



Full length article

## Artificial intelligence-enabled detection and assessment of Parkinson's disease using multimodal data: A survey

Aite Zhao <sup>a</sup>\*, Yongcan Liu <sup>a</sup>, Xinglin Yu <sup>a</sup>, Xinyue Xing <sup>a</sup>, Huiyu Zhou <sup>b</sup><sup>a</sup> College of Computer Science and Technology, Qingdao University, Qingdao, 266071, China<sup>b</sup> School of Computing and Mathematical Sciences, University of Leicester, Leicester, LE17RH, United Kingdom

## ARTICLE INFO

## Keywords:

Parkinson's disease  
Artificial intelligence  
Symptoms  
Multimodal  
Diagnosis  
Machine learning and deep learning

## ABSTRACT

Highly adaptable and reusable AI models are revolutionizing the diagnosis and management of Parkinson's disease (PD). A wide range of AI algorithms, including machine learning and deep learning techniques, are now being employed for PD diagnosis and treatment. These algorithms leverage multimodal data, such as gait patterns, hand movements, and speech characteristics, to predict the likelihood of PD, evaluate symptom severity, facilitate early detection, and assess the effectiveness of treatments, demonstrating their advanced diagnostic potential. This paper provides a comprehensive review of machine learning and deep learning approaches for PD detection and assessment over the past decade, emphasizing their strengths, addressing their limitations, and exploring their potential to inspire new research directions. Additionally, it offers a curated collection of publicly available multimodal datasets focused on PD motor symptoms and validates the performance of these algorithms on privately collected datasets. Experimental results reveal that AI technologies for PD assessment have progressed from traditional methods to convolutional and sequential neural networks, and further to Transformer-based models, achieving consistently improving accuracy and establishing benchmarks for future advancements.

## 1. Introduction

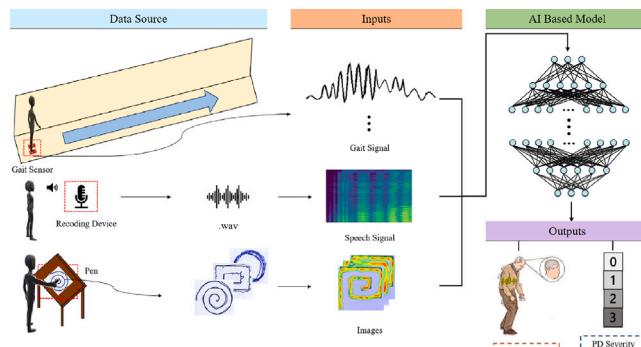
Parkinson's disease (PD) is the fastest-growing neurological disorder worldwide. As of 2020, over 1 million people in the United States are affected by PD [1], and China has approximately 3.5 million patients, a number expected to rise to 5 million by 2030 [2]. To date, no medication has been capable of reversing or preventing the disease's progression [1]. A critical challenge in PD's diagnosis, treatment, and management is the absence of an objective and efficient assessment tool [3–5]. PD is typically diagnosed through clinical symptoms, predominantly hand tremors, speech and gait disorders, and other motor impairments. However, diagnosis relies heavily on the subjective evaluations of physicians, and the symptoms fluctuate dynamically before and after treatment, necessitating time-consuming visual observations and questionnaires. Early-stage patients may exhibit subtle limb changes that are hard to detect, impeding early medical intervention and risking delayed diagnosis, which could exacerbate the condition. Consequently, there is an urgent demand for a novel paradigm in medical artificial intelligence for diagnosing and assessing PD, with the goal of streamlining the conventional processes of PD diagnosis, treatment, and rehabilitation.

Effective strategies for tracking the progression of PD over time remain elusive. Current PD symptom assessments are heavily dependent on patient self-reports or the qualitative evaluations provided by clinicians. The diagnosis often involves the use of questionnaires such as the Movement Disorder Society's Unified Parkinson's Disease Rating Scale (MDS-UPDRS) or the Hoehn-Yahr scale, both of which are semi-subjective and lack the sensitivity required to detect minor fluctuations in a patient's condition. Consequently, clinical trials for PD must span several years to ensure that changes in MDS-UPDRS scores are reported with adequate statistical confidence, a factor that escalates costs and prolongs the research timeline [6]. Thus, there is a pressing demand for an automated and objective method to assess disease progression, which could streamline clinical trials and enhance the timeliness and accuracy of PD management.

The existing literature has delved into various AI-assisted methodologies for the detection and assessment of PD [5], yet these have predominantly concentrated on singular modalities or discrete tasks within the diagnostic and therapeutic continuum of PD. This encompasses areas such as PD screening, the grading of symptom severity, and the tracking of specific biomarkers. For instance, the analysis

\* Corresponding author.

E-mail addresses: [zhaoaite@qdu.edu.cn](mailto:zhaoaite@qdu.edu.cn) (A. Zhao), [liuyongcan@qdu.edu.cn](mailto:liuyongcan@qdu.edu.cn) (Y. Liu), [yuxinglin@qdu.edu.cn](mailto:yuxinglin@qdu.edu.cn) (X. Yu), [xingxinyue@qdu.edu.cn](mailto:xingxinyue@qdu.edu.cn) (X. Xing), [hz143@leicester.ac.uk](mailto:hz143@leicester.ac.uk) (H. Zhou).



**Fig. 1.** The process of PD multi-modality and multi-symptom detection and evaluation.

and diagnosis based on PD brain imaging, the gauging of PD severity through gait force sensor data, the comprehension of PD-affected speech, and the evaluation of handwriting in PD patients have been explored. However, these studies are constrained by their reliance on a narrow range of data types and typically address only a single stage of PD's progression. Further, we have reviewed a substantial number of review articles, which include investigations into individual symptoms of PD. For instance, some studies presented the evidence that demonstrated the relationship between human gait and PD, and illustrated the role of different gait analysis systems based on vision or wearable sensors. It also provided a comprehensive overview of the available automatic recognition systems for the detection and management of PD [7,8]. Moreover, the current landscape of speech-based deep learning approaches for PD classification was explored by Gelderen et al. [9,10], none of these reviews can more intuitively reflect the changes in patients than the use of PD handwriting [11]. However, these review articles are mostly limited to data classification for a single symptom, with little involvement of new methods in dynamic multimodal learning and data fusion. The majority of these studies primarily reflect AI methods and auxiliary diagnostic strategies within the past five years. A broader temporal scope in the review would more effectively reveal the pace and trajectory of methodological evolution. In essence, the limitations of these studies conflict with the clinical practice of physicians, who typically assess patients' motor behavior based on multiple symptoms.

In this article, we provide a comprehensive review of the various artificial intelligence techniques based on machine learning (ML) and deep learning (DL) that have been applied to the diagnosis and assessment of Parkinson's disease over the past decade [12,13]. The methods encompass popular Transformer-based models, transfer learning, multimodal learning, and heterogeneous data fusion strategies, which are among the most advanced algorithms in recent years. Additionally, a vast collection of heterogeneous, multi-source Parkinson's disease datasets is provided for researchers' reference. Our review synthesizes findings from a broad spectrum of published research, shedding light on emerging trends in PD diagnostics. As illustrated in Fig. 1, the hybrid AI model is trained on various PD data through ML and DL techniques. Next, the model needs to access various sources of medical knowledge to perform medical reasoning tasks, thereby releasing rich features that can be used for downstream applications, such as severity grading for patient speech, gait and hand movement data. The subsequent sub-models then perform tasks or feedback that users can specify in real time. The article will proceed with a comprehensive overview of the integration of multi-source symptom data [14,15], highlighting the nuances and potentials of this approach in advancing PD diagnosis and assessment.

Gait abnormalities are a defining feature of motor symptoms in neurodegenerative diseases [16], with PD being no exception. PD is particularly characterized by distinctive changes in walking posture

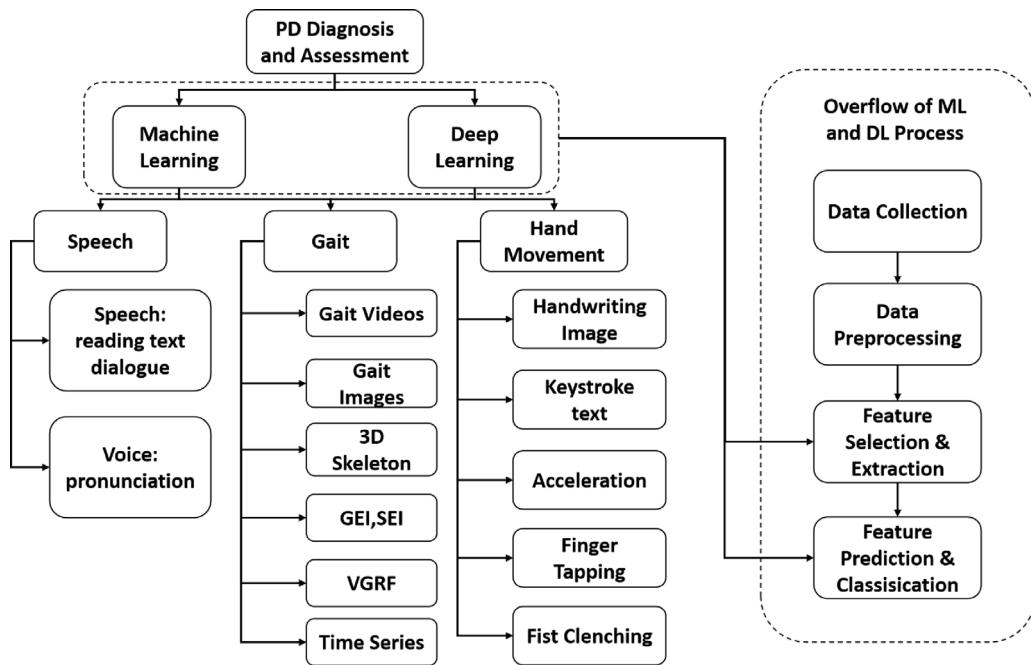
throughout the disease's progression, such as the well-known shuffling and freezing gaits. By employing machine learning techniques to analyze and decode the intricate dynamics and patterns within these abnormal gaits, we can achieve a more accurate differentiation among PD patients, taking into account the unique stages and varying levels of severity they present. The gait data of PD patients show considerable diversity, with each type of sensor capturing specific traits of movement patterns. This rich data can be harnessed to refine the diagnostic process, potentially leading to earlier and more precise identification of PD, as well as more personalized treatment strategies. To collect gait data, employing visual sensors like cameras offers a straightforward approach to capturing the comprehensive walking patterns of individuals. Subsequently, these data are meticulously analyzed and processed through sophisticated automated learning algorithms. Skeletal features serve as a pivotal reference in gait recognition, as they effectively filter out extraneous details, focusing solely on the joint coordinates of the human skeletal structure. This refined approach allows for a more precise identification of gait variations. Furthermore, data gleaned from force sensors, exemplified by the vertical ground reaction force (VGRF), offer a nuanced perspective on the subtleties of lower limb dynamics. This approach has garnered considerable research attention across various regions globally, focusing on the analysis of gait through force sensor data.

