, GSE20292, GSE20163, GSE20164, and GSE49036, Average age (68.2 ± 7.2) Average age (66.0 ± 12.8)
162	Clinical	65	65	9	56	9	56	Parkinson's patients were selected from Juntendo University Hospital, Tokyo, Japan, The first cohort included de novo PD patients (HC: average age (62.2 ± 11.8) PD: average age (61.7 ± 11.4) The second cohort included male PD patients with and without medication (HC: average age (66.8 ± 9.08) PD: average age of (64.2 ± 10.6)

Table 15 (continued) | Biomarker datasets

Article	Repository/ Source	#PD	#HC	#PDF	#PDM	#HCF	#HCM	Information
163	UC San Diego dataset[267,286,287]	15	16	8	7	9	7	Collected from the University of San Diego, EEG was recorded in the resting state
	270	27	27	17	10	17	10	Collected from the University of New Mexico (UNM)
164	PPMI[220]	294	154	99	195	58	96	PD: average rage (61 ± 9.7) HC: average rage (60.3 ± 11)
	PDEP[288]	263	115	112	151	64	51	PD: average rage (64.3 ± 8.6) HC: average rage (63.6 ± 9.5)
165	National Health Insurance Service-Health Screening (NHIS-HEALS) database[289]	1102	1102	505	597	492	610	Adults aged 40 and older, data includes lab and anthropometric measures, sex, lifestyle, socioeconomic status
166	PPMI[220]	423	–	–	–	–	–	The dataset consisted of de novo PD patients
167	Loyola University Chicago (LUC), Clinical	29	165	11	18	64	101	ECG data from individuals aged 26–89 years, MLH dataset: collected 2015–2020, LUC dataset: collected 2014–2020
	University of Tennessee-Methodist Le Bonheur Healthcare (MLH), Clinical	131	1058	54	77	496	562	
168	PPMI[220]	697	–	–	–	–	–	Patients assessed before dyskinesia onset

In this table, we detailed all the datasets in papers and compared participant demographics (number, gender, and health status: #PD => Parkinson's Disease, #HC => Number of Healthy Control Participants, #PDF => Number of Parkinson's disease Female Participants, #PDM => Number of Parkinson's disease Male Participants, #HCF => Number of Healthy Control Female Participants, #HCM => Number of Healthy Control Male Participants).

Table 16 | A comparison of biomarkers papers

Article	Main idea	Tools	Applied algorithms	Advantages	Disadvantages
1	Ensemble algorithm for PD diagnosis using protein sequences	• Python	• DT • NB • SVM • KNN • LR • RF • AdaBoost • GB	• High accuracy	• Not analyzing parameters such as specificity and MCC
6	PD diagnostic model using ML and immune infiltration data	• R programming	• ANN • RF • RMA	• Identifying the crucial genes • High AUC (ROC)	• Low scalability
145	DL-based PD diagnosis using EEG data	• Not mentioned	• BLSTM	• High specificity • High precision • High F1-score • High accuracy • High sensitivity • High MCC	• Limited diversity in age groups and races
146	PD diagnosis using resting-state EEG signals	• Python	• KNN • RF • ET • LightGBM • XGBoost • QDA	• High accuracy	• Low scalability • Manual removal of artifacts • Limited spatial resolution
147	CRNN-based PD diagnosis using EEG signals	• Python	• GRU • CNN • CRNN	• High accuracy • High recall • High precision • Automatic feature learning without additional processing of data	• Low scalability • Limited interpretability
148	PD detection via ASGCNN on EEG data	• Not mentioned	• FASTER algorithm • ASGCNN • LSTM • Adam optimizer • ICA	• High accuracy • High precision • High recall • High F1-score • High Kappa	• Low scalability
149	Metabolomics-based PD diagnosis using ML	• MetDNA • Compound discoverer software • QIAGEN • IPA	• PLS-DA • RF • XGBoost • LASSO • Ridge regression • QC-RLSC	• High accuracy • High AUC (ROC)	• Low scalability
144	PD diagnosis using gene expression data and ML	• R programming	• LASSO • Ridge regression • NB • RF • KNN	• High accuracy • Integrating datasets from a variety of sources	• Lack of analysis of other types of PD • Not integrating additional datasets

Table 16 (continued) | A comparison of biomarkers papers

Article	Main idea	Tools	Applied algorithms	Advantages	Disadvantages
			• DT • SVM • LR		
150	EEG-driven PD classification using ML	• MATLAB	• FastICA algorithm • CNN • SVM	• High accuracy • Short training time	• Lack of subgroup evaluation based on PD clinical stages • Low scalability • Lack of evaluation of advanced feature selection methods
151	DCNN-based model to study patho-electrophysiology via EEG	• MATLAB • FieldTrip	• DCNN • Grad-CAM	• Generalizability	• Limited to small dataset performance • Low accuracy
152	PD diagnosis using Hjorth features and ML approaches	• MATLAB • EEGLab	• SVM • KNN • RF	• High accuracy • High TPR • High AUC	• Low scalability • Limited comparison with other feature extraction techniques, such as the coefficients of the fast Fourier transform
153	CNN-GA-KNN model for improving STN localization using LFP	• MATLAB	• CNN • ResNet18 • VGG16 • KNN • GA	• High accuracy • High sensitivity • High AUC (ROC)	• Low scalability
154	PD detection using ML and non-motor symptom data	• Python • Weka	• AdaBoost • Bootstrap aggregating • DT • KNN • MLP • NB • RF • RIPPER • SVM	• High accuracy • High F1-score	• Limited to non-motor information • Small sample size • Bias in data collection
155	DL model for PD detection using metabolic fingerprint	• Not mentioned	• LASSO • LR • XGBoost • SVM • RF • AdaBoost	• High AUC • High specificity • High sensitivity	• Low scalability
156	EEG analysis methods for PD diagnosis and monitoring	• MATLAB • Python • Wavelet toolbox	• DRSN • TQWT • WPT	• High performance in the analysis of non-stationary signals • High accuracy • High F1-score • High recall • High specificity • High precision	• Decreased accuracy with more categories • No assessment of short- and long-term memory networks • Low scalability • Lack of standardized EEG data
157	DL-driven architecture for identifying emotional states in PD	• MATLAB	• CRNN • 1D-CNN • LSTM • ELM	• High accuracy	• Low scalability
158	ML/DL-based model for PD detection using EEG	• Not mentioned	• 1D-PDCovNN Model • MLA • XGBoost • ICA	• High accuracy	• Lack of detecting PD by the model in similar signals • Low scalability
159	EEG-driven PD diagnosis using DNN	• Not mentioned	• CNN • DNN • STMIM • Grad-CAM	• High accuracy	• Not considering different subtypes of PD • Limited clinical data were taken into account
160	ML approach for ICD detection in PD using EEG data	• MATLAB • EEGLab	• SVM • SVR • mRMR	• Low cost	• Low scalability
161	Biomarker and therapeutic gene identification for PD diagnosis	• R programming	• RF • SVM-RFE • PCA • LASSO	• High accuracy	• Low scalability • No assessment of gene-level mechanistic function
162	ML-based PD diagnosis using sebum RNA profiles	• Python • R programming	• ERT • PCA	• High AUC (ROC)	• Low scalability
163	PD diagnosis using ML methods and DWT	• MATLAB	• DWT • LR • LDA • KNN	• High accuracy • Less memory usage • Low execution time	• Low scalability

Table 16 (continued) | A comparison of biomarkers papers

Article	Main idea	Tools	Applied algorithms	Advantages	Disadvantages
			• SVM • RF	• Fewer parameters • Low complexity	
164	ML-driven prediction and detection of PD subtypes	• Python	• Non-negative matrix factorization • GMM • RF • LightGBM • XGBoost	• High AUC (ROC)	• Low scalability
165	ML-based PD prediction using NHIS screening data	• Not mentioned	• LR • RF • Neural networks • GBM • DT • NB • XGBoost	• High AUC (ROC) • High accuracy • Cost-effective	• Restriction of data to a particular nation
166	ML-based prediction of cognitive outcomes in PD patients	• R programming • PRSice-2 software	• ElasticNet • RF • SVM • Conditional inference forest	• High sensitivity	• Low scalability
167	AI-based PD detection using ECG data	• Python • Epic Software	• LightGBM • CNN	• High specificity	• Small sample size for validation
168	Dyskinesia prediction in PD via ML	• Not mentioned	• RF • CART • Adaboost • DT • LR • MLP • SVM	• High AUC (ROC)	• Not analyzing parameters such as F1-score and sensitivity

This table Includes key concepts (main ideas), the utilized tools, the applied algorithms, advantages, and disadvantages.

Lal et al.[146] developed and evaluated an architectural pipeline for PD diagnosis using resting-state EEG. This paper suggested utilizing ML models with a focus on the KNN classifier. In addition, their method achieved high accuracy. Also, Soojin Lee et al.[147] presented a model based on a CRNN composed of RNN, CNN, and gated recurrent units (GRUs) utilizing resting-state EEG. Additionally, CRNN displayed high recall, precision, and accuracy.

Chang et al.[148] developed an attention-based sparse graph convolutional neural network (ASGCNN) approach for early PD detection using electroencephalography (EEG) signals. This study uncovered statistically significant disparities between patients with PD and healthy individuals. The proposed method obtained high levels of recall, accuracy, precision, and F1-score, according to the findings of the study. Additionally, Wang et al.[149] suggested a urine sample-based method for detecting PD to identify specific metabolites as biomarkers for PD and a predictive model use ensemble ML techniques. This study identified eight metabolites that could distinguish between PD patients and healthy people using a combination of metabolomics and ML assessments. Based on the results, this method had a high AUC and accuracy.

Bhandari et al.[144] presented an approach for diagnosing PD using blood-based gene data. This study employed ridge regression and LASSO for feature selection, alongside various ML methods—particularly LR and SVM—for data classification. In addition, the SHAP method was used to identify the essential genes responsible for diagnosing PD. According to the results, the proposed method provided high accuracy. Yang and Huang[150] conducted a study to evaluate the efficacy of CNN and SVM in categorizing individuals with PD based on resting-state EEG data. The results demonstrated that the CNN approach outperformed SVM by effectively identifying important features, shortening the training time, and achieving higher levels of accuracy.

Shabanpour et al.[151] developed a multivariate and data-driven model utilizing DCNN to analyze EEG data and identify spatial oscillatory patterns associated with PD. The model was intended to improve understanding of the brain physiology in PD while creating clinically interpretable topographical maps. In the same vein, Oliveira Coelho et al.[152] proposed a diagnostic model for PD using Hjorth features derived from EEG signals. Patients with PD exposed to auditory stimuli had their data analyzed using SVM, KNN, and RF. Results demonstrated the model's high accuracy in distinguishing Parkinson's patients from healthy individuals, particularly when SVM was utilized.

Hosny et al.[153] suggested a DL model founded on CNN-GA-KNN that improved the localization of the subthalamic nucleus (STN) using local field potentials (LFP) in patients with PD. The suggested model used a CNN to extract features and the GA to select features. In addition, KNN was used for classification. Furthermore, the results showed that this model was highly accurate. Additionally, Martinez-Eguiluz et al.[154] developed ML models for the early diagnosis of PD using non-motor symptoms, such as autonomic dysfunction and depression. They evaluated the models using two databases, namely PPMI and Biocruces. SVM and MLP demonstrated the most promising results among all the algorithms.

Xu et al.[155] proposed a DL-based system that investigated the metabolites and small molecules in saliva to detect PD at early stages (Hoehn-Yahr stage 1-2.5). The mentioned algorithm used 312 samples for validation purposes, and the results demonstrated its performance in terms of AUC, sensitivity, and specificity. Also, Zhang et al.[156] introduced two methods for classifying clinical sleep EEG data, namely wavelet packet transform with deep residual shrinkage network (WPT-DRSN) and tunable Q-factor wavelet transform with deep residual shrinkage network (TQWT-DRSN). This model integrated time-frequency analysis and DL, demonstrating promising outcomes in classifying non-stationary signals. In addition, REM sleep behavior disorder (RBD) was investigated in conjunction with PD, and the proposed model demonstrated high accuracy in early detection of PD and disease tracking.

Dar et al.[157] introduced an architecture called 1D-CRNN-ELM, which combined a CRNN and an ELM to detect six fundamental emotions in individuals with PD. The suggested architecture achieved high accuracy in classifying emotions and showed its effectiveness in EEG-based signals emotion recognition. Also, Nour et al.[158] suggested an approach that

Table 17 | Biomarker's evaluation metrics

Article	Accuracy	Sensitivity	Specificity	F1-scores	Precision	AUC (ROC)	FPR	MCC	Standard deviation	Kappa coefficient
1	+	+		+	+	+				
6		+	+			+				
145	+	+	+	+	+			+		
146	+	+		+	+	+	+			
147	+	+		+	+	+				
148	+	+		+	+					+
149	+					+				
144	+	+			+	+	+			
150	+									
151	+					+				
152	+	+	+			+				
153	+	+	+			+			+	
154	+	+		+	+					
155	+	+	+	+	+	+				
156	+	+	+	+	+	+				+
157	+	+	+	+	+					
158	+	+		+	+	+				+
159	+					+				
160	+	+	+							
161	+					+				
162	+	+	+	+		+			+	
163	+	+	+	+		+				
164						+				
165	+					+				
166	+	+	+			+		+		
167	+	+	+			+				
168	+	+	+			+	+			

Plus (+) means the metric that has been evaluated in the paper, and blank cells mean the metric has not been evaluated in the paper.

classified PD using ensemble learning through EEG signals. The mentioned models were investigated in terms of several evaluation metrics, such as kappa score, ROC curve, and accuracy, demonstrating their performance. Moreover, Chu et al.[159] offered an advanced framework for EEG microstates utilizing DNN to identify patients with PD in the initial stages. Also, brain regions were investigated for any probable relationship with PD. The assessments depicted that the model could increase accuracy.

Lin et al.[160] developed a method for assessing and estimating the severity of impulse control disorder (ICD) comorbidity in patients with PD using ML. Furthermore, EEG measurements were obtained utilizing an inexpensive headset, which enabled the device to be implemented in routine environments. Moreover, Wu et al.[161] investigated the potential biomarkers and therapeutic target genes for PD and confirmed the findings using experimental approaches. Also, this study used different ML algorithms, including LASSO and RF. In another research, Uehara et al.[162] introduced an approach for diagnosing PD using sebum RNA profiles. In addition, they analyzed the profiles using ML methods. This approach showed high AUC (ROC). Additionally, Aljalal et al.[163] demonstrated the efficiency of various entropy measures and discrete wavelet transform (DWT) combined with ML methods for diagnosing PD using EEG data. Also, the suggested method showed high accuracy.

Using longitudinal data from two PDBP and PPMI cohorts, Dadu et al.[164] developed ML-based models to detect distinct subgroups of PD and predict disease progression. Based on the study findings, AUC (ROC) was high, indicating a framework for predicting PD progression up to five years before diagnosis. Besides, Park et al.[165] presented a cost-effective method for predicting PD risk and enabling early detection through ML techniques and longitudinal health screening data. Cholesterol levels, blood pressure, and hemoglobin levels were identified as the most critical predictors. Moreover, the proposed method achieved a high AUC (ROC) for the neural network model. Further, Harvey et al.[166] introduced an ML-based approach to forecast cognitive outcomes in PD patients. In addition, the proposed approach obtained high sensitivity. Karabayir et al.[167] also provided a 1D-CNN-based method for predicting prodromal PD up to five years before clinical diagnosis by using 10-second ECG data. According to the findings, the method showed high specificity. Moreover, Leal et al.[168] developed a model to predict PD patients who are at risk of developing dyskinesia using clinical and behavioral data. The presented model using the RF classifier showed high AUC (ROC).

Qualitative analysis of biomarker-based approaches. We analyzed biomarker-based approaches thoroughly and introduced their strengths, weaknesses, opportunities, and threats to the classification in the context of PD diagnosis using ML:

- Strengths: The qualitative review led to two outcomes. First, having a strong biological view of PD may lead to early detection of the disease. Second, these approaches may be utilized for personal treatment and monitoring of disease progression.
- Weaknesses: While biomarker data collection is invasive, it may result in patient willingness to do tests. Many variables among different populations, such as genes, may hinder the generalization of models. These methods are expensive; indeed, they need many experts and resources to be accomplished.