PD often affects the hands of patients, frequently resulting in a characteristically flexed posture due to muscle rigidity, which impedes the full extension of the palms. Moreover, the small joints between the fingers can become hyperextended, further complicating the action of making a fist. To tackle these pathological manifestations, our project will systematically collect data on patients' hand movements. We will analyze the early impacts of PD on the kinematic characteristics of hand and finger movements and evaluate the severity of these impairments. As widely recognized, biomarkers indicative of PD can be identified through various modes of human-computer interaction, including the assessment of grip strength, the finger tapping test (FTT), and the observation of hand and finger movements.

Speech disorders in PD present as a distinctive complication, characterized by a sudden emergence and a gradual progression of communication barriers. Patients often encounter a reduction in vocal clarity and volume, hoarseness, and in severe cases, even a complete loss of voice. Some individuals may also develop stuttering, which severely hampers their social interactions. Moreover, uncontrollable vocal tremors and indistinct articulation can render patients incomprehensible to their colleagues and family, potentially leading to emotional distress and social withdrawal. The co-occurrence of speech disorders in Parkinson's patients exacerbates their suffering, significantly diminishing their quality of life and leaving their families feeling helpless and at a loss for support. Therefore, scientific research should be conducted on language disorders in different stages of PD, using speech data from PD patients to assist doctors in early diagnosis of the disease and assess its severity.

The primary objective of this study is to provide an in-depth analysis of the diverse methodologies employing ML and DL for the diagnosis and evaluation of PD. As illustrated in Fig. 2, the classification of modalities for diagnosing and assessing PD utilizes two artificial intelligence algorithms that operate on multimodal data derived from three distinct symptoms. Following the process of data analysis and processing, feature selection and extraction are conducted to yield classification and prediction outcomes that aid in the diagnosis and treatment of PD. By meticulously examining the extensive body of published research from a multitude of scholars, this study distills deeper insights into current research trends and underscores the evolving paradigms in PD diagnosis.

This work introduces various research results and studies on the use of machine learning and deep learning techniques for diagnosing PD published in various electronic database search engines (such as IEEE, Google Scholar, Web of Science, etc.) between 2014 and 2024. In recent



**Fig. 2.** Classification of modalities for the diagnosis and assessment of Parkinson's Disease (PD).

years, it has been observed that the number of PD diagnostic articles based on deep learning techniques has steadily increased. From the literature review, it can be noted that most authors only investigated the contribution of AI methods to the diagnosis of PD. But this review compares and analyzes the experimental results of various researchers, provides algorithms for multimodal learning and analysis of multimodal datasets, and compares feature selection models, classifiers, and sensors, providing new perspectives for future work.

Hence, the significance of this comprehensive review paper are as follows:

- This article summarizes the contributions of traditional machine learning techniques and deep learning models to the analysis and processing of multimodal, multi symptom, and heterogeneous data.
- Emphasize the functionality and effectiveness of feature selection, extraction, and recognition models, as they play a crucial role in improving the accuracy of PD diagnostic assessment.
- This study introduces multi-source and multimodal datasets of speech, gait, and hand movements. Enable readers to have a more detailed understanding of the acquisition process and experimental conditions of multimodal data.
- For researchers interested in developing PD prediction models through various AI based approaches in the future, this review paper may be a valuable source.

In summary, AI-based diagnostic methods are enhancing health-care systems and emerging as a cutting-edge research focus. However, most current solutions suffer from biases and inaccuracies due to reliance on single-modality data. Additionally, limitations in processing heterogeneous, multimodal disease data restrict their scalability and interdisciplinary applications.

To address these challenges, this study focuses on integrating multimodal, heterogeneous data from Parkinson's disease (PD) patients, using gait, hand movements, and speech as entry points. The developed AI-assisted diagnostic model effectively extracts spatiotemporal features and identifies relationships and structural characteristics between data types, thereby quantifying and regularizing disease data to achieve precise diagnosis.

Currently, research on multimodal medical diagnostics remains underdeveloped globally. Leveraging vast medical data and advanced

technologies is essential to promote knowledge integration in PD diagnosis and form a comprehensive knowledge system to support clinical decision-making. This effort is crucial for achieving global leadership in the intersection of medicine and AI, and holds significant theoretical and practical value for building intelligent medical systems and refining AI-assisted diagnostic frameworks.

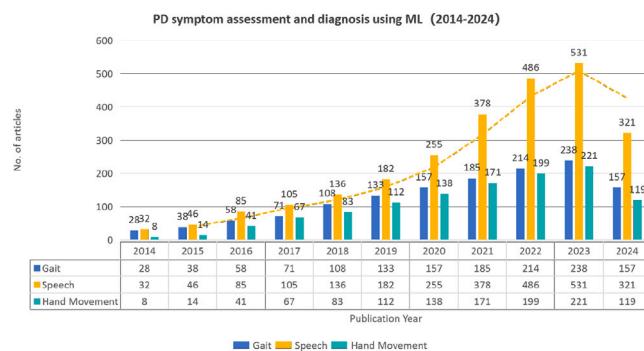
The structure of this article unfolds as follows: Initially, it presents an overview of the significance of PD, highlighting both its motor and non-motor symptoms. The subsequent section delves into the utilization of machine learning methodologies for the diagnosis and assessment of PD symptoms. Moving forward, the third segment offers an in-depth examination of various deep learning approaches employed in the diagnosis and evaluation of the condition. The Section 4 introduces and scrutinizes publicly accessible datasets pertaining to the clinical manifestations of PD. Conclusively, the final section conducts a scholarly analysis of the principal challenges and future directions within this domain.

## 2. PD diagnosis and assessment through ML-based approaches

This section evaluates the efficacy of machine learning techniques in analyzing three key clinical manifestations of Parkinson's disease, examining both singlemodal and multimodal analytical perspectives (see Fig. 3).

By comparing the three biometric features and their symptom manifestations, we can see that from 2014 to 2024, machine learning algorithms mainly focused on speech recognition, and the number of new papers added each year has consistently ranked first over the past decade. This is also related to the singularity and ease of collecting voice data. There are relatively few papers on hand movement disorders and tremor detection, as the precision and subtle changes in hand or finger movements are difficult to grasp and closely related. Gait involves the entire human body and has various sources and modes, which determines its important position in the diagnosis and treatment of Parkinson's disease.

Machine learning is a branch of artificial intelligence that enables computer systems to learn from data and make decisions or predictions without the need for explicit programming. Machine learning techniques include supervised learning, unsupervised learning, semi



**Fig. 3.** Single modal PD symptom assessment and diagnosis methods using ML (2014–2024).

supervised learning, and reinforcement learning. Each method has its specific application scenarios and advantages [17].

Machine learning can analyze the motor and non-motor symptoms of Parkinson's disease patients, such as tremors, rigidity, and bradykinesia; By analyzing patients' physiological and behavioral data, machine learning models can help doctors make more accurate diagnoses, track disease progression, and provide personalized treatment plans for patients. The application of machine learning in Parkinson's disease detection is becoming increasingly widespread. By analyzing large amounts of medical data, it improves the accuracy of diagnosis and provides better treatment and management plans for patients. With the advancement of technology, the application of machine learning in this field will become more profound and extensive in the future.

### 2.1. Single modal PD symptom analysis

Over the past decade, machine learning has emerged as a beacon of innovation in the realm of Parkinson's disease diagnosis and assessment. Especially in the processing of singlemodal normalized data, its characteristics of few parameters, high efficiency, and high accuracy make it dominant in PD data processing.

In recent years, the spotlight has been on the Support Vector Machine (SVM) as a prominent machine learning method. SVM is a versatile linear classifier that excels in binary classification tasks through the application of supervised learning. Its emphasis on precision in classification makes it a pivotal tool in the diagnosis and evaluation of Parkinson's disease, addressing the complexities of this condition with remarkable efficacy.

Firstly, SVM has made significant contributions to speech recognition and classification for PD patients. Shahbakhi et al. proposed a new algorithm for Parkinson's disease diagnosis based on biomedical speech signal analysis, using genetic algorithm to select optimized features from all extracted features. Then, SVM was used to distinguish between healthy individuals and Parkinson's disease patients [18–21]. The extraction of MFCC features from general speech data was a key feature widely used for extracting audio signals, as it can effectively represent the spectral characteristics of audio signals, especially performing well in noisy environments. SVM was also used to distinguish samples represented by MFCC features [22–24,26,34,35]. It appeared in the decision layer for feature classification, and other classifiers with the same functionality included Logistic Regression (LR), Naive Bayes (NB), k-nearest neighbors (KNN), multi-layer perceptron (MLP), Decision Tree (DT), Random Forest (RF) and XGBoost, which can process concatenated MFCC features to provide accurate classification results [24–29]. The ensemble learning algorithm [30–32,36–38] that collaborated with these classifiers using voting decisions to distinguish speech features had greater advantages. For example, an ensemble learning algorithm consisting of AdaBoost, random forest, and decision tree [32] developed on an imbalanced speech dataset selected 10 best

features and obtained 90% of the results for the optimal ensemble classifier and comprehensive evaluation. In addition to utilizing MFCCs, a variety of other acoustic features, such as Tunable Q-factor Wavelet Transform (TQWT), Linear Predictive Cepstral Coefficients (LPCC) were concurrently employed as inputs for classifiers, which were designed to discern the progression and variations in vocal impairments across various types of Parkinson's Disease (PD) patients [33,39–41].

As demonstrated in [Table 1](#), 16 literature on PD speech recognition were investigated, including raw data on voice and speech. The maximum number of participants is 188 PD patients and 64 healthy subjects, collected under different experimental conditions. Most of these data were preprocessed to extract features such as Mel-frequency cepstral coefficients (MFCC), Discrete Wavelet Transform (DWT), Tunable Q-Factor Wavelet Transform (TQWT), Linear Prediction cepstral coefficients (LPCC), Linear Predictive Coding (LPC), weighted linear prediction cepstral coefficients (WLPCC), etc., and then fed into support vector machine (SVM), random forest (RF), naive Bayes (NB), k-nearest neighbor (KNN), multi-layer perceptron (MLP), Adaboost, logistic regression (LR), decision tree (DT), extreme learning machine (ELM), and ensemble learning for classification. The majority of the predicted results obtained exceeded 90%. MFCC is a popular feature for speech processing due to its human-like sensitivity to frequency, key information extraction like pitch and timbre, and robustness to speaker volume and noise. It also reduces data dimensions while preserving essential details, enhancing algorithm efficiency and reducing computational load. These traits make MFCC a preferred choice for tasks like speech recognition and synthesis. Meanwhile, [Table 1](#) provides a comprehensive overview of feature utilization patterns across different studies, explicitly indicating whether multiple features were employed in each investigation while concurrently classifying the methodologies as either ensemble or non-ensemble approaches. Our analysis reveals that ensemble learning architectures coupled with multiple feature combinations demonstrate superior performance in Parkinson's disease detection through speech analysis, achieving statistically significant improvements in diagnostic accuracy compared to single-feature or non-ensemble approaches.

Furthermore, these methods had not only performed well in the field of PD speech recognition, but they also exhibited stable performance in the realm of PD gait recognition. Although PD gait data were multi-sensor and multimodal, mostly from conditional laboratory acquisition, a lot of work had been focused on a single modality. Gait data generally included three-dimensional joint points of skeleton, ground reflection force, and image data [42–44]. Most machine learning classifiers had determined the severity of PD or detected PD in a population by learning gait patterns or data characteristics [45–49]. An automatic gait classification system [49] had been proposed for identifying the severity level of PD, using four supervised machine learning classifier algorithms, namely, DT, SVM, ensemble classifier (EC), and Bayes classifier (BC), obtaining an accuracy of 99.4% on the GRF gait dataset. Machine learning strategies had also been employed to classify PD gait and healthy controls [45] based on GRF data, using RF, SVM, KNN, and kernel Fisher discriminant. ML algorithms had also been very sensitive to gait skeleton data [42,50,51]. Based on 3D-depth skeleton data, Kim et al. [42] had built a KNN classifier and an SVM classifier to classify abnormal gait from a walking person. Chavez et al. [50] had presented the utilization of KNN, SVM, and gradient boosting (GB) classifiers in distinguishing well-established gait features from the skeleton. For the RGB gait images and videos, machine learning approaches could capture changes of each frame and directly classify the PD feature [44,52,53]. The SVM had classified the levels with a promising area under the ROC of 80.88% on a video dataset of PD gait [44]. DT, NB, and RF had worked together on a PD RGB-D camera dataset collected using Kinect [53]. There had also been some machine learning methods applied to detect freezing of gait (FOG) and analyze gait energy images [54–57].

**Table 1**

Summarizes the diagnosis and assessment of PD based on speech data using machine learning techniques.

References	Dataset	No. of subjects	Feature	Method		Accuracy	Sensitivity	Specificity
				Multiple	Ensemble			
[18]	Speech	23-PD, 8-HC	Fundamental, frequency or pitch	✓	SVM	✗	94.50%	*
[19]	Voice	17-PD, 17-HC	MFCC	✗	SVM	✗	91.17%	*
[20]	Voice	17-PD, 17-HC	MFCC	✗	SVM(RBF)	✗	91.18%	*
[21]	Speech	20-PD, 18-HC	MFCC(DWT)	✗	Linear SVM	✗	72%	80%
[22]	Voice	188-PD, 64-HC	TQWT and MFCC	✓	KNN, SVM, RF, MLP	✗	94.7%	98.4%
[23]	Speech	20-PD, 14-HC	DWT, MFCC	✓	SVM	✗	91.18%	*
[24]	Speech	60-PD, 100-HC	MFCC	✗	LDA-SVM	✗	88%	73.33%
[25]	Voice	23-PD, 8-HC	Frequency, Jitter and Shimmer	✓	NB, LR, KNN, RF	✗	90.2%	*
[26]	Voice	188-PD, 64-HC	MFCC and Long-term feature	✓	RF	✓	88.84%	*
[27]	Voice	20-PD, 20-HC	Frequency, Jitter and Shimmer	✓	KNN	✗	94.5%	*
[28]	Voice	188-PD, 64-HC	MFCC, Wavelet and TQWT	✓	Decision forests	✓	94.12%	94%
[29]	Voice	40-PD, 40-HC	Acoustic feature	✗	SVM, LR, Discriminant Analysis, KNN, DT, RF, Bagging tree, NB, and AdaBoost	✓	97.22%	94.44%
[30]	Speech	20-PD, 20-HC	Acoustic feature	✗	Ensemble learning	✓	88%	97%
[31]	Speech	20-PD, 20-HC	Manual feature	✗	ELM+SVM	✗	97.50%	*
[32]	Speech	188-PD, 64-HC	753 features	✓	AdaBoost, RF, and DT	✓	95.20%	98.30%
[33]	Speech	20-PD, 10-HC	MFCC, LPCC, LPC, WLPCC	✓	KNN, PNN	✗	92.63%	*

**Abbreviations:** PD — Parkinson's disease; HC — healthy controls; SVM — Support Vector Machine; RBF — Radial Basis Function; MFCC — Mel-frequency cepstral coefficients; DWT — Discrete Wavelet Transform; TQWT — Tunable Q-Factor Wavelet Transform; KNN — K-nearest neighbors; LR — Logistic Regression; RF — Random Forest; MLP — Multi-layer Perceptron; LDA — Linear Discriminant Analysis; SVM(RBF) — Support Vector Machine with Radial Basis Function kernel; DT — Decision Tree; NB — Naive Bayes; AdaBoost — Adaptive Boosting; ELM — Extreme Learning Machine; PNN — Probabilistic Neural Network; LPCC — Linear Predictive Cepstral Coefficients; LPC — Linear Predictive Coding; WLPCC — Weighted Linear Predictive Cepstral Coefficients; DWT — Discrete Wavelet Transform.

**Table 2**

Summarizes the diagnosis and assessment of PD based on gait data using machine learning techniques.

References	Dataset	No. of subjects	Feature	Method		Accuracy	Sensitivity	Specificity
				Multiple	Ensemble			
[42]	Gait	*	Integrated gait features	✓	KNN, SVM	✗	96.52%	*
[43]	Gait	93-PD, 73-HC	VGRF, CNN, LSTM, temporal features	✓	LSTM, CNN	✓	98.88%	*
[44]	Gait	19-PD	Silhouette, and temporal features	✓	SVM	✗	70.83%	*
[45]	Gait	23-PD, 26-HC	Spatial-temporal gait features	✓	KFD, BA, KNN, SVM, RF	✓	92.6%	96%
[46]	Gait	15-PD, 16-HC, 20-HD, 13-ALS	Stride, swing and stance	✓	SVM(RBF)	✗	83.33%	85.71%
[47]	Gait	93-PD, 73-HC	Gait, tremor, kinetic, and temporal features	✓	SVM	✗	92.7%	*
[48]	Gait	29-PD, 18-HC	VGRF, COP, load distribution	✓	LDA, QDA	✗	95%	*
[49]	Gait	93-PD, 73-HC	Temporal and spatial features	✓	DT, SVM, EC	✗	*	*
[50]	Gait	*	Stride, swing and double support	✓	SVM, KNN, GB	✓	99%	97%
[51]	Gait	7-HC	Gait cycle, joint angles	✓	SVM, KNN, LSTM, CNN	✗	94.9%	*
[52]	Gait	10 mix	GEI, SEI, joint angles, CNN features	✓	VGG-19 CNN, LDA, SVM	✗	98.5%	*
[53]	Gait	30-PD, 30-HC	Gait spatiotemporal features, swing magnitude, asymmetry	✓	RF, SVM, DT, NB, LR	✓	84.5%	*
[54]	Gait	93-PD, 73-HC	Time domain, non-linear and frequency domain features	✓	MLP, RF, XGB, SVML, KNN, NB, SVMP	✓	91%	89%
[55]	Gait	30 mix	Gait energy images	✗	OSVM	✗	97.33%	98.25%
[56]	Gait	81-PD	Temporal and spectral features	✓	LR, RF, CNN	✓	*	96%
[57]	Gait	10-PD	Time and frequency domain features	✓	RF, NN, SVM(RBF), XGBoost	✓	79.85%	*

**Abbreviations:** PD — Parkinson's disease; HC — Healthy controls; VGRF — Vertical Ground Reaction Force; KFD — Kernel Fisher Discriminant; BA — Bayesian Algorithm; RF — Random Forest; KNN — K-nearest neighbors; DT — Decision Tree; EC — Ensemble Classifier; LDA — Linear Discriminant Analysis; QDA — Quadratic Discriminant Analysis; GB — Gradient Boosting; GEI — Gait Energy Images; SEI — Spatio-Temporal Energy Images; OSVM — One-Class Support Vector Machine; XGBoost — Extreme Gradient Boosting.

**Table 2** recaps the diagnosis and assessment of PD using machine learning techniques to analyze gait data. For gait data, most of the studies choose a variety of features in feature selection, so that the information used by the model is updated comprehensively. The table highlights in bold the algorithms that achieved the highest accuracy rate among all the studies listed. It reflects the critical role of model's

architecture selection in improving diagnostic performance. For example, studies using spatial-temporal features are able to capture subtle variations in gait, and deep learning classifiers such as CNN and LSTM can process these complex features to achieve high diagnostic accuracy. In the studies referenced as [47,49,54], the use of extensive datasets has allowed for the application of diverse algorithms, including SVM, KNN, DT, RF, MLP, and ensemble methods like XGBoost. In general,

**Table 3**

Summarizes the diagnosis and assessment of PD based on hand movement data using machine learning techniques.