Table 18 | Multimodal datasets

Article	Repository/Source	#PD	#HC	#PDF	#PDM	#HCF	#HCM	Information
170	PPMI[220]	885	–	–	–	–	–	Longitudinal data (years 0–4), 981 features: motor, non-motor, and imaging
171	PPMI[220]	423	196	146	277	70	126	7-year data collection from PPMI
	PDEP[288]	610	196	218	392	70	126	PDBP dataset used for validation
169	PPMI[220]	648	434	256	392	232	202	HC: average age 62.38 PD: average of 64.38
172	PhysioNet[250,251]	93	72	32	40	35	58	HC: average age (63.65 ± 8.58) PD: average age (66.30 ± 9.45)
15	PPMI[220]	73	59	21	52	16	43	Only individuals with complete MRI, SPECT, and CSF data were included
173	290	93	73	–	–	–	–	an average age of 66.3 years
	Sarkar dataset[192]	20	20	6	14	10	10	Audio features: Jitter, Shimmer, HNR, Auto-Correlation
	HandPD[291]	74	18	15	59	12	6	From São Paulo State University, Brazil
174	PPMI[220]	396	168	136	260	59	109	58 individuals with SWEDD (35 M, 23 F) with an average age of (60.6 ± 10) years, Participants lacking clinical and imaging features HC: average age (61.1 ± 11.3) PD: average age (61.7 ± 9.65)
175	Sarkar dataset[192]	20	20	6	14	10	10	26 features, including parameters related to time and frequency
	292	28	28	–	–	–	–	From Dandenong Neurology, Melbourne, VIC, Australia
	HandPD[291]	74	18	15	59	12	6	From São Paulo State University, Brazil
	NewHandPD[235]	31	35	–	–	–	–	A total of 66 individuals
	293	–	–	–	–	–	–	–
	294	–	–	–	–	–		1108 images, 659 Covid patients, 277 bacterial pneumonia cases, 175 healthy lungs
176	PPMI[220]	264	–	155	109	–	–	264 patients with known LRRK2
		129	–	79	50	–	–	–
	295,296	–	–	–	–	–	–	–
177	PPMI[220]	421	213	–	–	–	–	SWEED Group: 81 individuals, 465 M and 250 F
178	PPMI[220]	460	160	178	282	53	107	Data was obtained on January 6, 2023, from the PPMI database, A total of 675 entities, SWEED Group: 55 individuals (23 F, 32 M)
179	Clinical	50	25	–	–	–	–	14 neurocognitive tests, health questionnaires, and movement evaluations
180	PPMI[220]	427	171	–	–	–	–	The model was trained with PPMI data and validated with PDBP data
	PDEP[288]	804	442	–	–	–	–	

In this table, we detailed all the datasets in papers and compared participant demographics (number, gender, and health status: #PD => Parkinson's Disease, #HC => Number of Healthy Control Participants, #PDF => Number of Parkinson's disease Female Participants, #PDM => Number of Parkinson's disease Male Participants, #HCF => Number of Healthy Control Female Participants, #HCM => Number of Healthy Control Male Participants).

- Opportunities: Resolving these concerns may involve combining biomarker data with imaging and other clinical data, potentially leading to higher accuracy. Employing advanced technologies and methods such as genomics and proteomics may accelerate identifications. Moreover, scientists may use non-invasive approaches such as saliva and urine biomarkers to increase the willingness among patients.
- Threats: However, the quality of the collected data varies due to differences in data collection and storage methods. Also, there are ethical barriers to regulating and deploying biomarker data.

Multimodal

Multimodal datasets provide a complete PD diagnosis. These datasets contain biomarkers, medical imaging, movement patterns, and acoustic sounds. Improving the accuracy of early PD diagnosis through refined ML models requires the integration of various data input sources[169], as a holistic perspective is essential. Table 18 presents the details of the dataset, including data collection periods, the number of participants, database names, multimodal data types, and supplementary information. Table 19 covers the multimodal dataset-based articles, including their main ideas, advantages and limitations, applied tools, and algorithms. Furthermore, the evaluation metrics are outlined in Table 20. In Section "Multimodal-based approaches", multimodal-based approaches are reviewed. Section "Qualitative analysis of multimodal-based approaches" also includes a qualitative analysis of approaches, such as strengths, weaknesses, opportunities, and threats.

Multimodal-based approaches. Salmanpour et al.[170] identified optimal feature combinations to evaluate and predict subtypes of PD using feature

Table 19 | A comparison of multimodal datasets

Article	Main idea	Tools	Applied algorithms	Advantages	Disadvantages
170	PD subtype classification and forecasting using ML	• MATLAB	• PCA • KMeans algorithm • Infinite latent feature selection • Relief algorithm • FSKL-LLC • UFS-MCC • RDFSA • SF-PC • UFS-ASL • UFS-OL • LASSO • DT • SVM • KNN • LDA • NPNN • ECOCMC • MLP_BPC • RF • Ensemble Leaner Classifier	• Evaluation of diverse data • Utilizing various ML algorithms	• Low scalability
171	Statistical model for PD progression analysis	• Not mentioned	• IO-HMM • A contrastive latent variable model	• Analyzing longitudinal data	• Not generalizing to all Parkinson's patients • Not evaluating metrics such as sensitivity and specificity
169	Early-stage PD diagnosis using ML	• Not mentioned	• SVM • RF • LR • DT • ET • GNB • LightGBM • SGD • AdaBoost • KNN	• High accuracy	• Lack of MRI and other informative modalities • LSTM and RNN may be more accurate
172	Hybrid approach for PD identification and staging with ML	• Not mentioned	• DT • Multi-variate regression model	• High accuracy	• Limited subject group • Missing additional factors
15	DL frameworks for PD detection using multimodal features	• SPM • MATLAB	• Relief algorithm • CNN • SSAE	• High accuracy • High sensitivity • High specificity • High F1-score • High geometric-mean	• High complexity multi-modal feature-based approach • Not evaluating imbalanced datasets • Low scalability
173	Improved KNN algorithm for PD diagnosis	• Python	• KNN	• High accuracy • Lower margin of error • Improved performance with larger sample sizes • Compatible with both odd and even k-values	• Lack of reducing time complexity • Not evaluating the sensitivity to overlapping data
174	ML-based PD diagnosis using images and clinical data	• Python	• EBM	• High accuracy • High AUC (ROC)	• Low scalability • Imbalanced datasets make bias • Lack of correlation between data • MRI and other features not included
175	Quantum ReLU–based model for PD detection	• Python • MATLAB	• QReLU • CNN • M-QReLU	• High reliability • High accuracy	• Lack of evaluation the noisy text data • No assessment in small clinics with limited computation
176	Hybrid ML approach for PD diagnosis using pathogenic/non-pathogenic data	• Not mentioned	• AdaBoost • Bagging classifier • BNB • DT • ET • GNB • GB • KNN • LDA • LR • MLP • PAC • RF • Ridge Classifier	• High accuracy	• Limited to initial clinical variables

Table 19 (continued) | A comparison of multimodal datasets

Article	Main idea	Tools	Applied algorithms	Advantages	Disadvantages
			• SVM • Ensemble Voting • QDA		
177	PD prediction improvement using ant colony optimization	• MATLAB	• Ant colony optimizations • Regression neural network	• High accuracy • High F1-score • High sensitivity • High specificity	• Not evaluating metrics such as recall and TNR
178	1D-CNN-based detection of PD and SWEDD	• Not mentioned	• CNN	• High accuracy • High recall • High F1-score • High precision	• Low scalability
179	ML-based classification of PD and its stages using digital health data	• Python	• DT • CART	• High accuracy • High recall • High precision	• Low scalability
180	ML-based PD detection using multivariate data	• Python • GenoML	• MLP • LR • GB • AdaBoost • SGD • SVM • KNN • LDA • QDA • Bagging classifier • XGBoost • ERT • SHAP	• High AUC (ROC)	• Not involving other kinds of data, such as MRI or voice data

This table includes key concepts (main ideas), the utilized tools, the applied algorithms, advantages and disadvantages.

selection techniques and longitudinal datasets. In addition, various feature selection algorithms (FSAs), clustering, feature extraction algorithms (FEAs), and classification algorithms were utilized. Also, based on the findings, merging non-imaging data with SPECT-based radiomic features and the optimal use of hybrid ML systems (HMLs) enhanced the identification and prediction of PD subtypes in the fourth year. Similarly, Severson et al.[171] presented a statistical progression model using longitudinal data that accounted for intra-individual variations, medication effects, and inter-individual differences to better understand the heterogeneous symptoms and progression of PD. They also used a contrastive latent variable model and a customized input-output hidden Markov model as part of their method.

Junaid et al.[169] proposed an interpretable ML framework that utilizes multimodal data, including medication history, patient characteristics, and motor and non-motor data, to detect and predict early signs of PD. According to the results, the techniques were effective and accurate. Also, Khera and Kumar[172] proposed a hybrid strategy to classify the severity of the PD based on ML approaches. According to the evaluation results, the model could detect PD accurately. Pahuja and Prasad[15] presented the modal-level and feature-level frameworks based on DL architectures to improve the diagnosis of PD. In addition, this study used multi-modal features, including biological (CSF) and neuroimaging. The results also demonstrated that the provided frameworks had acceptable accuracy.

Richa Indu et al.[173] introduced a modified KNN algorithm based on handwriting, gait, and voice parameters for PD diagnosis. The authors enhanced this algorithm by incorporating the δ-neighborhood for predicting the class of test samples and the concept of weights. The findings showed that the proposed approach was accurate. In another study, Sarica et al.[174] suggested an ML technique using an explainable boosting machine (EBM) to classify SWEDD, PD, and HC. Additionally, they used imaging and clinical data in order to train their model. Despite the relatively small dataset, the findings demonstrated that the model performed remarkably well, achieving a high AUC-ROC score.

Parisi et al.[175] proposed two activation functions, namely quantum rectified linear unit (QReLU) and modified-QReLU (m-QReLU), aimed at enhancing the performance of CNN in tasks such as medical image classification, PD diagnosis, and COVID-19 detection. The results indicated that the proposed approach exhibited high accuracy and reliability. Also, Hajianfar et al.[176] offered a hybrid ML system to detect two essential genes, namely leucine-rich repeat kinase 2 (LRRK2) and glucocerebrosidase (GBA), using the PPMI database to diagnose PD. Also, several feature selection and feature extraction algorithms were used to reduce the number of variables and tackle overfitting. The evaluation results demonstrated high accuracy in performance.

Kanagaraj et al.[177] offered a method for enhancing the accuracy of PD diagnosis by utilizing PPMI data and an ant colony optimization approach. The proposed method resulted in more accurate predictions with fewer features, enhancing computational efficiency. Additionally, it demonstrated high accuracy. Also, Aggarwal et al.[178] proposed a method using 1-D CNN and data augmentation to classify non-PD, PD, and scans without evidence of dopamine deficit (SWEDD) and avoid misdiagnosis. Their approach achieved high-performance in precision, F1-score, recall, and accuracy. In another study, Templeton et al.[179] applied ML to classify PD and its stages through tablet-based neurocognitive assessments. The model utilized multimodal data, including speech, memory, evaluations of motor function, and the CART algorithm. Furthermore, the proposed model achieved high accuracy. Makarious et al.[180] presented a model for diagnosing and managing PD before patients recognize the signs and symptoms. This method was developed to identify PD by combining genetic, clinical, and demographic data with ML methods, such as AdaBoostClassifier and GenoML. According to the results, this method had a high AUC (ROC).

Qualitative analysis of multimodal-based approaches. Through critical review and analysis of the papers using multimodal-based approaches, we listed some of the most important strengths, weaknesses, opportunities, and threats of the classification:

- Strengths: The primary advantage is the accumulation of both motor and non-motor symptoms, which drives the resolution. Also, in this category, all the forms of datasets, such as acoustic, imaging, movement, and biomarker, are combined to increase accuracy.

Table 20 | Multimodal evaluation metrics

Article	Accuracy	Sensitivity	Specificity	F1-score	Precision	AUC (ROC)	FPR	MAE	Correlation coefficient	RMSE	Standard deviation	G-mean	Model likelihood	NPV
170	+								+					
171													+	
169	+	+		+	+									
172	+			+				+		+				
15	+	+	+	+								+		
173	+	+			+	+	+							
174	+	+	+	+		+								
175	+	+		+	+	+	+							
176	+										+			
177	+	+	+	+	+	+	+							
178	+	+	+	+	+									
179	+	+			+									
180	+	+	+		+	+								+

Plus (+) means the metric that has been evaluated in the paper, and blank cells mean the metric has not been evaluated in the paper.

- Weaknesses: However, these approaches require significant financial resources, as gathering clinical data involves various tests, and acquiring movement datasets may necessitate high-tech devices. Also, these need to be done with substantial computational resources and expertise, which makes the entire process too expensive.
- Opportunities: Paying attention to the mentioned scenarios may pave the way for future researchers. Developing new fusion techniques for multimodal datasets may optimize the process. Also, multimodal analysis may reveal a new pattern of biomarkers that were not obvious through single-modal analysis.
- Threats: The risk of data overfitting is inevitable and should be considered due to model complexity and limited data.

Analysis of results

The results of our systematic review procedure are discussed in this section. Section "Overview of the selected studies" presents an overview of the selected studies, while Section "Objectives" discusses the advantages, limitations, and differences among the various PD datasets and ML techniques. Analyzing classifications used in the dataset and helping to draw a future path for researchers are the ultimate goals of this study.

Overview of the selected studies

The goal of this study is to investigate state-of-the-art PD datasets currently used in diagnosis with the aid of ML methods. To achieve this, we have considered the following complementary questions (CQs):

- *CQ1*: Which groups are actively involved in diagnosing PD with ML applications?
- *CQ2*: Which publishing channels distributed the most papers?
- *CQ3*: How are the publications and studies distributed on diagnosing PD with ML approaches per publisher?
- *CQ4*: How are the publications and studies distributed on diagnosing PD with ML approaches per year?
- *CQ5*: How do the studies address the Parkinsonian syndrome classification years before diagnosing PD and performing complex classification tasks?

These questions aim to provide a clearer understanding of the current landscape of PD diagnosis using ML and identify future trends and potential research directions.

- *CQ1*: Which groups are actively involved in diagnosing PD with ML applications?

After synthesizing and selecting the papers, we extracted the authors' affiliations. Table 21 presents a comprehensive list of universities and institutes that have contributed at least twice in this field. Researchers from the University of British Columbia, Canada; Skolkovo Institute of Science and Technology in Russia; the University of Surabaya in Indonesia; RMIT University in Australia; FAU in Germany; and Stanford University in United States published a significant number of research papers on PD diagnosis through ML approaches.

- *CQ2*: Which publishing channels distributed the most papers?
  Table 22 shows that most JCR-indexed journal papers on PD diagnosis using ML approaches are published in BSPC, SR, NCA, MTAP, IEEE TONSRE, IEEE JOBHI, CIBM, and NPD. Table 22 provides a list of JCR-indexed journals that have published at least two related papers, including the publisher's name and abbreviation, as well as the journal's impact factor.
- *CQ3*: How are the publications and studies distributed on diagnosing PD with ML approaches per publisher?
  Figure 5 presents the distribution of publishers across the given subject annually. Figure 5A shows that Elsevier is the leading publisher, accounting for over one-third of the reviewed papers. In second place is Springer with 31%, followed by IEEE with 17% and Nature with 11%. Wiley has contributed 4%, and both Taylor & Francis and ACM each have a share of 1% of the total publications. Figure 5B–E showed the trends and number of studies about diagnosing PD with ML over the mentioned timeline based on publishers.
  The charts show that Wiley published four papers in 2021, while only two were published in 2023. Moreover, Elsevier's contribution was greater over these years, except in 2024, when only two papers were published by April. Springer maintained a steady pace, contributing at least ten papers annually, except in 2024, with only 5 papers published until April 2024. Meanwhile, IEEE showed a contribution in 2023, with a total of 12 publications. In 2021, *Nature* published only one paper, while during the rest of the timeframe, it showed a steady contribution of at least four papers annually. On the other hand, ACM and Taylor & Francis have the smallest share of releases, with just one publication each, making them the two publishers with the fewest papers.
- *CQ4*: How are the publications and studies distributed on diagnosing PD with ML applications per year?

Figure 5 illustrates a steady increase in publications, rising from 37 in 2021 to 42 in 2022, with an overall upward trend observed between 2021 and

Table 21 | Active groups and their research focus

University/Institute	Studies	Research focus
University of British Columbia, Canada	90,147,170,176	Medical images, Multimodal, Biomarkers
Skolkovo Institute of Science and Technology, Russia	113,119,139	Movement data
University of Surabaya, Indonesia	53,67,79	Acoustic data
RMIT University, Australia	53,67,79	Acoustic data
Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Germany	66,115,133	Movement data, Acoustic data
Stanford University, United States	147,151,167	Biomarkers
A. I. Burnazyan Federal Medical and Biophysical Center, Russia	113,139	Movement data
Medical Valley Digital Health Application Center, Germany	115,133	Movement data
Luxembourg Institute of Health, Luxembourg	115,133	Movement data
Tsinghua University, China	55,61	Acoustic data
Shandong First Medical University and Shandong Academy of Medical Sciences, China	140,148	Biomarkers, Movement data
Xi'an Jiaotong University, China	117,135	Movement data
Imam Khomeini Marine Science University, Iran	55,70	Acoustic data
Wenzhou University, China	55,70	Acoustic data
Hong Kong Baptist University, China	100,149	Biomarkers, Medical image
Sungkyunkwan University, South Korea	91,169	Multimodal, Medical image
Benha University, Egypt	153,169	Biomarker, Multimodal
Technical University of Kosice, Slovakia	79,121	Acoustic data, Movement data
University College London, UK	85,89	Medical image
Chongqing University, China	62,68	Acoustic data
University of Electronic Science and Technology of China (UESTC), China	58,85	Biomarker, Acoustic data
Chang Gung Memorial Hospital at Linkou, Taoyuan, Taiwan	85,150	Biomarker, Medical image
University of Sfax, National Engineering School of Sfax (ENIS), Tunisia	56,71	Acoustic data
Agri Ibrahim Cecen University, Turkey	57,106	Medical image, Acoustic data
Jamia Millia Islamia, India	65,101	Medical image, Acoustic data
University of Genoa, Genoa, Italy	98,136	Medical image, Movement data
Michael J. Fox Foundation, USA	136,171	Multimodal, Movement data
University of Alberta, Edmonton, Alberta, Canada	94,136	Medical image, Movement data
University of Illinois at Urbana-Champaign, USA	164,180	Biomarker, Multimodal
National Institutes of Health, USA	164,180	Biomarker, Multimodal
UCL Queen Square Institute of Neurology, UK	164,180	Biomarker, Multimodal
University College London, UK	164,180	Biomarker, Multimodal
Virginia Commonwealth University, USA	164,180	Biomarker, Multimodal
Georgia Institute of Technology, USA	52,180	Multimodal, Acoustic data

2024. In the year 2023, the total number of papers reached its highest point of 43. Even though just 11 papers were published up to April 2024, the research in this field is improving.

- *CQ5*: How do the studies address the Parkinsonian syndrome classification years before diagnosing PD and performing complex classification tasks?

According to Fig. 6, of the 133 studies reviewed, only 20% focused on the early diagnosis of PD, and none specifically addressed the classification of Parkinson's syndromes. This highlights a significant gap in research within this field. While many studies focused on binary classification tasks, such as distinguishing between PD and HC, these approaches may not adequately capture the complexity and challenges involved in clinical diagnosis. In complex cases such as those mentioned, there is a need for longitudinal datasets and subtle signal changes. Future research should focus on the Parkinsonian syndrome classification years before diagnosing PD to increase the real-world effectiveness of ML in PD diagnosis.

However, several studies[2,4,15,54,59,65,71,72,77,87,92,124,130,134,143,148,154–156,164,165,167,169,170,174,179,180] also focus on more complex tasks, such as diagnosing the early stages of PD using non-motor symptoms. These symptoms include olfactory impairment, depression, rapid eye movement, sleep behavior disorder (RBD), and data obtained from fMRI, saliva, or blood tests. Additionally, research has aimed to identify PD from similar conditions, such as SWEDD. These approaches highlight the potential of ML to tackle various diagnostic challenges beyond simple binary classification.

Table 23 shows that the acoustic data and biomarkers category contains the most papers on the early diagnosis of PD, while the medical imaging category has the fewest articles.

Objectives

In this section, we answer some research questions by providing statistical and analytical investigation results according to Section "Planning".

- *RQ1*: What types of datasets are used to diagnose PD?

In our detailed review in Section "Classification of the selected studies", we have distinguished the datasets used in the papers into five discernible categories–namely movement information, medical imaging, acoustic

Table 22 | Distribution of papers by publication channel

Publisher	Publication channel	Abbreviation	Count	Impact factor
IEEE	IEEE Transactions on Neural Systems and Rehabilitation Engineering	IEEE TONSRE	6	4.8
	IEEE Journal of Biomedical and Health Informatics	IEEE JOBHI	5	6.7
	IEEE Access	IA	4	3.4
	IEEE Transactions on Instrumentation and Measurement	IEEE TOIM	2	5.6
Elsevier	Biomedical Signal Processing and Control	BSPC	12	4.9
	Computers in Biology and Medicine	CIBM	5	7
	Parkinsonism & Related Disorders	PRD	4	3.1
	Expert Systems with Applications	ESWA	4	7.5
	NeuroImage: Clinical	NIC	3	3.4
	Computational Biology and Chemistry	CBC	2	2.6
	Applied Acoustics	AA	2	3.4
	Journal of Neuroscience Methods	JONM	2	2.7
	Computer Methods and Programs in Biomedicine	CMPIB	2	4.9
Springer	Neural Computing and Applications	NCA	7	4.5
	Multimedia Tools and Applications	MTAP	6	3
	Soft Computing	SC	3	3.1
	Physical and Engineering Sciences in Medicine	PESIM	2	2.4
	European Radiology	ER	2	4.7
	Annals of Nuclear Medicine	AONM	2	2.5
	Brain Imaging and Behavior	BIAB	2	2.4
	Neurological Sciences	NS	2	2.7
Wiley	Movement Disorders	MD	2	7.4
Nature	Scientific reports	SR	10	3.8
	npj Parkinson's disease	NPD	5	6.7

information, biomarkers, and multimodal– which has been provided in Fig. 4.

• *RQ2*: Which category of datasets is used the most to diagnose PD?

Based on our detailed review in the fourth section and the answer to *RQ1*. This paper categorizes datasets into five groups. According to Fig. 7, about 25% of the studies utilized movement data, while acoustic data was slightly less common, appearing in 23% of the cases. Medical imaging and biomarkers jointly ranked in the third place with two tenths, respectively. Also, multimodal with 11% is a novel approach that drives scientists' attention.

• *RQ3*: What tools are used the most in assessing ML approaches in diagnosing PD?

Figure 8 shows a collection of the tools demonstrated in Tables 6, 9, 12, 15 and 18. It shows the wide range of software and tools that experts have used in their work. Python ranks the most considerable proportion, accounting for 42% of usage, closely followed by MATLAB at 23%. R was used in 8% of the studies, while FSL, with 3%, was allocated the shortest rank. This data provides insight into the preferred computational tools in ML for assessing PD.

• *RQ4*: What metrics are significantly used to assess the ML techniques in diagnosing PD?

Applying ML for PD diagnosis has been assessed using various techniques and measures. The measures include accuracy, TPR (sensitivity), precision, TNR, F1-score, AUC (ROC), and MCC. Tables 7, 10, 13, 16 and 19 offer thorough assessments applied in review papers. In addition, Tables 6, 9, 12, 15 and 18 detail the benefits and downsides of each categorization approach. For a more detailed view, our statistical analyses are visually represented in Figs. 9 and 10, which show the comprehensive percentage distribution of the evaluation metrics and the parameter estimations by category, respectively. Figure 9 indicates that accuracy is the most frequently reported metric, representing 20% of the cases, followed by sensitivity at 18%. Precision, F1-score, and specificity are also significant, with 12% each. At the same time, balanced accuracy stood at the lowest end with only 1%.

According to Fig. 10, Studies involving acoustic data mostly focused on accuracy (21.13%), with a similarly strong emphasis on Sensitivity (20.42%). A minor 9.86% of studies considered AUC (ROC), and none of the papers evaluated balanced accuracy. As the focus shifted to movement data, about one-fifth of papers aimed to enhance accuracy. In medical imaging, the focus on accuracy peaked at 18.05%, and sensitivity reached 17.29%, while AUC and balanced accuracy were considered in 9.02% and 0.75% of the studies, respectively. Biomarker research displayed accuracy with 21.74% derived the most attention; also, AUC received significant attention at 18.26% compared to another group. When it comes to multimodal, accuracy accounted for 21.05% of studies. However, balanced accuracy rated just 3.51%, and it was the highest among other categories. This suggests a relatively balanced focus across evaluation metrics. Across all fields, accuracy consistently emerged as a critical metric in PD detection using ML approaches, although other metrics, such as balanced accuracy, F1-score, and AUC, might better represent model performance.

The focus on some evaluation criteria for ML models in PD diagnosis may lead to an inadequate evaluation of the model's actual performance. Figure 9 indicates that accuracy is the most frequently used criterion, warranting further investigation. Focusing insufficiently on class balance in the imbalanced dataset can reduce the effectiveness of accuracy as a performance evaluation metric. In medical data, the number of positive cases (patients) is typically much lower than that of negative cases (non-patients). Using accuracy as a metric does not effectively reflect a model's ability to identify minority classes correctly. Balanced accuracy is a more suitable measure for evaluating an imbalanced dataset, which appeared in only 1% of the reviewed articles. Similarly, F1-score and AUC (ROC) are widely used but remain less common than accuracy. Many reviewed papers in our study

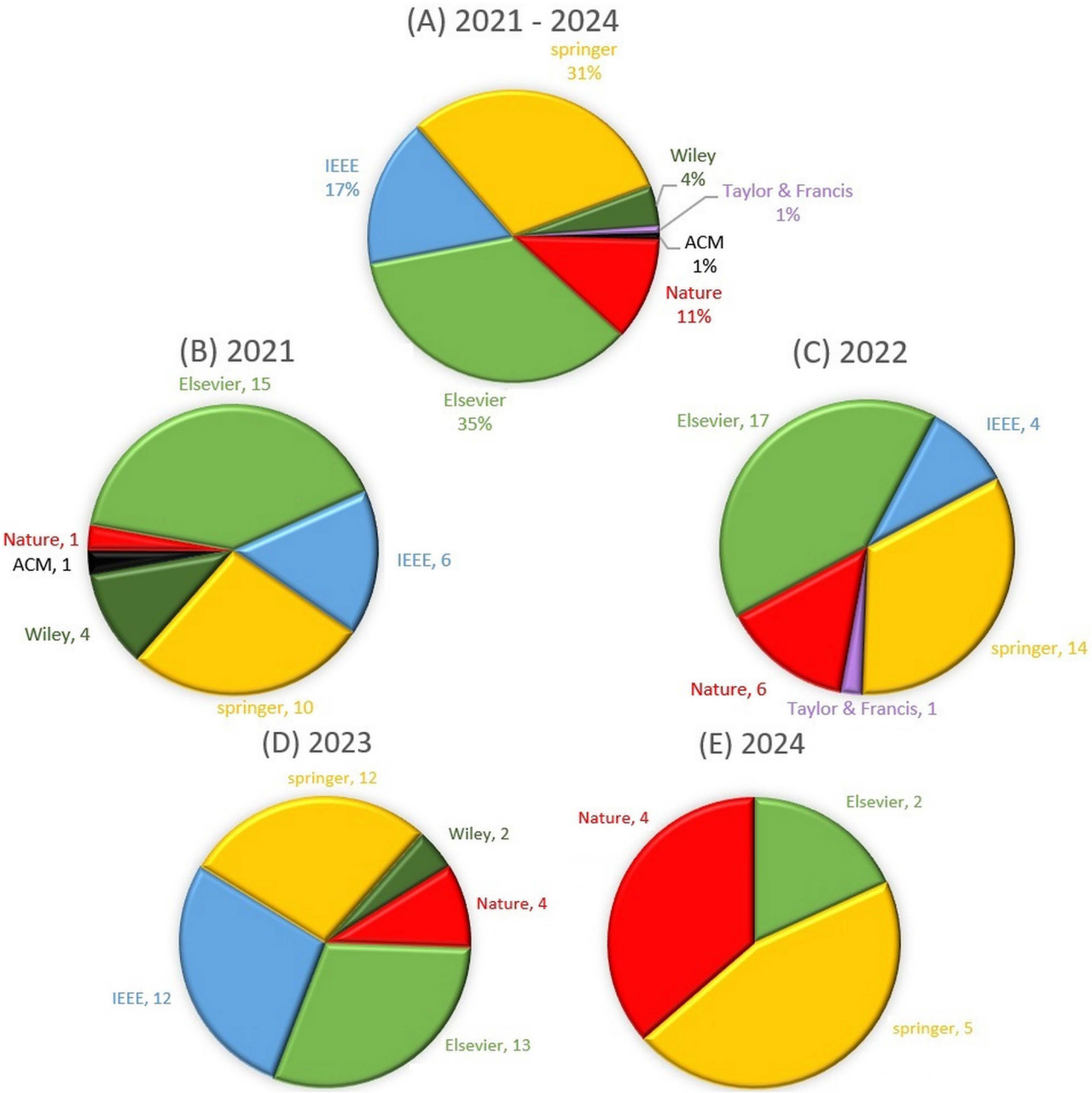


Fig. 5 | Annual distribution of the studied papers by publisher. Panels **A–E** present the relative share of different publishers for each year. **A** Year 2021-2024, **B** Year 2021, **C** Year 2022, **D** Year 2023, **E** Year 2024. The pie charts illustrate the percentage of papers published by each publisher in the corresponding year.

used techniques such as oversampling, class weighting, and robust algorithms to handle imbalanced data. Although accuracy and F1-score are often preferred due to their ease of use, it is imperative to employ more robust measures such as balanced accuracy and AUC (ROC) to assess model performance fully.

• *RQ5*: What ML algorithms have been considered the most in diagnosing PD?

Figure 11 shows the several ML approaches that significantly impact the diagnosis of PD. In this paper, we provide information based on the following categories: DL algorithm, regression algorithm, ensemble algorithm, reduction algorithm, clustering algorithm, ANN, optimization algorithm, feature selection algorithm, signal processing algorithm, and others. Figure 11 demonstrates that the most frequently used algorithms among all published papers were DL algorithms, which had a 20% share, followed by ensemble algorithms with 16%. SVMs, with 12%, ranked third in usage. Algorithms such as ANN, optimization, reduction, and regression algorithms each had usage rates ranging from 7% to 9%. In contrast, the least utilized approaches were signal processing techniques, clustering algorithms, and feature selection methods, with usage rates of 1%, 1%, and 2%, respectively.

In addition, the algorithmic advantages across different study topics for PD diagnosis are clearly shown in Fig. 12A–E. Figure 12A shows that DL was the most frequently used algorithm, appearing 16 times across the studies. Ensemble methods, SVM, ANN, and optimization algorithms followed, with usage rates of 12%, 11%, 10%, and 9%, respectively. Regarding medical imaging, DL algorithms recorded the first usage with 21 instances, according to Fig. 12B, while the reduction algorithm had the lowest usage. Figure 12C shows that both the DL algorithm and the ensemble algorithm were each used 17 times, whereas feature selection received the lowest contribution ranking. In the biomarker category (Fig. 12D), ensemble algorithms had the highest usage with 17 instances, followed by DL with 13, SVM with 11, and reduction algorithms with 10. In the final section of the

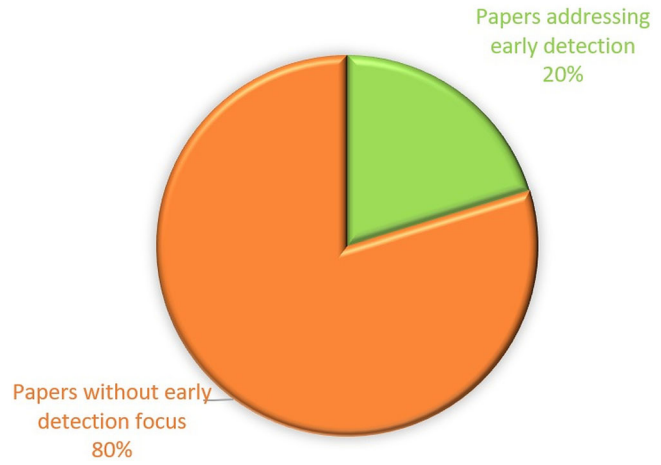


Fig. 6 | Distributing articles on early detection of PD.

Table 23 | Distribution of studies on the early detection of PD across categories

Ref	Category
54,59,65,71,72,77	Acoustic data
87,92	Medical images
4,124,130,134,143	Movement data
148,154–156,164,165,167	Biomarker's data
15,169,170,174,179,180	Multimodal data

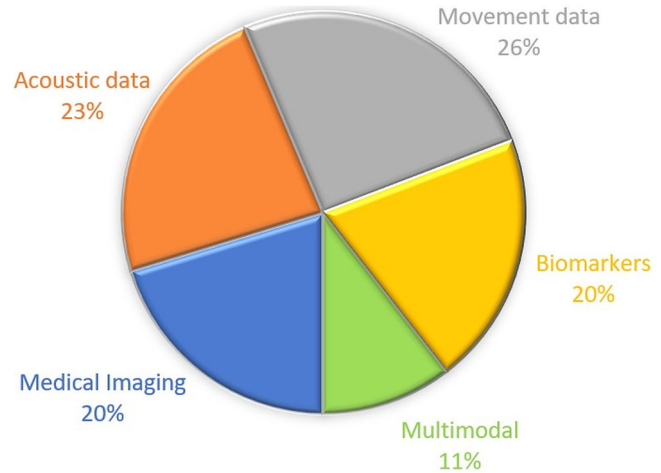


Fig. 7 | Percentage of papers included in this study for each classification.

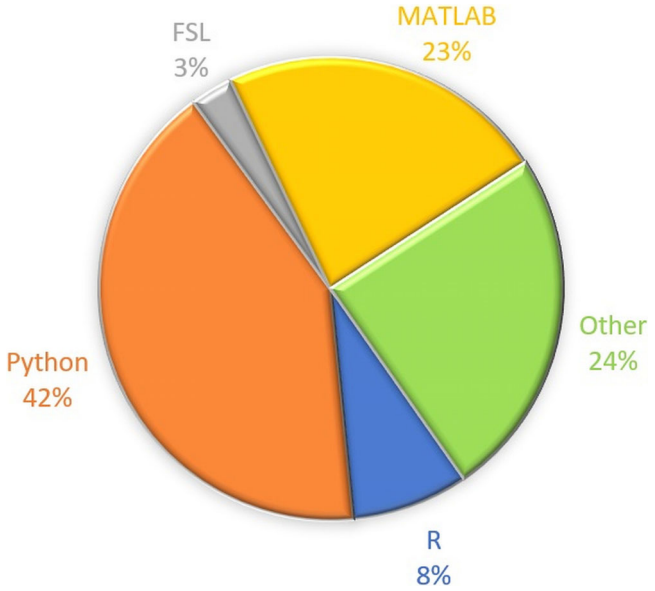


Fig. 8 | The proportion of tools utilized in the analyzed papers.

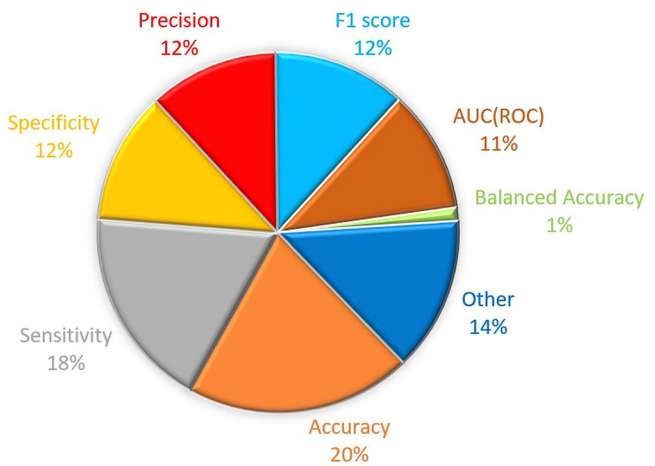


Fig. 9 | Percentage of evaluated parameters in studied papers.

figure, the ensemble ranked first, followed closely by regression algorithms in second place.