References	Dataset	No. of subjects	Feature	Method		Accuracy	Sensitivity	Specificity
				Multiple	Ensemble			
[11]	Voice, handwriting	*	Voice dataset, handwriting dataset features	✓	SVM, RF, KNN, CNN	✓	95%	*
[58]	Handwriting	37-PD,38-HC	Kinematic, Spatio-Temporal features	✓	KNN, SVM(RBF), NB, LDA, RF, ADA	✓	73.38%	68.87%
[59]	Handwriting	37-PD,18-HC	Visual-Based features, MRT, statistical features	✓	NB, PF, SVM(RBF)	✗	78.9%	*
[60]	Handwriting	*	HOG	✗	RF	✓	86.67%	*
[61]	Handwriting	24-PD, 20-HC	NNV, SDV, MV	✓	NB, AdaBoost, LR, SVN, RF	✓	91%	88%
[62]	Handwriting	55-PD,94-HC	Neuromotor, kinematic, nonlinear dynamic	✓	KNN, MLP, SVM	✗	97%	*
[63]	Handwriting	66-PD,24-HC	Movement Amplitude, Speed, Amplitude Decay	✓	LR	✗	86%	*
[64]	Finger tapping and postural stability	579-PD	Tapping rate, times, amplitude variation, body straightness	✓	DNN	✗	88%	*
[65]	Finger tapping	50-PD	Demographic factors and kinematic features	✓	DT	✗	80%	*

**Abbreviations:** HOG — Histogram of Oriented Gradients; RF — Random Forest; NNV — Normalized Velocity Variability ; SDV — Standard Deviation of Velocity; MV — Mean Velocity; KNN — K-nearest neighbors; MLP — Multi-layer Perceptron; SVM — Support Vector Machine; CNN — Convolutional Neural Network; RBF — Radial Basis Function; LDA — Linear Discriminant Analysis; ADA — Adaptive Discriminant Analysis; NB — Naive Bayes; PF — Probability Forest; DNN — Deep Neural Network; DT — Decision Tree.

the model based on ensemble learning is better than the non-ensemble model, and the performance of these algorithms on larger datasets is especially noteworthy, as it indicates their potential for real-world applicability and reliability in diagnosing PD.

According to the UPDRS, hand motor symptoms generally occurred in the early stages of Parkinson's disease, characterized by static tremors and slow, clumsy movements. The severity of these symptoms was determined by the speed, amplitude, and rhythm of finger tapping, hand opening and closing, and hand spinning movements, or changes before and after rehabilitation training. Hand tremors directly affected the stability of handwriting [11]. Machine learning algorithms learned the movement trajectory and handwriting of hand movements to distinguish between PD patients of different severity levels or to conduct early screening for potential PD [58–62]. A random forest classifier [60] was employed for the detection of Parkinson's disease among patients, using two types of handwriting analysis, i.e., spiral and wave drawings. The RF, NB, boosted tree, logistic regression, and SVM classified Normalized Velocity Variability (NNV), the velocity's Standard Deviation (SDV), and Mean (MV), and the signal Entropy (ETP), of the PD line-drawing; NB had been verified to achieve the highest accuracy [61]. In addition to handwriting, finger-tapping and fist-clenching exercises were also important references for evaluating PD during the rehabilitation stage [63–69]. For example, SVM was used on PD video footages for UPDRS motor examination of finger-tapping (UPDRS-FT), which could evaluate the severity of PD [66]. XGBoost was proposed to distinguish Parkinsonian rest tremor and fist-clenching by analyzing brain activity, providing better results than other linear models [69].

Table 3 summarizes most works for the diagnosis and assessment of PD based on hand movement data using machine learning techniques. The table highlights in bold the highest reported accuracy achieved within the listed studies. A distinctive aspect of the table is the variety in the combination of feature extraction methods and classifiers. For example, the study [58] uses kinematic and spatio-temporal features with a range of classifiers, including KNN, SVM with RBF kernel, NB, LDA, RF, and ADA, achieving an accuracy of 73.38%. This demonstrates the potential of combining specific features with multiple classifiers to enhance diagnostic accuracy. The study [64] involved 579 PD patients and used features such as tapping rate, timing, amplitude variation, and body straightness with a DNN classifier, achieving an accuracy of 88%. In general, most studies using hand data to diagnose Parkinson's disease

have selected hand motion sensor data, and the performance is better than using only handwritten image features. Some studies [11,61] further improve the accuracy of diagnosis by ensemble learning.

## 2.2. Multimodal PD symptom analysis

Machine learning algorithms were not only suitable for single mode and single symptom but also suitable for multimodal data and multi-symptom analysis, as long as different modes were mapped in the same feature space [70–76]. Li et al. introduced a multi-modal data image encoding and fusion approach for diagnosing depression in PD patients, using a multi-modal dataset encompassing motion, facial expression, and audio data [71]. Fisher feature classifiers and deep canonical correlation analysis (DCCA) were applied on a multimodal PD gait dataset, fusing VGRF signal and temporal data for PD detection [77]. In the study of [72], a correlation analysis method was used to detect the relationship between brain regions and genes, using RF and DT for classification on functional magnetic resonance imaging (fMRI) and single nucleotide polymorphisms (SNPs). Another study [78] compared Random Forest, Decision Tree, Logistic Regression, Gradient Boosting, Support Vector Machine, Stacking, and Bagging Ensemble classifiers to two standard benchmark datasets of PD. However, these methods only considered a single symptom or task of Parkinson's disease and did not provide diagnosis and treatment plans based on multimodal data of multiple symptoms like a professional doctor [79]. In this study, Huo et al. [80] used machine learning algorithms to quantitatively evaluate bradykinesia, stiffness, and tremor in PD patients and tested their ability to predict the UPDRS score on a system embedded with a force-sensor, three inertial measurement units, and four custom mechanomyography sensors.

Although machine learning methods have emerged as essential tools for decision-making in various Parkinson's disease (PD) detection tasks, their limitations lie in analyzing and interpreting similar data, as well as distinguishing various salient features. The extraction and description of these salient features depend on the learning capabilities of deep neural networks from different perspectives.

## 3. PD diagnosis and assessment through DL-based approaches

As illustrated in Fig. 4, deep learning technology rapidly developed between 2015 and 2020 and required sufficient training data. It is

**Table 4**

Summarizes the diagnosis and assessment of PD based on speech data using deep learning techniques.

References	Dataset	No. of subjects	Feature	Method			
					Manual	Temporal	Accuracy
[81]	Speech	*	Phonological features, MFCC	✓	RNN, GRUs	✓	76%
[82]	Speech	70-PD, 70-HC	XLSR-PD	✓	Transformer	✓	87.1%
[83]	Speech	52-PD, 56-HC	*	✗	LSTM-based siamese network	✓	96.2%
[84]	Voice	188-PD, 64-HC	TQWT, MFCC and Concat	✓	CNN	✗	86.9%
[85]	Speech	50-PD, 50-HC	STFT, CWT	✓	CNN	✗	89%
[86]	Voice	500-PD, 500-HC	Joint Time-Frequency Analysis algorithm	✓	CNN	✗	90.45%
[87]	Voice	28-PD, 20-HC	MFCC	✓	TmmNet	✗	99.91%
[88]	Voice	104-PD, 77-HC	Mel-spectrograms	✓	Vision Transformer	✗	75.96%
[89]	Voice	188-PD, 64-HC	753 vocal features	✓	XGBoost & Transformer & MLP	✓	*
[90]	Speech	34-PD, 34-HC	128 MFCCs	✓	CNN, LSTM, E2E	✓	94.5%
[91]	Speech	55 subjects	*	✗	Swin transformer & BiLSTM-GRU	✓	97.64%
[92]	Speech	100-PD, 100-HC	128-d log Mel filterbank features	✓	EfficientNet & AST & PSLA	✗	92.15%
[93]	Speech	28-PD, 22-HC	Wav2vec	✓	Ensemble random forest	✗	95%
							97%
							*

**Abbreviations:** XLSR-PD — Cross-lingual Speech Representation for Parkinson's Disease; TQWT — Time-Frequency Weighted Transform; MFCC — Mel-Frequency Cepstral Coefficients; STFT — Short-Time Fourier Transform; CWT — Continuous Wavelet Transform; AST — Attention-based Speech Transformer; PSLA — Parallel Speech and Language Analysis; E2E — End-to-End; BiLSTM — Bidirectional Long Short-Term Memory; Wav2vec — Waveform-to-vector.

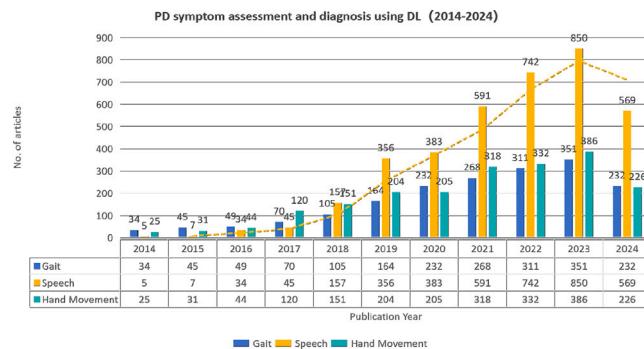


Fig. 4. Single modal PD symptom assessment and diagnosis methods using DL (2014–2024).

noteworthy that few deep learning models for speech recognition were developed in 2014 and 2015. LSTM entered a high-speed development stage from 2014 to 2015 and was applied in various time series fields. Speech data, being temporal in nature, are well-suited for LSTM analysis. The development of algorithms such as NLP led to a steady increase in the number of related papers, which reached a peak in 2023. The number of studies on the other two symptoms is also high, as most hand movement and gait data are collected using visual sensors and presented in the form of images and videos. Convolution-based deep learning algorithms can directly extract spatial features from two-dimensional and three-dimensional data, which contain more structural information and thus require more complex algorithms for processing. The development of temporal networks, convolutional networks, and attention mechanisms not only laid the foundation for models like Transformers but also fueled the growth of deep learning algorithms, leading to a higher average number of publications than machine learning algorithms in the 2020s.

### 3.1. Single modal learning for PD symptom

This section investigates the impact of gait, hand movements, and speech performance based on single-data (from multi-sensor sources) on the diagnosis and treatment of Parkinson's disease.

In recent years, a large number of deep models extracted MFCC features from speech disorder data, driving in-depth research on PD [82, 94]. Between 2014 and 2019, due to the popularity of CNNs (Convolutional Neural Networks) and LSTMs (Long Short-Term Memory networks) and their widespread application in other pattern recognition

fields, PD speech, characterized by its time-domain and frequency-domain features, could be served as input for these models to extract temporal and spatial features [83–85]. These two were also combined into a fusion model to extract spatiotemporal features of PD speech simultaneously [86]. The LSTM variant, GRU [81], was also employed to assess speech impairments by computing static features from a complete utterance. After 2020, some complex models and Transformer-based learning frameworks gradually emerged, and the strategy of incorporating attention mechanisms made PD speech recognition more accurate [9, 87–90, 95, 96]. Some transformer-based hybrid models also emerged, which performed better than single models. For example, the use of BiLSTM-GRU fusion with Swin transformer achieved a 97.64% classification effect on spoken language of patients with pronunciation disorders [91]. The triple multimodal network [87] that simultaneously integrated LSTM, GRU, CNN, and transformers represented a leap forward in multi-model fusion for PD speech signal processing. It was applicable to multiple PD speech datasets and, for the first time, comprehensively considered the spatiotemporal features, salient regions, and time-frequency domain features of PD speech. Moreover, there were also some DL-based PD speech assessment systems that had been put into PD self-test and rehabilitation processes [82, 92, 97]. Abner et al. [82] explored the usefulness of using Wav2Vec self-supervised speech representation as the speech feature of dysarthria in training ASR systems and used a transformer-based context network for feature representation and classification. Due to the limited amount of PD speech data, self-supervised and unsupervised models flourished [93, 95, 97, 98]. For instance, Mian et al. proposed an unsupervised autoencoder for PD speech feature selection and passed the compressed features to supervised machine-learning (ML) algorithms. From then on, pre-trained models gradually became the basis for PD speech feature extraction and parameter fine-tuning.

Table 4 summarizes the diagnosis and assessment of PD based on speech data using deep learning techniques. It presents an overview of the dataset, the extraction feature (handcrafted or non-handcrafted), the classifier model (temporal or non-temporal), and performance metrics – including accuracy, sensitivity, and specificity – for each referenced study. Bold text indicates the best-performing model in that study, and the performance measures are for that model. Based on the data presented in the table, it can be inferred that attention-based models (e.g., Transformer) and LSTM-based models generally outperform traditional models (e.g., CNN, RNN) across a range of speech datasets. Most achieved more than 92% accuracy and more than 90% sensitivity. For the diagnosis and evaluation of PD on speech datasets, researchers seem more inclined to apply deep learning technology in the classification stage, while for feature extraction, common features such as

**Table 5**

Summarizes the diagnosis and assessment of PD based on gait data using deep learning techniques.

References	Dataset	No. of subjects	Feature	Method		Accuracy	Sensitivity	Specificity
				Manual	Temporal			
[99]	Gait	15-PD, 16-HC	*	X	Dual Channel LSTM	✓	97.33%	*
[100]	Gait	9 subjects	*	X	AGS-GCN	X	100%	*
[101]	Gait	26-PD, 23-HC	Analysis of gait features from video	✓	DT	X	93.75%	*
[102]	Gait	93-PD, 73-HC	*	X	1D-Convnets, FC	X	98.7%	97.8% 99.2%
[103]	Gait	18 subjects	*	X	CNN and LSTM	X	93.7%	*
[104]	Gait	15-PD, 16-HC	*	X	CorrMNN, Multi-Switch Discriminator	X	99.86%	*
[105]	Gait	93-PD, 73-HC	*	X	LSTM	✓	98.6%	96.20% 98.08%
[106]	Gait	93-PD, 73-HC	*	X	CNN, CNN-LSTM	X	95%	*
[107]	Gait	93-PD, 73-HC	*	X	TF, FF, Uni-&Bi-directional LSTM	✓	99.19%	99.12% 99.72%
[108]	Gait	*	TF, Spectrogram images	✓	BiLSTM, CNNs (AlexNet, GoogleLeNet, VGG16)	X	99.91%	99.93% 99.97%
[109]	Gait	3996-PD, 12216-HC	*	X	FuseLGNet	✓	99.78%	*
[110]	Gait	93-PD, 73-HC	*	X	Transformer	✓	95.2%	98.1% 86.8%
[111]	Gait	43 subjects	*	X	Multi-Speed Transformer	✓	96.9%	96.9% 97.1%
[112]	Gait	*	*	X	Transformer, GAU	✓	97.4%	*
[113]	Gait	93-PD, 73-HC	*	X	ConvNet-Transformer	✓	97%	98.7% 86.1%
[114]	Gait	52 subjects	*	X	Multi-Level Fine-Tuned Transformer	✓	99.83%	*
[115]	Gait	93-PD, 73-HC	*	X	Transformer	✓	98.86%	98.86% *

**Abbreviations:** AGS-GCN — Attention Enhanced Gait-Structural Graph Convolutional Network; FC — Fully Connected Layer; CorrMNN — Correlative Memory Neural Network; TF — Time-frequency Features; FF — Fuzzy Features; FuseLG — Fusion of Local and Global Information; GAU — Gated Attention Unit.

MFCC and TQWT are used. Comparing to manual feature extraction, automatic feature learning can automatically capture complex features within the data, thereby enhancing the performance of the model. Temporal models are better at handling the temporal characteristics of speech data. Models such as RNN, GRU, LSTM, and BiLSTM can capture the temporal features and long-term dependencies in speech data, achieving better results when processing sequential data. This, to some extent, mitigates the issue of poor generalization to out-of-distribution data that deep models often face. Some studies use the Transformer model for feature extraction and achieve better results, but this does not mean that common features are not suitable for speech data. In the table, the TmmNet model achieves 99.91% accuracy and 99.94% sensitivity by inputting MFCC features. Additionally, efforts in feature extraction using large deep learning models such as Wav2vec may be more beneficial for speech diagnosis of PD. Even if an ensemble random forest is used as the classifier, it still achieves a commendable performance of 95% accuracy and 97% sensitivity.

Furthermore, a number of intelligent automatic learning approaches were widely adopted in gait data-based PD detection. The gait recognition algorithms based on GCN (Graph Convolutional Network), CNN (Convolutional Neural Network), and LSTM (Long Short-Term Memory) networks remained the mainstream approach, and researchers paid close attention to gait images, force sensing signals, and accelerations [99–108,116]. Among them, CNN and LSTM were the most common methods due to their effectiveness in extracting spatiotemporal features. Kumar et al. [106] used CNN to diagnose PD and a CNN-LSTM model using the Hoehn & Yahr rating scale led to severity rating prediction. An LSTM classifier was presented for gait-based PD diagnosis and severity rating [105]. The emergence of the Transformer also promoted the development of gait recognition in PD detection [109–114,117,118]. For instance, Zhao et al. had been deeply involved in the field of severity assessment and detection of Parkinson's disease based on Transformer models for a long time and had made outstanding contributions [114,115,118]. A Multi-Level Fine-tuned Transformer [114] was trained on sequential gait data to capture and analyze the overall spatiotemporal information for the recognition of normal and abnormal gait. Moreover, they also designed a FuseLGNet [118] that fused local and global information of the gait data for PD detection. Considering the temporal feature of the gait cycle, an attention-based temporal network [115] was proposed for the time series of GRF (Ground Reaction Forces) for PD severity rating.

**Table 5** summarizes the diagnosis and assessment of PD based on gait data using deep learning techniques. It presents an overview of the dataset, the feature extraction method (handcrafted or non-handcrafted), the classifier model (temporal or non-temporal), and performance metrics – including accuracy, sensitivity, and specificity – for each referenced study. Bold text indicates the best-performing model in that study, and the performance measures are for that model. Deep learning models for Parkinson's disease (PD) diagnosis based on gait data have demonstrated satisfactory performance, with most models exceeding 95% accuracy, studies using temporal models typically have higher accuracy, sensitivity, and specificity. Notably, the AGS-GCN model achieved a perfect accuracy of 100%, potentially attributed to the limited size of the dataset utilized in this study. Compared to the summary in **Table 4**, for PD diagnosis and evaluation of gait data, researchers have focused on the feature extraction stage, employing simple global average pooling layers, fully connected layers, or Softmax layers on the classifier. Some deep learning works do not have a clear division between feature extraction and classification, opting for an end-to-end design. In such cases, Transformer-based models or LSTM-based models, along with some self-designed blocks, are used to gradually extract features and ultimately provide classification results, with the model performance also being good. For example, the FuseLGNet model performs exceptionally well on large sample data through convolutional blocks and FuseLGNet-specific blocks, achieving an accuracy of 99.78%. Some studies in the table use the same dataset, and as observed through the performance metrics, the Transformer-based model achieves the best performance, while the CNN-based and LSTM-based models exhibit similar performance levels. Overall, employing a Transformer for feature extraction and classification may yield superior performance when applied to gait datasets.

Traditional classifiers remained a practical tool for pattern recognition, utilizing the kinematic parameters of hand movements collected by a Leap Motion sensor for PD detection [119]. Additionally, using wearable devices with inertial sensors to obtain acceleration, a CNN-based model was proposed to detect PD tremor [120]. The 3D trajectory of hand motion with time series served PD severity rating [121]. Moreover, the processing of hand motion images and videos based on deep models was highly complex and contained a wealth of disease-related information, making it a focal point for researchers [122–125,129]. A tree-structure-guided graph convolutional network (TSG-GCN) [123]

**Table 6**

Summarizes the diagnosis and assessment of PD based on hand movement data using deep learning techniques.