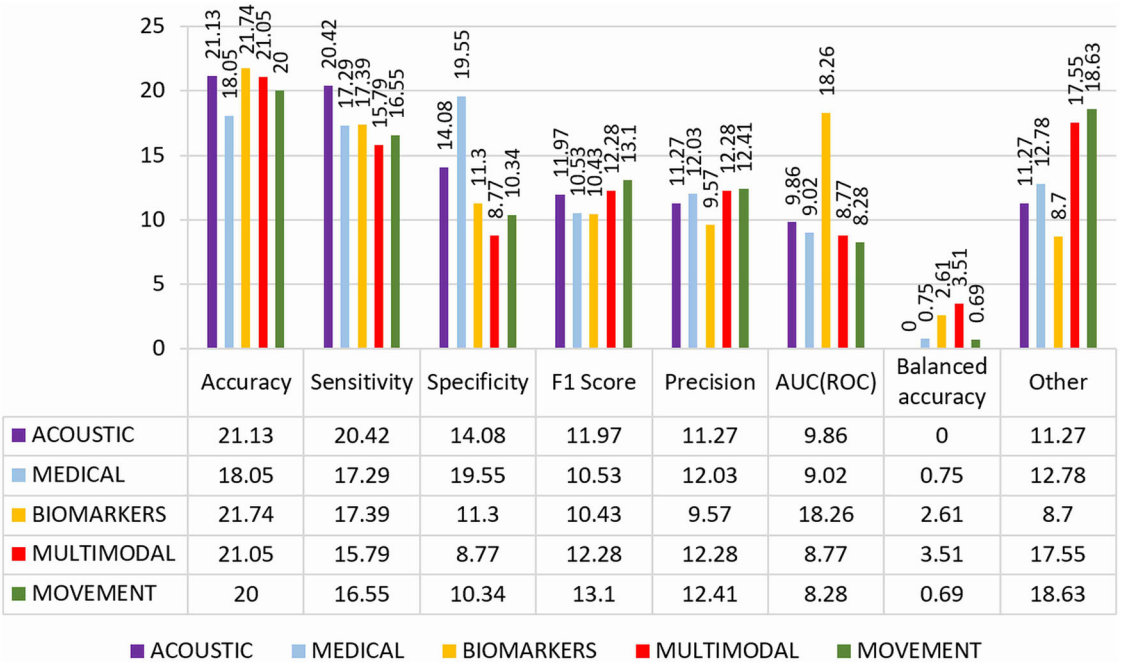
- *RQ6*: What validation methods are used in studies diagnosing PD with ML?

Figure 13 highlights that the reviewed studies commonly employ evaluation methods such as cross-validation and train-test split. About 49% of the studies used cross-validation to evaluate model performance. A total of 29% of the studies used Train-Test Split as an evaluation method, and only 19% used independent test sets to test the generalizability of the models. This is considered a significant limitation in evaluating the generalizability of the models. Furthermore, 3% of the studies did not directly mention the assessment method, raising concerns about the reliability of the results. As a result of the lack of transparency in this field, the model's performance in real-life conditions may be overestimated. In general, data leakage is a common pitfall in ML studies. This occurs when the model unintentionally learns from information in the test dataset. To ensure the reliability of the results, appropriate methods for data segmentation and careful evaluation strategies should be employed.

Open issues, challenges, and future trends

- *RQ7*: What are the major challenges, future trends, and open issues in diagnosing PD with ML?

We propose future research areas categorized into three groups: open issues, future trends, and challenges, see Fig. 14. This is motivated by the growing need to identify clinical and practical approaches for diagnosing PD using ML, as well as the review and analysis of data from selected papers. Each group is detailed below. The primary challenge in diagnosing PD via ML approaches is the absence of a predefined method, making it essential to employ ML approaches to evaluate clinical and non-clinical data from potential patients, healthy individuals, and those already diagnosed with PD. Therefore, managing data imbalance, regulating and normalizing, and overcoming ethical and legal obstacles remain the primary issues in this subject. Additionally, several problems in PD diagnosis remain unresolved, including the lack of advanced methods, the limited use of wearable technology, non-generalizable datasets, and insufficient language diversity in voice datasets. Furthermore, prospects for future research include the use of



	Accuracy	Sensitivity	Specificity	F1 Score	Precision	AUC(ROC)	Balanced accuracy	Other
ACOUSTIC	21.13	20.42	14.08	11.97	11.27	9.86	0	11.27
MEDICAL	18.05	17.29	19.55	10.53	12.03	9.02	0.75	12.78
BIOMARKERS	21.74	17.39	11.3	10.43	9.57	18.26	2.61	8.7
MULTIMODAL	21.05	15.79	8.77	12.28	12.28	8.77	3.51	17.55
MOVEMENT	20	16.55	10.34	13.1	12.41	8.28	0.69	18.63

Fig. 10 | The percentage of evaluation parameters in each category.

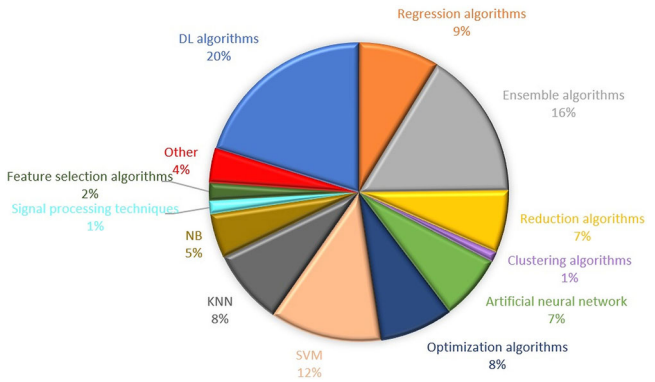


Fig. 11 | Percentage of algorithms used in reviewed studies.

genetic data, 3D dopamine imaging, and multimodal analysis. This introduction is followed by a detailed discussion of the challenges, potential developments, and unresolved concerns.

Open issues

- *Generalizable Datasets:* Many research have pointed out that there is an abundance of appropriate datasets for their evaluations[8,57,72,77,84,86,91,97,102,113,115,117,122,124,127–129,131,134,135,139,147,148,150,151,153,154,174,175]. In ML, one of the most fundamental principles is that the more data available for training, the more accurate model's results will be. In addition, the authors did not assess their ML methods on any other datasets. The absence of evidence demonstrating the performance of the models on other data types may have resulted in misleading findings, even if the models may have performed successfully on some of them. Consequently, larger and more diverse cohorts should be included in research to ensure that the results are both reliable and generalizable to a broader population. The low sample sizes and lack of generalizability might represent an ongoing, unresolved issue.
- *Model Generalization:* The capacity of models to generalize across diverse demographic and clinical contexts is a common source of worry. Indeed, the lack of thorough investigation using varied datasets limits our understanding of how well these models perform under diverse conditions. To address this, researchers are looking at ways to strengthen models by cross-validation and verifying them on different external cohorts[106,115,117,130,147,148,150,151]. Consequently, this ongoing issue may play a significant role in future research.
- *Lack of advanced analysis*: The need for more advanced analytical tools to manage complicated data aspects, particularly non-linear correlations and interactions among variables, is often brought up in discussions. This open issue will be discussed often by advocating the use of advanced ML and DL frameworks. This will allow for a better grasp of the intricacy involved[113,115,117,119,123,124,127–131,133–135,140,172].
- *Limited language diversity*: Most of the presented models, which analyze vocal datasets[73,74,77,138], have only been assessed by a single language, which again may produce inaccurate findings since other languages were unable to operate well with their models. Indeed, the amount of sound that may be made in different languages is varied; as a consequence, their method and results are only valid for a restricted number of individuals. In addition, the ethnicity and race of the participants have not been explored as a parameter in almost all of the datasets. These two factors have a direct bearing on the conclusions that may be drawn about a PD diagnosis since members of the same population might exhibit symptoms of this disease in the same manner. This open issue should be resolved to obtain more reliable results.
- *Personalized treatment*: The development of predictive models using ML has the potential to significantly enhance the early identification and treatment of PD, eventually leading to improved patient outcomes. Among the significant challenges that still need to be solved is the development of individualized prediction models that account for specific characteristics of an individual, such as their age, gender, medical history, genetics, and lifestyle variables. The possible outcome of this is that it might lead to more accurate predictions and improved treatment programs that are specifically customized to meet the requirements of those who are at risk for developing PD. Specifically, studies[74,128,151] raised questions in this regard.
- *Wearable technologies*: ML can evaluate patient body vibrations and wearable sensor data to diagnose better and monitor illness development. Another unresolved challenge is using wearable devices that

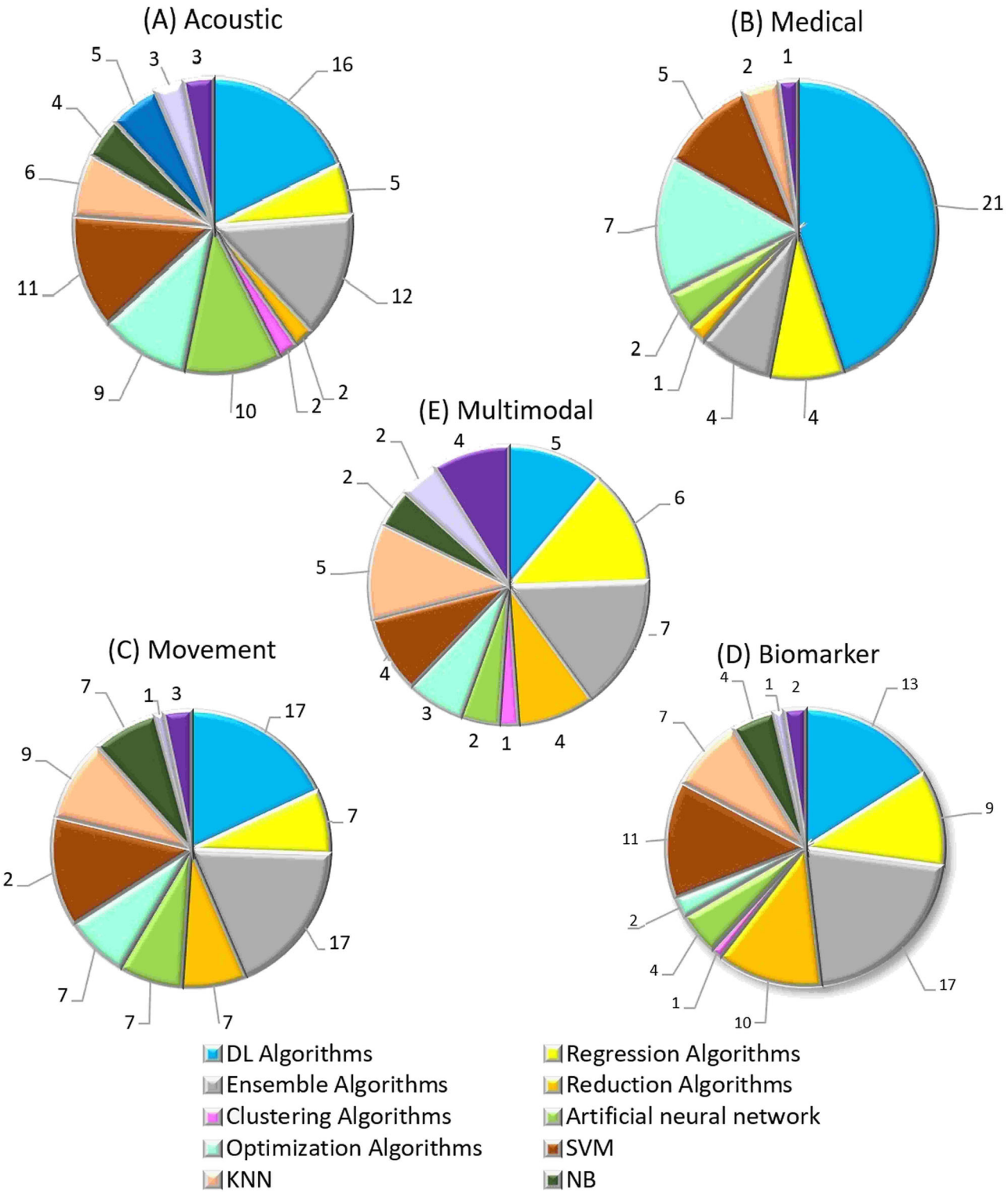


Fig. 12 | The percentage of algorithms used in each category. Panels **A**–**E** depict the proportion of algorithms employed within the five categories of the proposed taxonomy: **A** Acoustic, **B** Medical, **C** Movement, **D** Biomarker, and **E** Multimodal. Each pie chart reflects the proportion of algorithms utilized within the corresponding category, based on the reviewed studies.

continually monitor motor and non-motor symptoms in PD patients and giving real-time data to dynamically change therapy. Wearable sensors, such as smart watches, smart bands, and other available wearable sensors[9,181–183], can be utilized as a straightforward and accessible diagnostic tool for PD. Nevertheless, substantial work remains in this area. Wearable devices for continuous monitoring pose challenges such as ensuring data accuracy, enhancing patient comfort to improve compliance, securing sensitive health data, extending battery life, and effectively integrating collected data into clinical practice. Addressing these challenges is essential for using wearables to monitor PD and other neurological disorders, enhancing treatment options and patient outcomes. A gadget should be comfortable,

washable, and able to detect sickness phases and drug effects. For example, a wearable wristband collects data continuously over time and identifies the various symptoms of PD[74,122,131,135].

- *Multimodal analysis*: Multimodal ML algorithms are one way to use data from multiple datasets to diagnose PD more accurately. The method incorporates data from multiple datasets, such as wearable sensors, medical imaging, and clinical data, and optimized algorithms for each source. Although there are several studies on various biological signals to evaluate mobility impairments in PD patients, most studies only analyze a single method. Literature indicates that the multimodal method is more precise than utilizing each method independently. For example, in[184] and[185], the authors assessed existing PD staging systems and determined that incorporating additional modalities and functionalities beyond a single modality, such as motor symptoms, is essential. This approach would enable a more objective scoring system, leading to a more comprehensive evaluation of patients' symptoms and facilitating personalized treatment for each individual[186]. In general, existing signal processing and classification techniques incorporating data from many sensors have not been adequately evaluated. Although significant progress has been documented in multiple studies, there is currently no multimodal fusion system that can accurately forecast illness severity and track disease progression[8,62,92,102,119,131,140,174]. This open issue may play a vital role in future research.

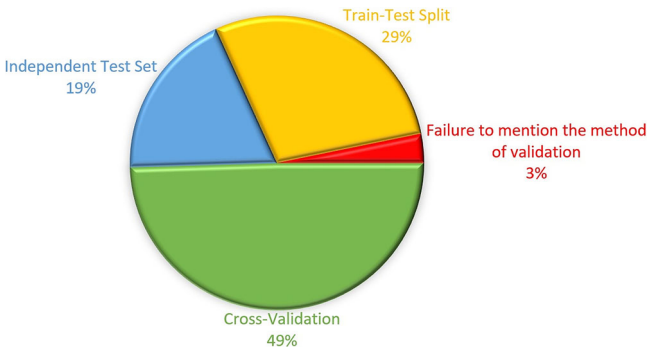


Fig. 13 | Validation methods used in the reviewed studies.

Fig. 14 | An overview of open issues, challenges, and future trends in diagnosing PD with ML.

Challenges

- *Data availability*: The limited availability of comprehensive datasets of good quality hinders research endeavors. Obtaining longitudinal data, diverse patient groups, and multimodal datasets that include clinical, genetic, and imaging data is a significant difficulty for many research studies. The majority of currently available data are either difficult to proprietary or access, belonging to hospitals or enterprises, and not openly accessible. This challenge affects the robustness, generalizability, and depth of research findings. Without diverse and extensive datasets, researchers cannot validate their models across different conditions, leading to less reliable and applicable results.
- *Regularization and normalization*: The regularization and normalization processes are crucial DL tools for PD identification. These strategies ensure the model learns valuable medical data patterns, improves generalizability, and prevents overfitting. This improves diagnostic accuracy and reliability. Even if these methods are not adopted widely by the DL research community, they have the potential to become a fruitful field of study in the following years.
- *Handling data imbalance*: An issue that often arises is the presence of imbalanced datasets, in which certain classes or outcomes are underrepresented. As a consequence of this imbalance, models may be slanted toward the class that comprises the majority. This, in turn, may have an effect on the accuracy of projections produced for underrepresented groups within the population. The inability of a model to accurately represent minority groups might result in diagnostic tools that are less practical and clinical insights that are prejudiced[130,154,175].
- *Computational resource limitation*: Advanced ML and DL methods often require substantial computational resources, which may not be available in all research settings, particularly in smaller clinics or institutions in developing regions. This restricts the ability of some researchers to employ the most advanced methodologies, potentially leading to slower progress or less innovative outcomes.
- *Ethical and regulatory hurdles*: In the healthcare industry, obtaining authentic patient data while safeguarding privacy is a significant challenge. This difficulty is compounded by the lack of balanced neurological disease datasets, which can adversely affect the efficiency of ML models. Furthermore, there are many obstacles to overcome when negotiating the ethical and regulatory environment, particularly when working with sensitive health data and using AI-driven technologies in clinical settings. The time required to obtain necessary approvals and ensure compliance with ethical standards and regulations can delay research progress and the implementation of findings in clinical practice.
- *Clinical adoption and usability*: Designing tools and models that are technically sound, user-friendly, and practically applicable in clinical workflows is a common obstacle. Even with strong technical performance, the lack of usability and integration into clinical practice can limit the adoption and impact of research innovations.
- *Real-testbed evaluation*: Clinical testing and evaluation of ML models is essential. The move from controlled research to real-world clinical practice involves significant problems. To ensure these models are productive and safe in real-world settings, patient variety, data heterogeneity, and ethical implications must be considered. Data scientists and healthcare practitioners must collaborate to create clinically relevant and accurate models. Healthcare experts can illuminate PD diagnosis, patient care, and the nuances of the disease. Early identification and management of PD can improve the lives of patients with this severe disorder.
- *Robust validation methods*: To address the identified methodological challenges, future studies should emphasize robust validation methods, such as external validation using completely independent datasets and k-fold cross-validation. Furthermore, clear documentation of the data preprocessing process, and the dataset segmentation methods is essential to reduce the risk of data leakage. The use of standard datasets to compare model performance, along with advanced techniques such as nested cross-validation, can enhance the validity of results. By concentrating on these methods, researchers can contribute more effectively, repeatedly, and transparently to this field.