References	Dataset	No. of subjects	Feature	Method		Accuracy	Sensitivity	Specificity
				Manual	Temporal			
[119]	Signal	16-PD, 16-HC	Speed, Frequency, and Amplitude Estimates	✓	KNN, SVM, DT and RF	✗	98.4%	*
[120]	Signal	5-PD, 5-HC	Acceleration Signals Features	✓	CNN	✗	97.32%	*
[121]	Video	12-PD, 12-HC	3D Motion Trajectory	✓	Two-channel LSTM	✓	95.7%	95.8%
[122]	Signal	*	Slow Motion Features	✓	SAG-RNN	✓	93.55%	*
[123]	Video	637-PD, 116-HC	Tri-directional Skeleton Tree Features	✓	TSG-GCN	✗	73.71%	*
[124]	Signal	100-PD, 35-HC	Time, Frequency, Spectrum, and Autocorrelation	✓	SVM, KNN, CNN, XGBoost and LGBM	✗	92.59%	88%
[125]	Signal	85-PD, 70-HC	Accelerometer Signals Features	✓	LGBM, SVM, KNN and XGBoost	✗	82.41%	*
[126]	Image	68-PD, 73-HC	Spatial Features	✗	CNN	✗	94.7%	*
[127]	Image	105-PD, 43-HC	Spatial Features	✗	CAS Transformer	✓	92.68%	*
[128]	Image & Signal	37-PD, 38-HC	Kinematic and CNN-based Features	✗	DIT	✓	98.87%	*
[111]	Video	20-PD, 23-HC	Skeleton Motion Features	✓	Multi-speed Transformer	✓	96.9%	96.9%
Abbreviations: SAG-RNN — Self-Attention-Guided Recurrent Neural Network; TSG-GCN — Tree-Structure-Guided Graph Convolution Network; CAS Transformer — Coordinate Attention Enhanced Swin Transformer; DIT — Digital Ink Transformer.								

with contrastive learning was proposed to learn spatio-temporal features from 753 RGB video samples for the assessment of parkinsonian hand movements. CNNs, HMMs (Hidden Markov Models), and RNNs (Recurrent Neural Networks) also outperformed other deep models for rating hand movement disorders [126,129]. Methods based on self-attention mechanisms and Transformers were constantly emerging [111,127,128]. Wang et al. [127] proposed a Coordinate Attention Enhanced Swin Transformer (CAS Transformer) model that extracted the fuzzy edge features of handwriting images for PD detection.

Table 6 presents studies based on hand movement data using DL methods. Moreover, the features used in each study are classified according to whether the features are manually extracted. According to the architecture of the model, it is divided into temporal model or non-temporal model. The accuracy column represents the best result in the paper, and the corresponding optimal method has been highlighted. The results indicate that the accuracy of deep learning models based on hand movement data has reached a plateau. Aside from one case where data-related issues resulted in poor performance, other studies employing deep learning techniques on image data, sensor signal data, and video data achieved high levels of accuracy. Specifically, a study with a large number of participants achieved only a 73.71% classification accuracy using a three-way skeleton tree feature extraction method and a TSG-GCN classifier on a dataset that included 637 Parkinson's disease patients and 116 healthy controls. This lower accuracy may be due to the complex and chaotic nature of the dataset's features, which may have introduced significant noise. In contrast, three algorithms achieved optimal performance on different types of hand movement data by adopting superior model frameworks (Transformer, multi-rate Transformer) or utilizing rich intrinsic data features. Among them, the DIT model learns the raw pen-tip kinematics data through a multi-head self-attention module and integrates multi-scale convolutional blocks in the embedding layer to extract helical features, combining the two types of motion features to achieve an accuracy of 98.87%. Another study employed multiple classifiers and determined the optimal classifier (SVM), feature combination, and number of combined features by comparing different motion signal feature combinations. This study achieved an accuracy of 98.4%. The last study is based on human skeletal features shown in videos and uses Multi-speed Transformer to capture joint sequences at different speeds, achieving the best performance of 96.9% with hand video data.

These methods all focus on one biological feature, one symptom or one modality of data, but there is still a lack of comprehensive consideration for the performance and development of PD.

### 3.2. Single-task-based research on PD

This section investigates the impact of single-task-based (PD detection/Severity Rating/Disease Tracking) gait, hand movements, and speech performance on the diagnosis and treatment of Parkinson's disease.

A machine learning-based telehealth infrastructure was proposed for remote symptom assessment, capable of detecting and classifying hand movement tasks based on the UPDRS [130]. For different hand movement tests in the UPDRS, CNNs were also utilized to detect and monitor PD from drawing movements [131]. Due to the prevalence of gait recognition algorithms, fusion networks had gradually been adopted for PD detection. A novel hybrid model was proposed to learn the gait differences between neurodegenerative diseases by fusing and aggregating data from multiple sensors [77]. This model was capable of capturing temporal and structural information of bimodal gait data, i.e., images or signals. The vertical ground reaction force (VGRF, in Newton) was also input into an associated spatio-temporal capsule network for the severity rating of Parkinson's disease [132]. A number of hybrid models [133–135] had gradually emerged in the application of PD speech recognition. Specifically, an audio spectrogram transformer [133] was proposed to analyze multimodal PD speech and handwriting data. An ensemble model [134] was designed for the classification of PD speech data, combining a deep sample learning algorithm with a deep network, realizing deep dual-side learning.

Compared to Table 6, the evaluation results of single-task models using multimodal data in Table 7 show a certain degree of advancement and a clear trend toward stability. We can see that using deep learning models to extract features from data can obtain better performance, and temporal models have advantages over non-temporal models for signal sequence data. Notably, one study developed an automated learning system and trained a relevant spatiotemporal capsule network (ASTCapsNet) on a multisensor dataset to analyze multimodal information for gait recognition. The system designed a low-level feature extractor based on recurrent memory units and relational layers, and a high-level feature extractor for spatiotemporal feature extraction of gait. Subsequently, a Bayesian model was used to make decisions on class labels, achieving an accuracy of 97.31% on a PD-related dataset. Additionally, based on spatiotemporal features, a study proposed a feature aggregation method for analyzing gait differences using data from multiple sensors. This approach utilized an SFE extractor to generate representative features from images or signals and designed a new Correlation Memory Neural Network (CorrMNN) structure to extract temporal features from two modalities, capturing temporal information. Finally, an HMM switch was embedded to associate observations with single state estimates, achieving an accuracy of 98.93%.

**Table 7**

Single task of PD based on multimodal data.

References	Dataset	No. of subjects	Feature	Method		Accuracy	Sensitivity	Specificity
				Manual	Temporal			
[130]	Hand Signal	4-PD, 5-HC	Speed, Amplitude, Hesitation, Halts and Decrementing Amplitude Coordinates, Pressure, and Grip Angle	✓	KNN, SVM, DT	✗	93%	*
[131]	Hand Signal	62-PD, 15-HC	Spatio-temporal Gait Features	✓	CNN	✗	96.5%	*
[77]	Gait Signal	93-PD, 73-HC	Spatio-temporal Gait Features	✗	HMM	✓	98.93%	*
[132]	Gait Signal	93-PD, 73-HC	Spatio-temporal Gait Features	✗	ASTCapsNet	✓	97.31%	*
[133]	Handwriting & Voice	81-PD, 85-HC	Spatial Features and MFCC	✓	DeiT, AST	✓	92.37%	*
[134]	Speech	34-PD, 20-HC	Original Features and Deep Features	✗	Deep dual-side learning ensemble model	✓	99.67%	99.35%
								99.7%

**Abbreviations:** HMM — Hidden Markov Model; ASTCapsNet — Associated Spatio-Temporal Capsule Network; DeiT — Data-efficient Image Transformer; AST — Audio Spectrogram Transformer.

Based on multimodal speech data, another study further developed a deep dual-side learning ensemble model. This model performed deep feature learning using an embedded stack group sparse autoencoder and integrated these features with original speech features to create mixed feature data. A classification model was then trained on this feature space, achieving the highest accuracy of 99.67%, making it one of the most accurate algorithms in related research.

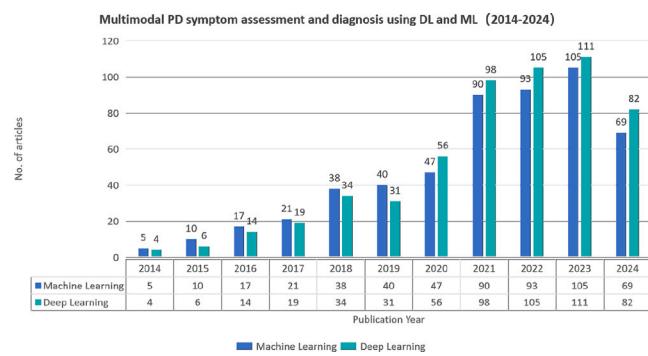
The methods and datasets mentioned have indeed played a crucial role in advancing the diagnosis and treatment of Parkinson's disease (PD). However, studies that focus on single modalities and single tasks may encounter limitations in terms of applicability and flexibility. They might not fully accommodate the complexity and variability of the disease across different stages and settings. Additionally, their scalability might be limited, which could hinder their ability to encompass the intricate details of PD in diverse patient populations. To address these limitations, there is a need for more comprehensive, multimodal, and scalable approaches that can provide a more nuanced understanding of the disease.

### 3.3. Multimodal learning for PD symptom

Deep learning methodologies play a major role in multimodal and multi-symptom recognition for Parkinson's Disease, significantly enhancing the diagnostic process by leveraging the rich information encapsulated within diverse data types.

Deep learning algorithms adeptly learn temporal, spatial, and structural information from gait, hand movements, and speech in PD patients, integrating this multimodal information to provide diagnostic results. For instance, a hierarchical architecture [136] combined Hidden Markov Models (HMMs), three single-symptom models, and machine learning to quantify multi-symptom severity in PD patients. A capsule network [132] was trained on multimodal PD gait data to extract spatiotemporal and structural information. CNNs, LSTMs, ResNets, and Transformer-based hybrid models have been widely applied in multimodal PD symptom assessment [111,137–143]. A Spatio-Temporal Siamese Neural Network [142] was proposed for multimodal handwriting recognition, based on images and signals, and included a Siamese bidirectional memory neural network (SiamBiMNN) and a Siamese octave convolutional neural network (SiamOctCNN). A Perceiver architecture-based multimodal deep learning framework [140] was presented for analyzing gait time-series signals and hand-crafted features from GRF sensors, allowing them to complement each other and improve the predictive ability for PD diagnosis and PD symptom severity estimation. Furthermore, a multi-speed transformer architecture [111] within a Deep Learning and Shallow Learning framework served for neurodegenerative disease classification and activity recognition.