Future trends

- *Multimodal brain MRI*: By revealing the structural and functional changes in the brain associated with PD, multimodal brain MRI can enhance the accuracy of PD diagnosis. In addition to aiding in early diagnosis and identifying prodromal stages of PD, advanced MRI techniques—when analyzed with ML— can improve diagnostic precision by integrating additional data layers from various MRI modalities[187]. Moreover, integrating MRI data with additional biomarkers, such as blood or cerebrospinal fluid indicators, through the use of ML models has the potential to dramatically improve the early identification and differential diagnosis of PD. The combination of multimodal MRI and ML offers new avenues for improving PD diagnosis.
- *3D Dopaminergic imaging*: Future research in PD should focus on developing 3D Dopaminergic imaging to map brain dopamine activity more precisely. This method has the potential to increase the accuracy of early diagnosis, develop individualized treatment methods, and evaluate the effectiveness of treatment procedures in real-time. While accurate visualization of dopaminergic neurotransmitter oscillations is expected to enhance the diagnosis of PD progression, it may also aid in the development of more effective medicines. To gain a better understanding of PD diagnosis and progression, three-dimensional maps of dopaminergic neurotransmitter activity in the brain are produced using advanced imaging methods[87,104].
- *Genetic and biomarker discovery*: Future research on PD should focus on genetic profiling and biomarker discovery to enhance diagnostic precision and personalized treatments. Researchers can identify unique genetic markers and better understand their influence on disease progression and treatment response by analyzing genetic data across diverse populations and correlating it with clinical outcomes. Through the use of this integrated method, new treatment targets may be discovered, enabling earlier diagnosis and more successful, individualized therapy of PD. To facilitate an earlier and more accurate diagnosis, it is necessary to carry out genetic profiling to find biomarkers predictive of the progression of PD[54,86,96,102,107,144].
- *Drug effect*: The use of ML in research on the influence of medications on neuromelanin in PD has the potential to dramatically improve the accuracy and speed of these studies. ML algorithms can analyze complex imaging data to find subtle changes in neuromelanin content and distribution in the SN as a result of therapies. These algorithms may be trained on datasets that are generated by modern 3D imaging techniques, which enables more accurate segmentation and analysis. In addition, ML models may be evaluated over a wide variety of external cohorts to evaluate their robustness. This helps ML models to accurately predict and generalize the effects of drugs on neuromelanin. This approach has the potential to develop therapeutic solutions that are more precisely targeted and a more comprehensive understanding of the progression of PD[87].
- *Symptom variability:* The diagnosis of PD and the subsequent surveillance of its progression may prove challenging due to the extensive range of individual symptoms. As a result of the fact that symptoms might differ in terms of their presentation, frequency, and intensity, developing an applicable diagnostic paradigm remains a challenging endeavor. This issue may be resolved, and PD diagnosis and therapy could be significantly enhanced by integrating a variety of data types into ML models. These data types may include imaging data, audio recordings, and clinical evaluations. The evaluation of all the patient symptoms can lead to the development of more personalized and effective medicines.

- *Advancing early detection and classification*: Future studies should prioritize the early detection of PD before severe symptoms appear, as well as the differentiation of various Parkinson's syndromes. Researchers can improve generalization across different populations by employing longitudinal data and multimodal datasets, along with advanced ML approaches, such as transfer learning and ensemble methods. Additionally, exploring tasks beyond simple binary classification—such as tracking disease progression and classifying subsets—holds significant potential for enhancing clinical applications.
- *Adoption of evaluation metrics*: A robust evaluation criteria selection is essential for the development of PD diagnosis models. Although accuracy is widely used as a metric, its limitations in imbalanced datasets highlight the need for alternative evaluation measures. Studies should focus on metrics such as F1-score, balanced accuracy, AUC (ROC), and MCC, which provide a more comprehensive assessment of model performance and address the imbalanced class. Furthermore, researchers should justify their selection of specific criteria to enhance transparency and improve research reproducibility. This approach leads to standardization of reporting criteria and facilitates comparability of results among various studies.

Threats to validity and limitations

This study aims to offer a systematic and comprehensive review that compares and classifies various ML methods for diagnosing PD. Although the result of SLR is typically reliable from different aspects[188], having limitations in these papers is inevitable[34]. Consequently, the most substantial limitation of this study has been highlighted below.

- Papers were selected from multiple well-known databases, including Elsevier, IEEE, Taylor & Francis, Springer, ACM, Nature, and Wiley, but we cannot certify that all related papers were selected. As a result, as described in Stage 3.2 of the conducting phase, it is possible that some articles were overlooked throughout the paper selection process.
- We classified the selected papers into five groups: acoustic data, biomarkers, movement data, medical imaging, and multimodal dataset. There could be alternative potential categorization, though.
- This SLR is organized around six fundamental questions, while there may be other important inquiries to consider.
- Despite the extensive literature on using ML approaches for diagnosing PD, this SLR focused on papers listed in the JCR, disregarding reputable conference papers. Furthermore, articles published nationally, book chapters, short papers, conference papers, editorial papers, and works written in languages other than English were not considered.
- This SLR can be considered highly credible due to the defined review protocol, adherence to a systematic process, and collaboration with multiple researchers in this research.

Conclusion

A comprehensive SLR in diagnosing PD based on ML is conducted in this study. This paper examines the use of well-known ML algorithms, as well as the employed datasets, repositories, applied tools, evaluation factors, validation methods, and relevant algorithms, comparing their advantages and disadvantages. In the first phase, 729 papers were collected based on a research query from 2021 to April 2024. The top 133 papers were chosen for the investigation based on methodology and inclusion/exclusion criteria. Elsevier and Springer together constitute two-thirds of the proportion in this field, with the former accounting for 35% and the latter for 31% of the contribution. Nature contributed only 11% of the papers. A fraction of a sixth of all publications will be devoted to the IEEE. On the other hand, Taylor & Francis and the ACM each recorded the least share, reaching barely one percent. Our study addressed a significant gap in research on early detection of PD, as only 20% of the reviewed articles focused on this aspect. We offered a category in which the selected papers were organized into five unique groups: acoustic data (23%), biomarkers (20%), medical imaging (20%), movement data (26%), and multimodal (11%). Statistical analysis reveals that 20% of the papers aimed to improve accuracy, while 18% focused on enhancing sensitivity. Python and MATLAB were the preferred tools, with 42% of publications using Python and 23% using MATLAB. The most commonly used algorithms were DL-based algorithms, which utilized in 20% of the selected papers. Based on the reviewed papers, nearly half of the studies (about 49%) utilized cross-validation methods to assess model performance, while 29% depended on the train-test split technique. The study emphasized open challenges and future trends in identifying PD using ML techniques. Open issues in this field include generalizable datasets, model generalization, lack of advanced analysis, limited language diversity in vocal datasets, personalized treatment, wearable technologies, and multimodal analysis. Additionally, significant challenges in this field, include data availability, regularization and normalization, handling data imbalance, computational resource limitation, ethical and regulatory hurdles, clinical adoption and usability, and real-testbed evaluation. Future trends may focus on multimodal brain MRI, 3D dopaminergic imaging, genetic and biomarker discovery, drug effect, and symptom variability.

Data availability

All data generated or analyzed during this study are included in this published article.

Received: 16 September 2024; Accepted: 22 May 2025;
Published online: 01 July 2025

References

1. Arora, P., Mishra, A. & Malhi, A. Machine learning Ensemble for the Parkinson's disease using protein sequences. *Multimed. Tools Appl.* **81**, 32215–32242 (2022).
2. Dhanalakshmi, S., Das, S. & Senthil, R. Speech features-based Parkinson's disease classification using combined SMOTE-ENN and binary machine learning. *Health Technol.* **14**, 393–406 (2024).
3. Frasca, M. & Tortora, G. Visualizing correlations among Parkinson biomedical data through information retrieval and machine learning techniques. *Multimed. Tools Appl.* **81**, 14685–14703 (2022).
4. Nilashi, M. et al. Early diagnosis of Parkinson's disease: a combined method using deep learning and neuro-fuzzy techniques. *Comput. Biol. Chem.* **102**, 107788 (2023).
5. Ghadimloozadeh, S., Sohrabi, M. R. & Fard, H. K. Development of rapid and simple spectrophotometric method for the simultaneous determination of anti-Parkinson drugs in combined dosage form using continuous wavelet transform and radial basis function neural network. *Optik* **242**, 167088 (2021).
6. Xie, S., Peng, P., Dong, X., Yuan, J. & Liang, J. Novel gene signatures predicting and immune infiltration analysis in Parkinson's disease: based on combining random forest with artificial neural network. *Neurol. Sci.* **45**, 2681–2696 (2024).
7. Abkenar, S. B. & Haghi Kashani, M. Chapter 30—Challenges of sensor network in smart hospitals. in *Sensor Networks for Smart Hospitals* (ed. Nguyen, T. A.) 617-636 (Elsevier, 2025).
8. Vidya, B. Gait based Parkinson's disease diagnosis and severity rating using multi-class support vector machine. *Appl. Soft Comput.* **113**, 107939 (2021).
9. Nikravan, M. & Haghi Kashani,M. Chapter 7—Smart medical sensor network. in Blockchain and Digital Twin for Smart Healthcare (ed. Nguyen, T. A.) 1-22 (Elsevier, 2025).
10. Nikravan, M. & Haghi Kashani, M. 6—Smart medical sensor network. in Blockchain and Digital Twin for Smart Healthcare (ed. Nguyen, T. A.) 99–120 (Elsevier, 2025).
11. Ajorloo, S., Jamarani, A., Kashfi, M., Kashani, M. H. & Najafizadeh, A. A systematic review of machine learning methods in software testing. *Appl. Soft Comput.* **162**, 111805 (2024).
12. Bazzaz Abkenar, S., Haghi Kashani, M. & Nikravan, M. Chapter 3—Telemedicine and remote patient monitoring. in *Smart Technologies for Sustainable Development Goals: Good Health and* Well-being (eds Anand, A. J. & Krishnan, S.) 1–21 (CRC Press, 2025).

13. Bazzaz Abkenar, S., Haghi Kashani, M. & Nikravan, M. Chapter 3—Feature engineering for threat detection. in *Handbook of AI-Driven Threat Detection and Prevention: A Holistic Approach to Security* (eds Bhambri, P. & Anand, A. J.) (CRC Press, 2025).
14. Kashani,M. H., Jamei,M., Akbari,M. & Tayebi,R. M. Utilizing bee colony to solve task scheduling problem in distributed systems. In *2011 Third International Conference on Computational Intelligence, Communication Systems and Networks,* pp. 298–303 (IEEE, 2011).
15. Pahuja, G. & Prasad, B. Deep learning architectures for Parkinson's disease detection by using multi-modal features. *Comput. Biol. Med.* **146**, 105610 (2022).
16. Nikravan, M., Haghi Kashani, M. & Bazzaz Abkenar, S. Chapter 10—Blockchain for secure health data management. in *Smart Technologies for Sustainable Development Goals: Good Health and Well-being* (Anand, A. J. & Krishnan, S.) 1–23 (CRC Press, 2025).
17. Nikravan, M., Haghi Kashani, M. & Bazzaz Abkenar, S. Chapter 4—Anomaly detection with artificial intelligence. in *Handbook of AI-Driven Threat Detection and Prevention: A Holistic Approach to Security* (eds Bhambri, P. & Anand, A. J.) 1–19 (CRC Press, 2025).
18. Kitchenham, B. *Procedures for Performing Systematic Reviews*, Vol. 33, 1–26 (Keele University, 2004).
19. Etemadi, M. et al. A systematic review of healthcare recommender systems: open issues, challenges, and techniques. *Expert Syst. Appl.* **213**, 118823 (2023).
20. Khachnaoui, H., Mabrouk, R. & Khlifa, N. Machine learning and deep learning for clinical data and PET/SPECT imaging in Parkinson's disease: a review. *IET Image Process* **14**, 4013–4026 (2020).
21. Salari, N. et al. The performance of various machine learning methods for Parkinson's disease recognition: a systematic review. *Curr. Psychol.* **42**, 16637–16660 (2022).
22. Tanveer, M., Rashid, A. H., Kumar, R. & Balasubramanian, R. Parkinson's disease diagnosis using neural networks: survey and comprehensive evaluation. *Inf. Process. Manag.* **59**, 102909 (2022).
23. Sigcha, L. et al. Deep learning and wearable sensors for the diagnosis and monitoring of Parkinson's disease: a systematic review. *Expert Syst. Appl.* **229**, 120541 (2023).
24. Skaramagkas, V., Pentari, A., Kefalopoulou, Z. & Tsiknakis, M. Multi-modal deep learning diagnosis of Parkinson's disease—a systematic review. *IEEE Trans. Neural Syst. Rehabil. Eng.* **31**, 2399–2423 (2023).
25. Amato, F., Saggio, G., Cesarini, V., Olmo, G. & Costantini, G. Machine learning- and statistical-based voice analysis of Parkinson's disease patients: a survey. *Expert Syst. Appl.* **219**, 119651 (2023).
26. Khanna, K., Gambhir, S. & Gambhir, M. Comparative analysis of machine learning techniques for Parkinson's detection: a review. *Multimed. Tools Appl* **82**, 45205–45231 (2023).
27. Keserwani, P. K., Das, S. & Sarkar, N. A comparative study: prediction of Parkinson's disease using machine learning, deep learning and nature inspired algorithm. *Multimed. Tools Appl* **83**, 69393–69441 (2024).
28. Islam, M. A., Hasan Majumder, M. Z., Hussein, M. A., Hossain, K. M. & Miah, M. S. A review of machine learning and deep learning algorithms for Parkinson's disease detection using handwriting and voice datasets. *Heliyon* **10**, e25469 (2024).
29. Sabherwal, G. & Kaur, A. Machine learning and deep learning approach to Parkinson's disease detection: present state-of-the-art and a bibliometric review. *Multimed. Tools Appl* **83**, 72997–73030 (2024).
30. Giannakopoulou, K.-M., Roussaki, I. & Demestichas, K. Internet of things technologies and machine learning methods for Parkinson's disease diagnosis, monitoring and management: a systematic review. *Sensors* **22**, 1799 (2022).
31. Rana, A. et al. Imperative role of machine learning algorithm for detection of Parkinson's disease: review, challenges and recommendations. *Diagnostics* **12**, 2003 (2022).
32. Zhang, J. Mining imaging and clinical data with machine learning approaches for the diagnosis and early detection of Parkinson's disease. *npj Parkinsons Dis.* **8**, 13 (2022).
33. Chandrabhatla, A. S., Pomeraniec, I. J. & Ksendzovsky, A. Co-evolution of machine learning and digital technologies to improve monitoring of Parkinson's disease motor symptoms. *NPJ Digital Med* **5**, 32 (2022).
34. Brereton, P., Kitchenham, B. A., Budgen, D., Turner, M. & Khalil, M. Lessons from applying the systematic literature review process within the software engineering domain. *J. Syst. Softw.* **80**, 571–583 (2007).
35. Haghi Kashani, M., Madanipour, M., Nikravan, M., Asghari, P. & Mahdipour, E. A systematic review of IoT in healthcare: applications, techniques, and trends. *J. Netw. Comput. Appl.* **192**, 103164 (2021).
36. Kitchenham, B. et al. Systematic literature reviews in software engineering—a systematic literature review. *Inf. Softw. Technol.* **51**, 7–15 (2009).
37. Calero, C., Bertoa, M. F. & Moraga, M. Á. A systematic literature review for software sustainability measures. in *2013 2nd International Workshop on Green and Sustainable Software (GREENS)*, 46–53 (IEEE, 2013).
38. Bazzaz Abkenar, S., Haghi Kashani, M., Akbari, M. & Mahdipour, E. Learning textual features for Twitter spam detection: a systematic literature review. *Expert Syst. Appl.* **228**, 120366 (2023).
39. Ahmadi, Z., Haghi Kashani, M., Nikravan, M. & Mahdipour, E. Fog-based healthcare systems: a systematic review. *Multimed. Tools Appl.* **80**, 36361–36400 (2021).
40. Khoshniat, N., Jamarani, A., Ahmadzadeh, A., Haghi Kashani, M. & Mahdipour, E. Nature-inspired metaheuristic methods in software testing. *Soft Comput* **28**, 1503–1544 (2024).
41. Jamarani, A. et al. Big data and predictive analytics: a systematic review of applications. *Artif. Intell. Rev.* **57**, 176 (2024).
42. Songhorabadi, M., Rahimi, M., MoghadamFarid, A. & Haghi Kashani, M. Fog computing approaches in IoT-enabled smart cities. *J. Netw. Comput. Appl.* **211**, 103557 (2023).
43. Nemati, S., Haghi Kashani, M. & Faghih Mirzaee, R. Comprehensive survey of ternary full adders: statistics, corrections, and assessments. *IET Circuits Devices Syst.* **17**, 111–134 (2023).
44. Haghi Kashani, M. & Mahdipour, E. Load balancing algorithms in fog computing. *IEEE Trans. Serv. Comput.* **16**, 1505–1521 (2023).
45. Sheikh Sofla, M., Haghi Kashani, M., Mahdipour, E. & Faghih Mirzaee, R. Towards effective offloading mechanisms in fog computing. *Multimed. Tools Appl.* **81**, 1997–2042 (2022).
46. Nikravan, M. & Haghi Kashani, M. A review on trust management in fog/edge computing: techniques, trends, and challenges. *J. Netw. Comput. Appl.* **204**, 103402 (2022).
47. Karimi, Y., Haghi Kashani, M., Akbari, M. & Mahdipour, E. Leveraging big data in smart cities: a systematic review. *Concurr. Comput. Pract. Exp.* **33**, e6379 (2021).
48. Fathi, M., Haghi Kashani, M., Jameii, S. M. & Mahdipour, E. Big data analytics in weather forecasting: a systematic review. *Arch. Comput. Methods Eng.* **29**, 1247–1275 (2021).
49. Rahimi, M., Songhorabadi, M. & Haghi Kashani, M. Fog-based smart homes: a systematic review. *J. Netw. Comput. Appl.* **153**, 102531 (2020).
50. Haghi Kashani, M., Rahmani, A. M. & Jafari Navimipour, N. Quality of service-aware approaches in fog computing. *Int. J. Commun. Syst.* **33**, e4340 (2020).
51. Bazzaz Abkenar, S., Haghi Kashani, M., Mahdipour, E. & Jameii, S. M. Big data analytics meets social media: a systematic review of techniques, open issues, and future directions. *Telemat. Inform.* **57**, 101517–110555 (2020).
52. Iyer, A. et al. A machine learning method to process voice samples for identification of Parkinson's disease. *Sci. Rep.* **13**, 20615 (2023).