In summary, the diagnostic and evaluative algorithms for Parkinson's disease are undergoing a significant transformation. The focus is



**Fig. 5.** Multimodal PD symptom assessment and diagnosis methods using DL and ML (2014–2024).

shifting from machine learning to the more profound insights of deep learning, from the guidance of supervised models to the autonomy of unsupervised ones. The learning paradigm is expanding from single-modal to the rich integration of multimodal approaches, broadening the understanding of the disease's manifestations. It is also transitioning from tackling one task at a time to handling multiple tasks concurrently, enhancing the algorithms' versatility. Furthermore, the move from a single model to an ensemble of models is aimed at achieving a more robust, accurate, and dependable diagnostic process, ensuring that the evaluation results are both exhaustive and credible.

Significantly, as shown in Fig. 5, multimodal PD diagnosis and evaluation algorithms require significantly fewer learning iterations compared to their single-modal counterparts. Prior to 2020, the number of new single-modal machine learning algorithms increased by 5–40 per year. In contrast, the number of multimodal deep learning algorithms fluctuated between 4–31, generally maintaining a lower overall level than that of the machine learning algorithms. Post-2020, there was a surge; the number of both types of algorithms doubled, with deep learning taking the lead.

As shown in Table 8, we have introduced a total of 13 pieces of literature on multimodal PD detection and evaluation. This includes 6 works based on gait multimodal recognition, 2 works on MRI image processing and medical signal fusion, and 5 works on multi-symptom and multimodal fusion evaluation. These works target various data types such as images, videos, biological signals, physical signals, and temporal data, employing deep learning (DL) for feature extraction and training, and machine learning (ML) for classification and evaluation. Most DL methods are based on hybrid neural networks, with a few relying on ensemble learning and other techniques. The accuracy of most multimodal learning algorithms exceeds 90%, and the performance of fusion models is generally superior to that of single models. The results of algorithms based on temporal networks and convolutional networks,

**Table 8**

Multimodal PD symptom analysis using machine learning and deep learning techniques.

References	Dataset	Modality	Method	Accuracy
[71]	Audio, motion and facial data	Image, signal	STN+RGA	95.08%
[72]	fMRI+SNP	Image	CERF	80.00%
[77]	Gait	VGRF+time series	CorrMNN	99.31%
[132]	Gait	Images, skeleton, force signal	ASTCapsNet	97.31%
[80]	Bradykinesia, rigidity and tremor	Force signal, IMU, MMG	Voting Classifier	85.40%
[136]	Bradykinesia and tremor	Accelerometer and gyroscope data	HMMs	89.28%
[137]	Motor and non-motor function data	Time-series, non timeseries	LGBM, RF	94.89%
[138]	MRI and CSF	Neuroimaging and biological features	CNN	93.33%
[111]	Gait	Videos, time series	Multi-speed transformer network	96.90%
[139]	Gait	Inertial, stride, and pressure data	CNN-BiLSTM	98.89%
[140]	Gait	GRF+time series	Cross-attention, Transformer	97.30%
[141]	Freezing of Gait	Acceleration of ankle, thigh, and trunk	MCT-Net	96.21%
[143]	Multimodal-medical-data	Medical history, medication use, neuropsychological assessments, functional evaluations and multimodal neuroimaging	3D transformer-based Models	96% (AUC)

**Abbreviations:** STN — Spatial-Temporal Network; RGA — Relation Global Attention; CERF — Cluster Evolutionary Random Forest; CorrMNN — Correlative Memory Neural Network; ASTCapsNet — Associated Spatio-Temporal Capsule Network; IMU — Inertial Measurement Unit; MMG — Mechanomyography; HMMs — Hidden Markov Models; LGBM — Light Gradient Boosting Machines; GRF — Ground Reaction Force; MCT-Net — Multi-Channel Time-series Neural Network.

as presented in the literature [77,139], outperform those in other studies. Due to their dependence on specific experimental conditions and datasets, other algorithms also exhibit outstanding advantages within their respective domains [144].

#### 4. Multi-source datasets

The training and evaluation processes of learning models, irrespective of the underlying methodology—be it supervised, unsupervised, or any alternative approach—rely on datasets that encapsulate the essence of the task at hand. Utilizing these datasets not only facilitates the assessment of a method's efficacy in addressing particular challenges but also enables a comparative analysis with alternative strategies.

In this section, we delve into a variety of datasets tailored to the analysis of different motor and non-motor symptoms. These include speech, gait, hand movement, and multimodal datasets, each reflecting the complexity of Parkinson's disease. Varied in origin, format, composition, scale, and modality, these datasets offer a rich tapestry for the exploration and understanding of the disease's manifestations.

##### 4.1. Speech datasets

**Oxford Parkinson's Disease Detection Dataset** [145]: This dataset is composed of a range of biomedical voice measurements from 31 individuals, 23 of whom have Parkinson's disease (PD). Each column in the table represents a specific voice measure, and each row corresponds to one of the 195 voice recordings from these individuals, as indicated in the "name" column. The primary objective of the data is to discriminate between healthy individuals and those with PD, based on the "status" column, which is set to 0 for healthy and 1 for PD.

**PDSpeech Dataset** [146]: This dataset consists of training and test files. The training data is derived from 20 people with Parkinson's (PWP), including 6 females and 14 males, and 20 healthy individuals, comprising 10 females and 10 males, who volunteered at the Department of Neurology in Cerrahpasa Faculty of Medicine, Istanbul University. From all subjects, multiple types of sound recordings were taken, including 26 voice samples with sustained vowels, numbers, words, and short sentences. A suite of 26 linear and time-frequency based features were extracted from each voice sample. The dataset also includes the UPDRS scores for each patient, as determined by an expert physician, allowing for its use in regression analysis.

**MDVR-KCL dataset** [147]: This dataset comprises voice files of early and late Parkinson's disease patients and healthy controls, recorded with mobile devices at King's College London (KCL) Hospital in Brixton, London, from September 26 to 29, 2017. The recordings

were made in a typical examination room with an area of about ten square meters and a typical reverberation time, utilizing a room with approximately ten square meters and a standard reverberation time for voice recording, with a duration of 500 ms. The recording was conducted in a natural setting, with participants placing the phone to their preferred ear and positioning the microphone close to their mouth. It can be assumed that all recordings were made within the reverberation radius, suggesting they can be considered "clean".

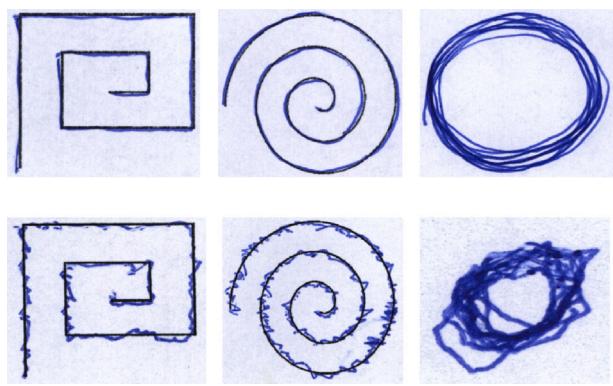
A Motorola Moto G4 smartphone was used as the recording device. Through a developed application, high-quality recordings with a sampling rate of 44.1 kHz and a bit depth of 16 bits (CD audio quality) were achieved in '.wav' format. The dataset includes data from 16 PD patients and 21 healthy controls, with each participant's scores labeled on the Hoehn & Yahr (H&Y) scale, as well as the UPDRS II part 5 and UPDRS III part 18 scales.

**IPVS dataset** [148]: This dataset includes voice recordings from 28 PD patients and 20 healthy controls, all of which were collected at a 16 kHz sampling rate in a quiet, echo-free, and temperature-controlled room. The microphone was positioned 15 to 25 cm from the participants. The participants performed various tasks, including two phonations of the vowels /a/, /e/, /i/, /o/, and /u/, and syllable execution of "ka" and "pa" for 5 s. In this study, the reading of phonetically balanced phrases and vowel recordings were utilized, with a phonetically balanced text read twice by each participant.

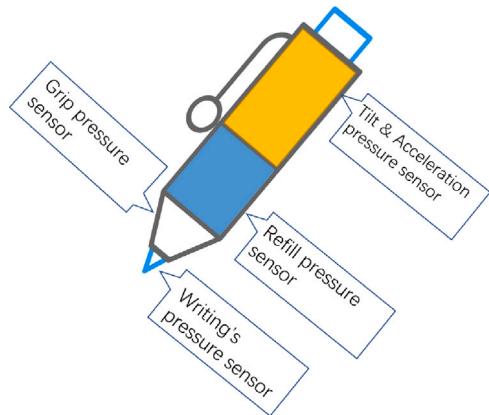
**PDVoice dataset** [149]: The data in this study were collected from 188 Parkinson's disease patients (107 males and 81 females) from the Neurology Department of Istanbul University, with ages ranging from 33 to 87 years old (mean age  $65.1 \pm 10.9$ ). The control group consisted of 64 healthy individuals (23 males and 41 females), with ages ranging from 41 to 82 years old (mean age  $61.1 \pm 8.9$ ). During the data collection process, the microphone was set to 44.1 kHz. After medical examination, sustained vocalizations of the vowel /a/ were collected from each subject, repeated three times.

##### 4.2. Hand-movement datasets

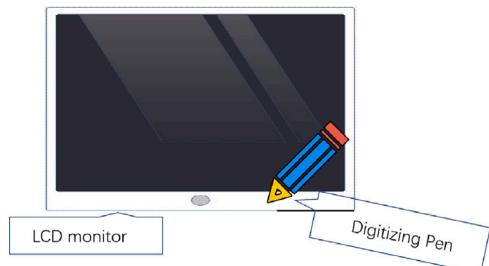
**NewHandPD dataset** [150]: This dataset includes images from two groups of individuals – healthy controls and PD patients – obtained during a handwriting test designed to assess their personal writing ability when filling out a form. The handwriting data was collected at the Botucatu Medical School of São Paulo State University, Brazil, and involved tasks unusual for PD patients, such as drawing "spirals", "meanders", and "circles" (Fig. 6). The dataset consisted of 66 participants (35 healthy controls and 31 PD patients), each asked to complete



**Fig. 6.** The handwriting data used in this project: We evaluate the proposed approach over the NewHandPD dataset, consists of the image data, i.e., spiral, meander, circled movements images.