53. Pah, N. D., Indrawati, V. & Kumar, D. K. Voice-based SVM model reliability for identifying Parkinson's disease. *IEEE Access* **11**, 144296–144305 (2023).
54. Dhar, J. An adaptive intelligent diagnostic system to predict early stage of Parkinson's disease using two-stage dimension reduction with genetically optimized lightgbm algorithm. *Neural Comput. Appl.* **34**, 4567–4593 (2022).
55. Yao, D., Chi, W. & Khishe, M. Parkinson's disease and cleft lip and palate of pathological speech diagnosis using deep convolutional neural networks evolved by IPWOA. *Appl. Acoust.* **199**, 109003 (2022).
56. Khaskhoussy, R. & Ayed, Y. B. Improving Parkinson's disease recognition through voice analysis using deep learning. *Pattern Recognit. Lett.* **168**, 64–70 (2023).
57. Celik, G. & Başaran, E. Proposing a new approach based on convolutional neural networks and random forest for the diagnosis of Parkinson's disease from speech signals. *Appl. Acoust.* **211**, 109476 (2023).
58. Ali, L. et al. A novel sample and feature dependent ensemble approach for Parkinson's disease detection. *Neural Comput. Appl* **35**, 15997–16010 (2022).
59. Masud, M. et al. CROWD: crow search and deep learning based feature extractor for classification of Parkinson's disease. *ACM Trans. Internet Technol.* **21**, 77 (2021).
60. Yücelbaş, C. A new approach: information gain algorithm-based k-nearest neighbors hybrid diagnostic system for Parkinson's disease. *Phys. Eng. Sci. Med.* **44**, 511–524 (2021).
61. Wang, X. et al. A Parkinson's auxiliary diagnosis algorithm based on a hyperparameter optimization method of deep learning. *IEEE/ACM Trans. Comput. Biol. Bioinform* **21**, 912–923 (2023).
62. Li, Y., Zhang, X., Wang, P., Zhang, X. & Liu, Y. Insight into an unsupervised two-step sparse transfer learning algorithm for speech diagnosis of Parkinson's disease. *Neural Comput. Appl.* **33**, 9733–9750 (2021).
63. Vital, T. P. R., Nayak, J., Naik, B. & Jayaram, D. Probabilistic neural network-based model for identification of Parkinson's disease by using voice profile and personal data. *Arab. J. Sci. Eng.* **46**, 3383–3407 (2021).
64. Jyotiyana, M., Kesswani, N. & Kumar, M. A deep learning approach for classification and diagnosis of Parkinson's disease. *Soft Comput* **26**, 9155–9165 (2022).
65. Masood, S., Maqsood, K. W., Pal, O. & Kumar, C. An ensemble-based feature selection framework for early detection of Parkinson's disease based on feature correlation analysis. *Math. Methods Appl. Sci.* https://doi.org/10.1002/mma.7835 (2021).
66. García, A. M. et al. Cognitive determinants of dysarthria in Parkinson's disease: an automated machine learning approach. *Mov. Disord.* **36**, 2862–2873 (2021).
67. Hireš, M. et al. Convolutional neural network ensemble for Parkinson's disease detection from voice recordings. *Comput. Biol. Med.* **141**, 105021 (2022).
68. Ma, J. et al. Deep dual-side learning ensemble model for Parkinson speech recognition. *Biomed. Signal Process. Control* **69**, 102849 (2021).
69. Quan, C., Ren, K., Luo, Z., Chen, Z. & Ling, Y. End-to-end deep learning approach for Parkinson's disease detection from speech signals. *Biocybern. Biomed. Eng.* **42**, 556–574 (2022).
70. Chen, F., Yang, C. & Khishe, M. Diagnose Parkinson's disease and cleft lip and palate using deep convolutional neural networks evolved by IP-based chimp optimization algorithm. *Biomed. Signal Process. Control* **77**, 103688 (2022).
71. Khaskhoussy, R. & Ayed, Y. B. Speech processing for early Parkinson's disease diagnosis: machine learning and deep learning-based approach. *Soc. Netw. Anal. Min.* **12**, 73 (2022).
72. Biswas, S. K. et al. Early detection of Parkinson disease using stacking ensemble method. *Comput. Methods Biomech. Biomed. Eng.* **26**, 527–539 (2023).
73. Liu, W. et al. Prediction of Parkinson's disease based on artificial neural networks using speech datasets. *J. Ambient Intell. Humaniz. Comput* **14**, 13571–13584 (2022).
74. Yuan, L., Liu, Y. & Feng, H.-M. Parkinson disease prediction using machine learning-based features from speech signal. *Serv. Oriented Comput. Appl* **18**, 101–107 (2023).
75. Kamalakannan, K., Anandharaj, G. & Gunavathie, M. A. Performance analysis of attributes selection and discretization of Parkinson's disease dataset using machine learning techniques: a comprehensive approach. *Int. J. Syst. Assur. Eng. Manag.* **14**, 1523–1529 (2023).
76. Saleh, S., Cherradi, B., El Gannour, O., Hamida, S. & Bouattane, O. Predicting patients with Parkinson's disease using Machine Learning and ensemble voting technique. *Multimed. Tools Appl* **83**, 33207–33234 (2023).
77. Devarajan, J. P., Sreedharan, V. R. & Narayanamurthy, G. Decision making in health care diagnosis: evidence from Parkinson's disease via hybrid machine learning. *IEEE Trans. Eng. Manag* **70**, 2719–2731 (2021).
78. Guatelli, R., Aubin, V., Mora, M., Naranjo-Torres, J. & Mora-Olivari, A. Detection of Parkinson's disease based on spectrograms of voice recordings and Extreme Learning Machine random weight neural networks. *Eng. Appl. Artif. Intell.* **125**, 106700 (2023).
79. Hireš, M., Drotár, P., Pah, N. D., Ngo, Q. C. & Kumar, D. K. On the inter-dataset generalization of machine learning approaches to Parkinson's disease detection from voice. *Int. J. Med. Inform.* **179**, 105237 (2023).
80. Eguchi, K. et al. Differentiation of speech in Parkinson's disease and spinocerebellar degeneration using deep neural networks. *J. Neurol.* **271**, 1004–1012 (2023).
81. Ali, L. et al. Parkinson's disease detection based on features refinement through L1 regularized SVM and deep neural network. *Sci. Rep.* **14**, 1333 (2024).
82. Johnson, K. A., Fox, N. C., Sperling, R. A. & Klunk, W. E. Brain imaging in Alzheimer disease. *Cold Spring Harb. Perspect. Med.* **2**, a006213 (2012).
83. Ding, C. W. et al. Prediction of Parkinson's disease by transcranial sonography-based deep learning. *Neurol. Sci.* **45**, 2641–2650 (2023).
84. Nakajima, K. et al. Diagnosis of Parkinson syndrome and Lewy-body disease using 123I-ioflupane images and a model with image features based on machine learning. *Ann. Nucl. Med.* **36**, 765–776 (2022).
85. Zhao, H. et al. Deep learning based diagnosis of Parkinson's Disease using diffusion magnetic resonance imaging. *Brain Imaging Behav.* **16**, 1749–1760 (2022).
86. Shibata, H. et al. Machine learning trained with quantitative susceptibility mapping to detect mild cognitive impairment in Parkinson's disease. *Parkinsonism Relat. Disord.* **94**, 104–110 (2022).
87. Gaurav, R. et al. NigraNet: an automatic framework to assess nigral neuromelanin content in early Parkinson's disease using convolutional neural network. *NeuroImage Clin.* **36**, 103250 (2022).
88. Nakano, T. et al. Neural networks associated with quality of life in patients with Parkinson's disease. *Parkinsonism Relat. Disord.* **89**, 6–12 (2021).
89. Dünnwald, M. et al. Fully automated deep learning-based localization and segmentation of the locus coeruleus in aging and Parkinson's disease using neuromelanin-sensitive MRI. *Int. J. Comput. Assist. Radiol. Surg.* **16**, 2129–2135 (2021).
90. Adams, M. P., Rahmim, A. & Tang, J. Improved motor outcome prediction in Parkinson's disease applying deep learning to DaTscan SPECT images. *Comput. Biol. Med.* **132**, 104312 (2021).
91. Shin, D. H. et al. Automated assessment of the substantia nigra on susceptibility map-weighted imaging using deep convolutional

neural networks for diagnosis of Idiopathic Parkinson's disease. *Parkinsonism Relat. Disord.* **85**, 84–90 (2021).
92. Xu, N., Zhou, Y., Patel, A., Zhang, N. & Liu, Y. Parkinson's disease diagnosis beyond clinical features: a bio-marker using topological machine learning of resting-state functional magnetic resonance imaging. *Neuroscience* **509**, 43–50 (2023).
93. Noella, R. S. N. & Priyadarshini, J. Diagnosis of Alzheimer's, Parkinson's disease and frontotemporal dementia using a generative adversarial deep convolutional neural network. *Neural Comput. Appl.* **35**, 2845–2854 (2023).
94. Camacho, M. et al. Explainable classification of Parkinson's disease using deep learning trained on a large multi-center database of T1-weighted MRI datasets. *NeuroImage Clin.* **38**, 103405 (2023).
95. Vyas, T. et al. Deep learning-based scheme to diagnose Parkinson's disease. *Expert Syst.* **39**, e12739 (2022).
96. Yasaka, K. et al. Parkinson's disease: deep learning with a parameter-weighted structural connectome matrix for diagnosis and neural circuit disorder investigation. *Neuroradiology* **63**, 1451–1462 (2021).
97. Dotinga, M. et al. Clinical value of machine learning-based interpretation of I-123 FP-CIT scans to detect Parkinson's disease: a two-center study. *Ann. Nucl. Med.* **35**, 378–385 (2021).
98. Piccardo, A. et al. The role of the deep convolutional neural network as an aid to interpreting brain [18F]DOPA PET/CT in the diagnosis of Parkinson's disease. *Eur. Radiol.* **31**, 7003–7011 (2021).
99. Sun, X. et al. Use of deep learning-based radiomics to differentiate Parkinson's disease patients from normal controls: a study based on [18F]FDG PET imaging. *Eur. Radiol.* **32**, 8008–8018 (2022).
100. Huang, W. et al. Auto diagnosis of Parkinson's disease via a deep learning model based on mixed emotional facial expressions. *IEEE J. Biomed. Health Inform.* **28**, 2547–2557 (2023).
101. Abdullah, S. M. et al. Deep transfer learning based Parkinson's disease detection using optimized feature selection. *IEEE Access* **11**, 3511–3524 (2023).
102. Pang, H. et al. Use of machine learning method on automatic classification of motor subtype of Parkinson's disease based on multilevel indices of rs-fMRI. *Parkinsonism Relat. Disord.* **90**, 65–72 (2021).
103. Balnarsaiah, B., Nayak, B. A., Sujeetha, G. S., Babu, B. S. & Vallabhaneni, R. B. Parkinson's disease detection using modified ResNeXt deep learning model from brain MRI images. *Soft Comput* **27**, 11905–11914 (2023).
104. Cui, X. et al. An adaptive weighted attention-enhanced deep convolutional neural network for classification of MRI images of Parkinson's disease. *J. Neurosci. Methods* **394**, 109884 (2023).
105. Wang, Y. et al. An automatic interpretable deep learning pipeline for accurate Parkinson's disease diagnosis using quantitative susceptibility mapping and T1-weighted images. *Hum. Brain Mapp.* **44**, 4426–4438 (2023).
106. Keles, A., Keles, A., Keles, M. B. & Okatan, A. PARNet: deep neural network for the diagnosis of parkinson's disease. *Multimed. Tools. Appl.* **83**, 35781–35793 (2023).
107. Khachnaoui, H., Chikhaoui, B., Khlifa, N. & Mabrouk, R. Enhanced Parkinson's disease diagnosis through convolutional neural network models applied to SPECT DaTSCAN images. *IEEE Access* **11**, 91157–91172 (2023).
108. Zhang, X. et al. Multi-level graph neural network with sparsity pooling for recognizing Parkinson's disease. *IEEE Trans. Neural Syst. Rehabil. Eng.* **31**, 4459–4469 (2023).
109. Veetil, I. K., Chowdary, D. E., Chowdary, P. N., Sowmya, V. & Gopalakrishnan, E. A. An analysis of data leakage and generalizability in MRI based classification of Parkinson's Disease using explainable 2D Convolutional Neural Networks. *Digital Signal Process* **147**, 104407 (2024).
110. Tran, C. et al. Deep learning predicts prevalent and incident Parkinson's disease from UK Biobank fundus imaging. *Sci. Rep.* **14**, 3637 (2024).
111. Archila, J., Manzanera, A. & Martínez, F. A multimodal Parkinson quantification by fusing eye and gait motion patterns, using covariance descriptors, from non-invasive computer vision. *Comput. Methods Prog. Biomed.* **215**, 106607 (2022).
112. Sotirakis, C. et al. Identification of motor progression in Parkinson's disease using wearable sensors and machine learning. *npj Parkinsons Dis.* **9**, 142 (2023).
113. Kovalenko, E. et al. Distinguishing between Parkinson's disease and essential tremor through video analytics using machine learning: a pilot study. *IEEE Sens. J.* **21**, 11916–11925 (2021).
114. Sun, H., Ye, Q. & Xia, Y. Predicting freezing of gait in patients with Parkinson's disease by combination of manually-selected and deep learning features. *Biomed. Signal Process. Control* **88**, 105639 (2024).
115. Ullrich, M. et al. Detection of unsupervised standardized gait tests from real-world inertial sensor data in Parkinson's disease. *IEEE Trans. Neural Syst. Rehabil. Eng.* **29**, 2103–2111 (2021).
116. Rezaee, K., Savarkar, S., Yu, X. & Zhang, J. A hybrid deep transfer learning-based approach for Parkinson's disease classification in surface electromyography signals. *Biomed. Signal Process. Control* **71**, 103161 (2022).
117. Zhao, H. et al. Severity level diagnosis of Parkinson's disease by ensemble K-nearest neighbor under imbalanced data. *Expert Syst. Appl.* **189**, 116113 (2022).
118. Borzì, L., Sigcha, L., Rodríguez-Martín, D. & Olmo, G. Real-time detection of freezing of gait in Parkinson's disease using multi-head convolutional neural networks and a single inertial sensor. *Artif. Intell. Med.* **135**, 102459 (2023).
119. Shcherbak, A., Kovalenko, E. & Somov, A. Detection and classification of early stages of Parkinson's disease through wearable sensors and machine learning. *IEEE Trans. Instrum. Meas.* **72**, 1–9 (2023).
120. Pedrero-Sánchez, J. F., Belda-Lois, J.-M., Serra-Añó, P., Inglés, M. & López-Pascual, J. Classification of healthy, Alzheimer and Parkinson populations with a multi-branch neural network. *Biomed. Signal Process. Control* **75**, 103617 (2022).
121. Gazda, M., Hireš, M. & Drotár, P. Multiple-fine-tuned convolutional neural networks for Parkinson's disease diagnosis from offline handwriting. *IEEE Trans. Syst. Man Cybern. Syst.* **52**, 78–89 (2022).
122. Ibrahim, A., Zhou, Y., Jenkins, M. E., Trejos, A. L. & Naish, M. D. Real-time voluntary motion prediction and Parkinson's tremor reduction using deep neural networks. *IEEE Trans. Neural Syst. Rehabil. Eng.* **29**, 1413–1423 (2021).
123. Kaur, R., Motl, R. W., Sowers, R. & Hernandez, M. E. A vision-based framework for predicting multiple sclerosis and Parkinson's disease gait dysfunctions—a deep learning approach. *IEEE J. Biomed. Health Inform.* **27**, 190–201 (2023).
124. Lin, C. H. et al. Early detection of Parkinson's disease by neural network models. *IEEE Access* **10**, 19033–19044 (2022).
125. Exley, T., Moudy, S., Patterson, R. M., Kim, J. & Albert, M. V. Predicting UPDRS motor symptoms in individuals with Parkinson's disease from force plates using machine learning. *IEEE J. Biomed. Health Inform.* **26**, 3486–3494 (2022).
126. Li, H., He, Q. & Wu, L. Detection of brain abnormalities in Parkinson's rats by combining deep learning and motion tracking. *IEEE Trans. Neural Syst. Rehabil. Eng.* **31**, 1001–1007 (2023).
127. Cesarelli, G. et al. Using features extracted from upper limb reaching tasks to detect Parkinson's disease by means of machine learning models. *IEEE Trans. Neural Syst. Rehabil. Eng.* **31**, 1056–1063 (2023).
128. Alissa, M. et al. Parkinson's disease diagnosis using convolutional neural networks and figure-copying tasks. *Neural Comput. Appl.* **34**, 1433–1453 (2022).

129. Wang, Q., Zeng, W. & Dai, X. Gait classification for early detection and severity rating of Parkinson's disease based on hybrid signal processing and machine learning methods. *Cogn. Neurodyn* **18**, 109–132 (2022).
130. Ma, Y.-W., Chen, J.-L., Chen, Y.-J. & Lai, Y.-H. Explainable deep learning architecture for early diagnosis of Parkinson's disease. *Soft Comput* **27**, 2729–2738 (2023).
131. Ferreira, M. I. A. S. N., Barbieri, F. A., Moreno, V. C., Penedo, T. & Tavares, J. M. R. S. Machine learning models for Parkinson's disease detection and stage classification based on spatial-temporal gait parameters. *Gait Posture* **98**, 49–55 (2022).
132. Valla, E., Nõmm, S., Medijainen, K., Taba, P. & Toomela, A. Tremor-related feature engineering for machine learning based Parkinson's disease diagnostics. *Biomed. Signal Process. Control* **75**, 103551 (2022).
133. da Rosa Tavares, J. E. et al. uTUG: an unsupervised timed up and go test for Parkinson's disease. *Biomed. Signal Process. Control* **81**, 104394 (2023).
134. Roy, S., Roy, U., Sinha, D. & Pal, R. K. Imbalanced ensemble learning in determining Parkinson's disease using Keystroke dynamics. *Expert Syst. Appl.* **217**, 119522 (2023).
135. Zhao, H. et al. Accurate identification of Parkinson's disease by distinctive features and ensemble decision trees. *Biomed. Signal Process. Control* **69**, 102860 (2021).
136. Mirelman, A. et al. Detecting sensitive mobility features for Parkinson's disease stages via machine learning. *Mov. Disord.* **36**, 2144–2155 (2021).
137. Kumar, K. & Ghosh, R. Parkinson's disease diagnosis using recurrent neural network based deep learning model by analyzing online handwriting. *Multimed. Tools Appl.* **83**, 11687–11715 (2023).
138. Zhao, A. & Li, J. A significantly enhanced neural network for handwriting assessment in Parkinson's disease detection. *Multimed. Tools Appl.* **82**, 38297–38317 (2023).
139. Talitckii, A. et al. Defining optimal exercises for efficient detection of Parkinson's disease using machine learning and wearable sensors. *IEEE Trans. Instrum. Meas.* **70**, 1–10 (2021).
140. Chen, M., Sun, Z., Xin, T., Chen, Y. & Su, F. An interpretable deep learning optimized wearable daily detection system for Parkinson's disease. *IEEE Trans. Neural Syst. Rehabil. Eng.* **31**, 3937–3946 (2023).
141. Varghese, J. et al. Machine learning in the Parkinson's disease smartwatch (PADS) dataset. *npj Parkinsons Dis.* **10**, 9 (2024).
142. Yang, Y. et al. MELPD-detector: multi-level ensemble learning method based on adaptive data augmentation for Parkinson disease detection via free-KD. *CCF Trans. Pervasive Comput. Interact.* **6**, 182–198 (2024).
143. Cuk, A. et al. Tuning attention based long-short term memory neural networks for Parkinson's disease detection using modified metaheuristics. *Sci. Rep.* **14**, 4309 (2024).
144. Bhandari, N., Walambe, R., Kotecha, K. & Kaliya, M. Integrative gene expression analysis for the diagnosis of Parkinson's disease using machine learning and explainable AI. *Comput. Biol. Med.* **163**, 107140 (2023).
145. Göker, H. Automatic detection of Parkinson's disease from power spectral density of electroencephalography (EEG) signals using deep learning model. *Phys. Eng. Sci. Med.* **46**, 1163–1174 (2023).
146. Lal, U., Chikkankod, A. V. & Longo, L. Fractal dimensions and machine learning for detection of Parkinson's disease in resting-state electroencephalography. *Neural Comput. Appl.* **36**, 8257–8280 (2024).
147. Lee, S., Hussein, R., Ward, R., Jane Wang, Z. & McKeown, M. J. A convolutional-recurrent neural network approach to resting-state EEG classification in Parkinson's disease. *J. Neurosci. Methods* **361**, 109282 (2021).
148. Chang, H., Liu, B., Zong, Y., Lu, C. & Wang, X. EEG-based Parkinson's disease recognition via attention-based sparse graph convolutional neural network. *IEEE J. Biomed. Health Inform.* **27**, 5216–5224 (2023).
149. Wang, X. et al. Urine biomarkers discovery by metabolomics and machine learning for Parkinson's disease diagnoses. *Chin. Chem. Lett.* **34**, 108230 (2023).
150. Yang, C.-Y. & Huang, Y.-Z. Parkinson's disease classification using machine learning approaches and resting-state EEG. *J. Med. Biol. Eng.* **42**, 263–270 (2022).
151. Shabanpour, M., Kaboodvand, N. & Iravani, B. Parkinson's disease is characterized by sub-second resting-state spatio-oscillatory patterns: a contribution from deep convolutional neural network. *NeuroImage Clin.* **36**, 103266 (2022).
152. Coelho, B. F. O. et al. Parkinson's disease effective biomarkers based on Hjorth features improved by machine learning. *Expert Syst. Appl.* **212**, 118772 (2023).
153. Hosny, M., Zhu, M., Gao, W. & Fu, Y. A novel deep learning model for STN localization from LFPs in Parkinson's disease. *Biomed. Signal Process. Control* **77**, 103830 (2022).
154. Martinez-Eguiluz, M. et al. Diagnostic classification of Parkinson's disease based on non-motor manifestations and machine learning strategies. *Neural Comput. Appl.* **35**, 5603–5617 (2023).
155. Xu, W. et al. Diagnosis of Parkinson's disease via the metabolic fingerprint in saliva by deep learning. *Small Methods* **7**, 2300285 (2023).
156. Zhang, R., Jia, J. & Zhang, R. EEG analysis of Parkinson's disease using time–frequency analysis and deep learning. *Biomed. Signal Process. Control* **78**, 103883 (2022).
157. Dar, M. N., Akram, M. U., Yuvaraj, R. & Khawaja, S. G. ul & Murugappan, M. EEG-based emotion charting for Parkinson's disease patients using convolutional recurrent neural networks and cross dataset learning. *Comput. Biol. Med.* **144**, 105327 (2022).
158. Nour, M., Senturk, U. & Polat, K. Diagnosis and classification of Parkinson's disease using ensemble learning and 1D-PDCovNN. *Comput Biol. Med.* **161**, 107031 (2023).
159. Chu, C. et al. An enhanced EEG microstate recognition framework based on deep neural networks: an application to Parkinson's disease. *IEEE J. Biomed. Health Inform.* **27**, 1307–1318 (2023).
160. Lin, W. C. et al. Electroencephalogram-driven machine-learning scenario for assessing impulse control disorder comorbidity in Parkinson's disease using a low-cost, custom lego-like headset. *IEEE Trans. Neural Syst. Rehabil. Eng.* **31**, 4106–4114 (2023).
161. Wu, J., Wu, W., Jiang, P., Xu, Y. & Yu, M. Identification of SV2C and DENR as key biomarkers for Parkinson's disease based on bioinformatics, machine learning, and experimental verification. *J. Mol. Neurosci.* **74**, 6 (2024).
162. Uehara, Y. et al. Non-invasive diagnostic tool for Parkinson's disease by sebum RNA profile with machine learning. *Sci. Rep.* **11**, 18550 (2021).
163. Aljalal, M., Aldosari, S. A., Molinas, M., AlSharabi, K. & Alturki, F. A. Detection of Parkinson's disease from EEG signals using discrete wavelet transform, different entropy measures, and machine learning techniques. *Sci. Rep.* **12**, 22547 (2022).
164. Dadu, A. et al. Identification and prediction of Parkinson's disease subtypes and progression using machine learning in two cohorts. *npj Parkinson. Dis.* **8**, 172 (2022).
165. Park, Y. H. et al. Machine learning based risk prediction for Parkinson's disease with nationwide health screening data. *Sci. Rep.* **12**, 19499 (2022).
166. Harvey, J. et al. Machine learning-based prediction of cognitive outcomes in de novo Parkinson's disease. *npj Parkinsons Dis.* **8**, 150 (2022).
167. Karabayir, I. et al. Externally validated deep learning model to identify prodromal Parkinson's disease from electrocardiogram. *Sci. Rep.* **13**, 12290 (2023).

168. Leal, D. A. B., Dias, C. M. V., Ramos, R. P. & Brys, I. Prediction of dyskinesia in Parkinson's disease patients using machine learning algorithms. *Sci. Rep.* **13**, 22426 (2023).
169. Junaid, M., Ali, S., Eid, F., El-Sappagh, S. & Abuhmed, T. Explainable machine learning models based on multimodal time-series data for the early detection of Parkinson's disease. *Comput. Methods Prog. Biomed.* **234**, 107495 (2023).
170. Salmanpour, M. R., Shamsaei, M. & Rahmim, A. Feature selection and machine learning methods for optimal identification and prediction of subtypes in Parkinson's disease. *Comput. Methods Prog. Biomed.* **206**, 106131 (2021).
171. Severson, K. A. et al. Discovery of Parkinson's disease states and disease progression modelling: a longitudinal data study using machine learning. *Lancet Digital Health* **3**, e555–e564 (2021).
172. Khera, P. & Kumar, N. Novel machine learning-based hybrid strategy for severity assessment of Parkinson's disorders. *Med. Biol. Eng. Comput.* **60**, 811–828 (2022).
173. Indu, R., Dimri, S. C. & Malik, P. A modified kNN algorithm to detect Parkinson's disease. *Netw. Model. Anal. Health Inform. Bioinforma.* **12**, 24 (2023).
174. Sarica, A., Quattrone, A. & Quattrone, A. Explainable machine learning with pairwise interactions for the classification of Parkinson's disease and SWEDD from clinical and imaging features. *Brain Imaging Behav.* **16**, 2188–2198 (2022).
175. Parisi, L., Neagu, D., Ma, R. & Campean, F. Quantum ReLU activation for convolutional neural networks to improve diagnosis of Parkinson's disease and COVID-19. *Expert Syst. Appl.* **187**, 115892 (2022).
176. Hajianfar, G. et al. Prediction of Parkinson's disease pathogenic variants using hybrid Machine learning systems and radiomic features. *Phys. Med. Eur. J. Med. Phys.* **113**, 102647 (2023).
177. Kanagaraj, S., Hema, M. S. & Guptha, M. N. Optimized supervised learning approach to predict Parkinson's disease with minimal attributes using PPMI Datasets. *Multimed. Tools Appl.* **83**, 48499–48520 (2023).
178. Aggarwal, N., Saini, B. S. & Gupta, S. A deep 1-D CNN learning approach with data augmentation for classification of Parkinson's disease and scans without evidence of dopamine deficit (SWEDD). *Biomed. Signal Process. Control* **91**, 106008 (2024).
179. Templeton, J. M., Poellabauer, C. & Schneider, S. Classification of Parkinson's disease and its stages using machine learning. *Sci. Rep.* **12**, 14036 (2022).
180. Makarious, M. B. et al. Multi-modality machine learning predicting Parkinson's disease, *npj*. *Parkinsons Dis.* **8**, 35 (2022).
181. Haghi Kashani, M. & Bazza Abkenar, S. Chapter 23—Applications of fog computing for smart sensor network. in *Sensor Networks for Smart Hospitals* (ed. Nguyen, T. A.) 505–523 (Elsevier, 2025).
182. Nikravan, M. & Haghi Kashani, M. Chapter 4—Smart sensor networks based on edge technologies. in *Blockchain and Digital Twin for* Smart Healthcare (ed. Nguyen, T. A.) 1–23 (Elsevier, 2025).
183. Nikravan, M., Jameii, S. M. & Kashani, M. H. An intelligent energy efficient QoS-routing scheme for WSN. *Int. J. Adv. Eng. Sci. Technol.* **8**, 121–124 (2011).
184. Rahimi, M., Al Masry, Z., Templeton, J. M., Schneider, S. & Poellabauer, C. A comprehensive multifunctional approach for measuring Parkinson's disease severity. *Appl. Clin. Inf.* **16**, 11–23 (2025).
185. Rahimi,M., Masry,Z. A., Templeton,J. M., Schneider,S. & Poellabauer,C. Beyond motor symptoms: toward a comprehensive grading of Parkinson's disease severity. in *Presented at the Proceedings of the 14th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics, Houston, TX, USA* (Association for Computing Machinery, 2023).
186. Templeton, J. et al. Modernizing the staging of Parkinson's disease using digital health technology. *J. Med. Internet Res.* **27**, e63105 (2025).
187. Heim, B., Krismer, F., De Marzi, R. & Seppi, K. Magnetic resonance imaging for the diagnosis of Parkinson's disease. *J. Neural Transm.* **124**, 915–964 (2017).
188. Zhang, H. & Babar, M. A. Systematic reviews in software engineering: an empirical investigation. *Inf. Softw. Technol.* **55**, 1341–1354 (2013).
189. Hernández-Mena, C. D. & Herrera-Camacho, J. Ciempiess: a new open-sourced mexican spanish radio corpus. in *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, 371–375 (European Language Resources Association (ELRA), 2014).
190. Orozco-Arroyave, J. R., Vargas-Bonilla, J. F., Vásquez-Correa, J. C., Castellanos-Domínguez, C. G. & Nöth, E. Automatic detection of hypernasal speech of children with cleft lip and palate from spanish vowels and words using classical measures and nonlinear analysis. *Rev. Facultad de. Ing. Univ. de. Antioquia* **80**, 109–123 (2016).
191. Orozco-Arroyave, J. R., Arias-Londoño, J. D., Vargas-Bonilla, J. F., Gonzalez-Rátiva, M. C. & Nöth, E. New Spanish speech corpus database for the analysis of people suffering from Parkinson's disease. in *LREC*, 342–347 (European Language Resources Association (ELRA), 2014).
192. Sakar, B. E. et al. Collection and analysis of a Parkinson speech dataset with multiple types of sound recordings. *IEEE J. Biomed. Health Inform.* **17**, 828–834 (2013).
193. Little, M., Mcsharry, P., Roberts, S., Costello, D. & Moroz, I. Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection. *Nat. Preced* **6**, 1–19 (2007).
194. Sakar, C. O. et al. A comparative analysis of speech signal processing algorithms for Parkinson's disease classification and the use of the tunable Q-factor wavelet transform. *Appl. Soft Comput.* **74**, 255–263 (2019).
195. Rahman, A. et al. Parkinson's disease diagnosis in cepstral domain using MFCC and dimensionality reduction with svm classifier. *Mob. Inf. Syst.* **2021**, 1–10 (2021).
196. Jaeger, H., Trivedi, D. & Stadtschnitzer, M. Mobile device voice recordings at King's College London (MDVR-KCL) from both early and advanced Parkinson's disease patients and healthy controls. *Zenodo* (2019).
197. Naranjo, L., Perez, C. J., Campos-Roca, Y. & Martin, J. Addressing voice recording replications for Parkinson's disease detection. *Expert Syst. Appl.* **46**, 286–292 (2016).
198. Li, Y. Z., Wang, X., Zhang, P. & X. Liu. Customized Subset of TIMIT with NOISEX-92 Augmentation (2021).
199. Kaggle. *Parkinson's Telemonitoring Dataset*. https://www.kaggle.com/datasets/mountainguest/parkinsons-telemonitoring (2019)
200. Little, M., McSharry, P., Hunter, E., Spielman, J. & Ramig, L. Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. *Nat. Preced* **56**, 1015–1022 (2008).
201. UCI Machine Learning Repository. *Parkinson's Disease Data Set*. https://archive.ics.uci.edu/dataset/174/parkinsons (2008).
202. Benba, A., Jilbab, A. & Hammouch, A. Discriminating between patients with Parkinson's and neurological diseases using cepstral analysis. *IEEE Trans. Neural Syst. Rehabil. Eng.* **24**, 1100–1108 (2016).
203. S. U. Institute of Phonetics, Saarbrücker Stimmdatenbank (Saarbrücken Voice Database).
204. Venegas, D. A. R. *Dataset of vowels*. https://www.kaggle.com/datasets/darubiano57/dataset-of-vowels (2018).
205. Tsanas, A. & Tsanas, A. eds. *UCI Machine Learning Repository*. https://archive.ics.uci.edu, (2014).
206. Khaskhoussy,R. & Ayed,Y. B. Detecting Parkinson's disease according to gender using speech signals. in *14th International Conference on Knowledge Science, Engineering and Management (KSEM2021)*, 414–425 (Springer International Publishing, 2021).
207. Xiao F. *Parkinson's Disease Diagnosis Dataset*. https://www.datafountain.cn/datasets/146 (DataFountain, 2020).

208. Giuliano, M. et al. Construction of a parkinson's voice database. in *International Conference on Industrial Engineering and Operations Management*, 940 (IEOM Society International Sao Paulo, 2021).
209. Rusz, J. et al. Imprecise vowel articulation as a potential early marker of Parkinson's disease: effect of speaking task. *J. Acoust. Soc. Am.* **134**, 2171–2181 (2013).
210. Dimauro, G., Di Nicola, V., Bevilacqua, V., Caivano, D. & Girardi, F. Assessment of speech intelligibility in Parkinson's disease using a speech-to-text system. *IEEE Access* **5**, 22199–22208 (2017).
211. Viswanathan, R. et al. Efficiency of voice features based on consonant for detection of Parkinson's disease. in *2018 IEEE Life Sciences Conference (LSC)*, 49–52 (IEEE, 2018).
212. Martínez, D., Lleida, E., Ortega, A., Miguel, A. & Villalba, J. Voice pathology detection on the Saarbrücken voice database with calibration and fusion of scores using multifocal toolkit. in *Advances in Speech and Language Technologies for Iberian Languages: IberSPEECH 2012 Conference*, *Madrid, Spain*, *November 21-23*, *2012. Proceedings* 99–109 (Springer, 2012).
213. Pützer, M. & Koreman, J. A German database of patterns of pathological vocal fold vibration. *Phonus* **3**, 143–153 (1997).
214. Khoshnevis, S. A. & Sankar, R. Diagnosis of Parkinson's disease using higher order statistical analysis of alpha and beta rhythms. *Biomed. Signal Process. Control* **77**, 103743 (2022).
215. Prior,F. et al. Voice samples for patients with Parkinson's disease and healthy controls. figshare, https://figshare.com/articles/dataset/Voice_Samples_for_Patients_with_Parkinson_s_Disease_and_Healthy_Controls/23849127 (2023).
216. Hughes, A. J., Daniel, S. E., Kilford, L. & Lees, A. J. Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases. *J. Neurol. Neurosurg. Psychiatry* **55**, 181 (1992).
217. Hughes, A. J., Ben-Shlomo, Y., Daniel, S. E. & Lees, A. J. What features improve the accuracy of clinical diagnosis in Parkinson's disease: a clinicopathologic study. *Neurology* **42**, 1142–1142 (1992).
218. Betts, M. J., Cardenas-Blanco, A., Kanowski, M., Jessen, F. & Düzel, E. In vivo MRI assessment of the human locus coeruleus along its rostrocaudal extent in young and older adults. *Neuroimage* **163**, 150–159 (2017).
219. Dünnwald, M., Betts, M. J., Düzel, E. & Oeltze-Jafra, S. Localization of the locus coeruleus in MRI via coordinate regression. in *Bildverarbeitung für die Medizin 2021: Proceedings, German Workshop on Medical Image Computing, Regensburg, March 7-9, 2021* 10–15 (Springer, 2021).
220. P. s. P. M. Initiative. Parkinson's Progression Markers Initiative (PPMI). (2011).
221. Marek, K. et al. The Parkinson progression marker initiative (PPMI). *Prog. Neurobiol.* **95**, 629–635 (2011).
222. Gorgolewski, K., Esteban, O., Schaefer, G., Wandell, B. & Poldrack, R. *OpenNeuro—A Free Online Platform for Sharing and Analysis of Neuroimaging Data* (Organization for Human Brain Mapping, 2017).
223. Laboratory of Neuro Imaging (LONI). *Image and Data Archive (IDA)*. https://ida.loni.usc.edu/login.jsp (2025).
224. Duchesne, S. et al. The Canadian Dementia Imaging Protocol: Harmonizing National Cohorts. *J. Magn. Reson. Imaging* **49**, 456–465 (2019).
225. Acharya, H. J., Bouchard, T. P., Emery, D. J. & Camicioli, R. M. Axial signs and magnetic resonance imaging correlates in Parkinson's disease. *Can. J. Neurol. Sci.* **34**, 56–61 (2007).
226. Lang, S. et al. Network basis of the dysexecutive and posterior cortical cognitive profiles in Parkinson's disease. *Mov. Disord.* **34**, 893–902 (2019).
227. Hanganu, A. et al. Mild cognitive impairment is linked with faster rate of cortical thinning in patients with Parkinson's disease longitudinally. *Brain* **137**, 1120–1129 (2014).
228. T. N. M. N. Institute-Hospital. *C-BIG Repository*. https://www.mcgill.ca/neuro/research/c-big-repository.
229. Badea, L., Onu, M., Wu, T., Roceanu, A. & Bajenaru, O. Exploring the reproducibility of functional connectivity alterations in Parkinson's disease. *PLoS ONE* **12**, 1–21 (2017).
230. Yoneyama, N. et al. Severe hyposmia and aberrant functional connectivity in cognitively normal Parkinson's disease. *PLoS ONE* **13**, e0190072 (2018).
231. Markiewicz, C. J. et al. ds000245: UCLA Consortium for Neuropsychiatric Phenomics LA5c Study. https://openneuro.org/datasets/ds000245/versions/00001 (2016).
232. Boelmans, K. et al. Brain iron deposition fingerprints in Parkinson's disease and progressive supranuclear palsy. *Mov. Disord.* **27**, 421–427 (2012).
233. LaMontagne, P. J. et al. IC-P-164: OASIS-3: longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and Alzheimer's disease. *Alzheimers Dement* **14**, P138–P138 (2018).
234. Wei, D. et al. Structural and functional MRI from a cross-sectional Southwest University Adult lifespan Dataset (SALD). *bioRxiv* https://doi.org/10.1101/177279 (2017).
235. Pereira, C. R., Weber, S. A., Hook, C., Rosa, G. H. & Papa, J. P. Deep learning-aided Parkinson's disease diagnosis from handwritten dynamics. in *2016 29th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, 340–346 (IEEE, 2016).
236. ul Haq, A. et al. A survey of deep learning techniques based Parkinson's disease recognition methods employing clinical data. *Expert Syst. Appl.* **208**, 118045 (2022).
237. Sheriff, I. Parkinson's Brain MRI Dataset. https://www.kaggle.com/datasets/irfansheriff/parkinsons-brain-mri-dataset (2022).
238. Sudlow, C. et al. UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med* **12**, e1001779 (2015).
239. Frenkel-Toledo, S. et al. Treadmill walking as an external pacemaker to improve gait rhythm and stability in Parkinson's disease. *Mov. Disord.* **20**, 1109–1114 (2005).
240. Drotár, P. et al. Decision support framework for Parkinson's disease based on novel handwriting markers. *IEEE Trans. Neural Syst. Rehabil. Eng.* **23**, 508–516 (2014).
241. Hausdorff, J. M. et al. Rhythmic auditory stimulation modulates gait variability in Parkinson's disease. *Eur. J. Neurosci.* **26**, 2369–2375 (2007).
242. Yogev, G. et al. Dual tasking, gait rhythmicity, and Parkinson's disease: which aspects of gait are attention demanding? *Eur. J. Neurosci.* **22**, 1248–1256 (2005).
243. Tsanas, A., Little, M., McSharry, P. & Ramig, L. Accurate telemonitoring of Parkinson's disease progression by non-invasive speech tests. *IEEE Trans. Biomed. Eng.* **57**, 884–893 (2010).
244. Bachlin, M. et al. Wearable assistant for Parkinson's disease patients with the freezing of gait symptom. *IEEE Trans. Inf. Technol. Biomed.* **14**, 436–446 (2009).
245. Alexander Koelewijn, D. H. & van den Bogert, A. J. Metabolic cost calculations of gait using musculoskeletal energy models: a comparison study. https://www.mad.tf.fau.de/research/datasets/#collapse_11 (2019).
246. Rodríguez-Martín, D. et al. Home detection of freezing of gait using support vector machines through a single waist-worn triaxial accelerometer. *PLoS ONE* **12**, e0171764 (2017).
247. Borzì, L. et al. Prediction of freezing of gait in Parkinson's disease using wearables and machine learning. *Sensors* **21**, 614 (2021).
248. Borzì, L. et al. Home monitoring of motor fluctuations in Parkinson's disease patients. *J. Reliab. Intell. Environ.* **5**, 145–162 (2019).
249. Borzì, L. et al. A new index to assess turning quality and postural stability in patients with Parkinson's disease. *Biomed. Signal Process. Control* **62**, 102059 (2020).

250. Goldberger, A. L. et al. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *Circulation* **101**, e215–e220 (2000).
251. Goldberger, A. L. et al. Gait in Parkinson's Disease Database (version 1.0.0). https://physionet.org/content/gaitpdb/1.0.0/ (2000).
252. Drotár, P. et al. Evaluation of handwriting kinematics and pressure for differential diagnosis of Parkinson's disease. *Artif. Intell. Med.* **67**, 39–46 (2016).
253. F. Machine Learning and Data Analytics Lab (MAD Lab). Research Datasets—Machine Learning and Data Analytics Lab. https://www.mad.tf.fau.de/research/datasets/.
254. Giancardo, L. et al. Computer keyboard interaction as an indicator of early Parkinson's disease. *Sci. Rep.* **6**, 34468 (2016).
255. Dhir, N., Edman, M., Sanchez Ferro, A., Stafford, T. & Bannard, C. Identifying robust markers of Parkinson's disease in typing behaviour using a CNN-LSTM network. (2020).
256. Mirelman, A. et al. Addition of a non-immersive virtual reality component to treadmill training to reduce fall risk in older adults (V-TIME): a randomised controlled trial. *Lancet* **388**, 1170–1182 (2016).
257. Mirelman, A. et al. Tossing and turning in bed: nocturnal movements in Parkinson's disease. *Mov. Disord.* **35**, 959–968 (2020).
258. Drotár, P. et al. Analysis of in-air movement in handwriting: a novel marker for Parkinson's disease. *Comput. Methods Prog. Biomed.* **117**, 405–411 (2014).
259. Zeid, S. S., ElKamar, R. A. & Hassan, S. I. Fixed-text vs. free-text keystroke dynamics for user authentication. *Eng. Res. J. Fac. Eng.* **51**, 95–104 (2022).
260. Iapa, A.-C. & Cretu, V.-I. Shared data set for free-text keystroke dynamics authentication algorithms. (2021).
261. Sun, Y., Ceker, H. & Upadhyaya, S. Shared keystroke dataset for continuous authentication. in *2016 IEEE International Workshop on Information Forensics and Security (WIFS)*, 1–6 (IEEE, 2016).
262. Lu, X., Zhang, S., Hui, P. & Lio, P. Continuous authentication by free-text keystroke based on CNN and RNN. *Comput. Secur.* **96**, 101861 (2020).
263. Hausdorff, J. M. *Gait in Parkinson's Disease Database* (PhysioNet, 2008).
264. National Center for Biotechnology Information (NCBI). GEO: Gene Expression Omnibus. https://www.ncbi.nlm.nih.gov/geo/.
265. Apweiler, R. et al. UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* **32**, D115–D119 (2017).
266. Anjum, M. F. et al. Linear predictive coding distinguishes spectral EEG features of Parkinson's disease. *Parkinsonism Relat. Disord.* **79**, 79–85 (2020).
267. Rockhill, A. P., Jackson, N., George, J., Aron, A. & Swann, N. C. UC San Diego resting state EEG data from patients with Parkinson's disease. (2020).
268. Pernet, C. R. et al. EEG-BIDS, an extension to the brain imaging data structure for electroencephalography. *Sci. Data* **6**, 103 (2019).
269. Cavanagh, J. F., Kumar, P., Mueller, A. A., Richardson, S. P. & Mueen, A. Diminished EEG habituation to novel events effectively classifies Parkinson's patients. *Clin. Neurophysiol.* **129**, 409–418 (2018).
270. U. o. N. M. P. M. Lab, Predict Dataset Repository, U. o. N. M. P. M. Lab, Ed., ed.
271. Scherzer, C. R. et al. Molecular markers of early Parkinson's disease based on gene expression in blood. *Med. Sci.* **104**, 955–960 (2007).
272. Scherzer, C. R. et al. GATA transcription factors directly regulate the Parkinson's disease-linked gene α-synuclein. *Proc. Natl Acad. Sci. USA* **105**, 10907–10912 (2008).
273. Calligaris, R. et al. Blood transcriptomics of drug-naïve sporadic Parkinson's disease patients. *BMC Genomics* **16**, 876 (2015).
274. Shamir, R. et al. Analysis of blood-based gene expression in idiopathic Parkinson disease. *Neurology* **89**, 1676–1683 (2017).
275. Locascio, J. J. et al. Association between α-synuclein blood transcripts and early, neuroimaging-supported Parkinson's disease. *Brain* **138**, 2659–2671 (2015).
276. Shehadeh, L. A. et al. SRRM2, a potential blood biomarker revealing high alternative splicing in Parkinson's disease. *PLoS ONE* **5**, e9104 (2010).
277. J. F. C. jcavanagh@unm.edu. *EEG: 3-Stim Auditory Oddball and Rest in Parkinson's* (OpenNeuro, 2021).
278. Railo, H., Nokelainen, N., Savolainen, S. & Kaasinen, V. Deficits in monitoring self-produced speech in Parkinson's disease. *Clin. Neurophysiol.* **131**, 2140–2147 (2020).
279. Open Science Framework. OSF Project Page (or actual project title if available). https://osf.io/pehj9/.
280. Cavanagh, J. F., Napolitano, A., Wu, C. & Mueen, A. The patient repository for EEG data+ computational tools (PRED+ CT). *Front. Neuroinform* **11**, 1–9 (2017).
281. Zheng, W. L., Liu, W., Lu, Y., Lu, B. L. & Cichocki, A. EmotionMeter: a multimodal framework for recognizing human emotions. *IEEE Trans. Cybern.* **49**, 1110–1122 (2019).
282. Miranda-Correa, J. A., Abadi, M. K., Sebe, N. & Patras, I. AMIGOS: a dataset for affect, personality and mood research on individuals and groups. *IEEE Trans. Affect. Comput.* **12**, 479–493 (2021).
283. Yuvaraj, R. et al. On the analysis of EEG power, frequency and asymmetry in Parkinson's disease during emotion processing. *Behav. Brain Funct.* **10**, 1–19 (2014).
284. Jackson, N., Cole, S. R., Voytek, B. & Swann, N. C. Characteristics of waveform shape in Parkinson's disease detected with scalp electroencephalography. *eNeuro* **6**, 1–11 (2019).
285. Lin, Y.-P. et al. Objective assessment of impulse control disorder in patients with Parkinson's disease using a low-cost LEGO-like EEG headset: a feasibility study. *J. Neuroeng. Rehabil.* **18**, 109 (2021).
286. Swann, N. C. et al. Elevated synchrony in P arkinson disease detected with electroencephalography. *Ann. Neurol.* **78**, 742–750 (2015).
287. OpenNeuro. Dataset ds002778 (Version 1.0.2). https://openneuro.org/datasets/ds002778/versions/1.0.2.
288. Rosenthal, L. S. et al. Parkinson's Disease Biomarkers Program. *Mov. Disord.* **31**, 915–923 (2016).
289. Seong, S. C. et al. Cohort profile: the National Health Insurance Service-National Health Screening Cohort (NHIS-HEALS) in Korea. *BMJ Open* **7**, 1–12 (2017).
290. Goldberger, A. L. et al. PhysioBank, PhysioToolkit, and PhysioNet. *Circulation* **101**, e215–e220 (2000).
291. Pereira, C. R. et al. A new computer vision-based approach to aid the diagnosis of Parkinson's disease. *Comput. Methods Prog. Biomed.* **136**, 79–88 (2016).
292. Zham, P., Kumar, D. K., Dabnichki, P., Poosapadi Arjunan, S. & Raghav, S. Distinguishing different stages of Parkinson's disease using composite index of speed and pen-pressure of sketching a spiral. *Front. Neurol.* **8**, 1–7 (2017).
293. Born, J. et al. POCOVID-Net: automatic detection of COVID-19 from a new lung ultrasound imaging dataset (POCUS). Preprint at *arXiv* https://doi.org/10.48550/arXiv.2004.12084 (2020).
294. Born, J. et al. Automatic DETECTIon of COVID-19 from ultrasound data. GitHub repository (2020).
295. Benamer, H. T. et al. Accurate differentiation of Parkinsonism and essential tremor using visual assessment of [123I]-FP-CIT SPECT imaging: the [123I]-FP-CIT study group. *Mov. Disord.* **15**, 503–510 (2000).
296. Siderowf, A. et al. PET imaging of amyloid with Florbetapir F 18 and PET imaging of dopamine degeneration with 18F-AV-133 (florbenazine) in patients with Alzheimer's disease and Lewy body disorders. *BMC Neurol.* **14**, 1–9 (2014).

Author contributions

S.S. and A.M. have equal contributions to this work: conceptualization, writing—original draft, investigation, visualization. S.B.A.:

conceptualization, review & editing, investigation. M.H.K.: supervision, project administration, investigation, validation. M.A.: validation, writing—review & editing. M.S.: validation, writing—review & editing. All authors read and approved the final manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to Mostafa Haghi Kashani.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.