**Fig. 7.** The biometric pen. It includes four sensors: tilt & acceleration sensor, refill pressure sensor, grip pressure sensor, and writing's pressure sensor.



**Fig. 8.** The digitized graphics tablet.

12 tests, including four spirals, four meanders, and two circular movements (one in the air and the other on paper). After the test, each person had nine images saved. The dataset also includes signal data from a smart pen used in the diadochokinesia test, where subjects performed hand-wrist movements while holding the pen with straight arms, and the generated signals were recorded. These signals were extracted from the BiSP® smart pen (Fig. 7), which has four sensors and six features, including microphone, finger grip, axial pressure of the ink refill, and tilt and acceleration in the "X", "Y", and "Z" directions.

**PARKINSON\_HW dataset [151]:** This dataset consists of 62 PD patients and 15 healthy controls who attended the Department of Neurology at Cerrahpasa Faculty of Medicine, Istanbul University. All participants underwent three handwriting tests: the Static Spiral Test, Dynamic Spiral Test, and Stability Test on Certain Point, using a Wacom Cintiq 12WX graphics tablet (Fig. 8). This tablet, integrated

with an LCD monitor displaying the PC's screen, was used with specially designed software for recording handwritten graphics and testing coordination in PD patients.

The Static Spiral Test (SST) was frequently used for clinical research, where patients were asked to trace three Archimedes spirals displayed on the drawing board using the software. The Dynamic Spiral Test (DST) differed by having the Archimedes spiral appear and disappear at certain intervals, challenging the patient to memorize and continue drawing the pattern. The Stability Test on Certain Point (STCP) aimed to measure the stability of the patient's hand or the degree of hand tremor, with the subject asked to hold a digital pen on a red dot in the middle of the screen.

**PD\_HW dataset [152]:** This PD and HC handwriting database comprises 25 individuals diagnosed with Parkinson's and 15 healthy individuals, all of whom visited the Department of Neurology at Cerrahpasa Faculty of Medicine, Istanbul University. A range of handwriting assessments was conducted, including the Static Spiral Test (SST), the Dynamic Spiral Test (DST), and the Stability Test on Certain Point (STCP). The dataset also contains images of the spirals drawn by the Parkinson's patients, suitable for image analysis, making the dataset valuable not only for handwriting analysis but also for regression analysis purposes.

**PD-Tappy Dataset [153]:** This PD and HC handwriting database comprises 25 individuals diagnosed with Parkinson's and 15 healthy individuals, all of whom visited the Department of Neurology at Cerrahpasa Faculty of Medicine, Istanbul University. A range of handwriting assessments was conducted, including the Static Spiral Test (SST), the Dynamic Spiral Test (DST), and the Stability Test on Certain Point (STCP). The dataset also contains images of the spirals drawn by the Parkinson's patients, suitable for image analysis, making the dataset valuable not only for handwriting analysis but also for regression analysis purposes.

**NeuroQWERTY MIT-CSXPD Dataset [154]:** This dataset includes keystroke records from 85 individuals, both with and without PD. It demonstrates that regular computer keyboard use can help identify early motor symptoms of PD. Participants were recruited from two Madrid-based movement disorder units, following protocols approved by the Massachusetts Institute of Technology and Spanish hospitals. Using a Lenovo G50-70 laptop with Manjaro Linux, subjects typed on a standard word processor, simulating their home typing habits. The keystroke data, collected with a high temporal resolution, aimed to capture typing pattern nuances that could indicate early Parkinson's Disease symptoms, highlighting the potential of everyday computer interactions as a diagnostic tool for motor impairments in PD.

**PADS Dataset [155]:** These evaluations were recorded using a smart device-based system consisting of two smartwatches and one smartphone. The smartwatches, worn on the patient's wrists, synchronously recorded 11 interactive exercise tasks designed by neurologists. A total of 5159 measurement steps involving 469 individuals were captured. The PADS dataset includes all acceleration and rotation sensor signals, along with detailed information on movement steps, demographics, medical history, and PD-specific non-movement symptoms. We believe this extensively annotated dataset provides a suitable foundation for the training, validation, and optimization of future motion analysis technologies and sensor-driven systems. Three participant groups were recorded: (1) Parkinson's disease patients, (2) Differential diagnosis (DD) including primary tremor, atypical Parkinson's disease, secondary causes of Parkinson's disease, and multiple sclerosis, and (3) Healthy control group (HC).

#### 4.3. Gait datasets

**PDgait dataset [156]:** This study utilized gait signals from PhysioNet, which consists of three PD gait sub-datasets contributed by three researchers. Ninety-three patients with idiopathic PD and 73 healthy controls (average age 66.3, 55% male) were asked to walk at their

usual, self-selected pace for about two minutes while wearing shoes equipped with 8 force sensors under each foot. The Vertical Ground Reaction Force (VGRF, in Newtons) was recorded by the force sensors as a function of time at 100 samples per second.

Each sample had 16 sensor outputs and two total outputs for each foot, with 19 parameters corresponding to each row in the dataset file: the timestamp, 16 VGRF foot parameters, and 2 total VGRF outputs. When subjects stood still, the 16 sensors recorded an initial value that changed as they began walking. With VGRF data, one can analyze the force record as a function of time and location, derive measures reflecting the center-of-pressure over time, and calculate stride or swing time for each foot. This allows for the study of stride-to-stride dynamics and the variability of these time series.

**NDDs dataset [157]:** This dual-modal dataset contains temporal and force-sensing gait data from 48 patients with neurodegenerative diseases (ALS, HD, and PD) and 16 healthy controls (CO). The data were collected simultaneously using different sensors in a foot-switch system, which provided accurate estimates of the start and end stages of successive steps. The system utilized commercially available transducers and can be easily reproduced in laboratory environments. It included two  $1.5 \text{ in}^2$  force sensing resistors and a  $390\Omega$  measuring resistor to acquire stride time intervals by measuring the variation in force changes during gait.

**tDCS FOG dataset [158]:** The tDCS FOG data were originally collected as part of a sham-controlled, double-blind, multi-site randomized trial examining the effects of transcranial direct current stimulation on freezing of gait (FOG). The competition dataset included 71 PD patients with mild-to-moderate symptoms (Hoehn and Yahr score of 1-3.570) and self-reported FOG (via the NFOG-Q), who performed videotaped FOG-provoking tests while wearing a sensor with a 3D accelerometer placed on the lower back. Patients were excluded from the original study if they could not ambulate 20 m unassisted, suffered from specific medical conditions detailed in the original work, or were unfit for the treatments. The FOG-provoking test protocol was conducted in the on-medication state and, if the patient agreed, also in the off-medication state. From the 71 tDCS FOG patients with three FOG class labels, 833 data series from 62 patients were included in the training data, 138 data series from 14 patients in the public test set, and 68 data series from 7 patients in the private test set.

**PD-FOG dataset [159]:** This dataset contains annotated readings from 3 acceleration sensors at the hip and leg of Parkinson's disease patients experiencing freezing of gait (FoG) during walking tasks. The Daphnet Freezing of Gait Dataset was devised to benchmark automatic methods for recognizing gait freeze from wearable acceleration sensors placed on legs and the hip. The dataset was recorded in a lab with a focus on capturing numerous freeze events. Participants performed various tasks: straight line walking, walking with turns, and a realistic activity of daily living (ADL) task, such as going into different rooms to fetch coffee and opening doors.

**PD-Video dataset [111]:** The dataset comprises 115 videos (61 of control subjects and 54 of people with diseases) from 43 subjects (22 females and 21 males, including 23 control subjects and 20 patients). The subjects walked a 4-m straight line on the floor. The camera was positioned perpendicularly about 4 m away from the line at a height of 2 m. The videos varied in length and were recorded in different settings with various backgrounds and scenes. Each video shows a person walking a linear path in both directions, from left to right and vice versa, recorded at 25 fps. Control subjects ranged from 30–75 years, while patients were between 65–90 years. With the assistance of trained neurologists and psychologists, the disease stage was assessed using the Mini-Mental State Examination, classifying each subject as normal, mild, or severe.

#### 4.4. Multimodal datasets

**PD-Posture-Gait dataset [160]:** This dataset offers a multi-camera, multimodal dataset for vision-based applications, utilizing a wheeled robotic walker equipped with a pair of affordable cameras. Depth data were acquired at 30 fps from a total of 14 healthy participants walking at three different gait speeds, across three different walking scenarios/paths at three different locations. Concurrently, accurate skeleton joint data were recorded using an inertial-based commercial motion capture system, providing a reliable ground-truth for both classical and novel (e.g., machine learning-based) vision-based applications. In total, the database contains approximately 166K frames of synchronized data, equating to 92 min of total recording time. This dataset may contribute to the development and evaluation of: (i) classic or data-driven vision-based pose estimation algorithms; (ii) applications in human detection and tracking, and movement forecasting; (iii) and gait/posture metrics analysis using a rehabilitation device.

**REMAP dataset [161]:** This study recruited 24 participants, 12 with PD (mean age 61.25; 7 males, 5 females) and 12 healthy control volunteers (mean age 59.25, 3 males, 9 females). The controlled dataset contains pseudonymous data, including full skeleton data for all Sit-to-Stand (STS) and turning episodes. It also contains bilateral wrist-worn accelerometry data for these episodes from all participants, along with accelerometry data for non-turning, non-STS action episodes. The reason for including accelerometry data in the controlled dataset is due to the proven re-identifiability risk of accelerometry, which contravenes the study consent conditions relating to data sharing. We include individual-level demographic and clinical rating scale score outcomes given in ranges. The open dataset contains anonymous data, either because it is provided at the cohort-level (e.g., demographic data).

**Multimodal-FOG dataset [162]:** This study gathered and presented a new multimodal dataset by combining rich physical and physiological sensor information. The multimodal data, including electroencephalogram (EEG), electromyogram (EMG), electrocardiogram (ECG), skin conductance (SC), and acceleration (ACC) in walking tasks, were collected using a high-quality hardware system with integrated commercial and self-designed sensors. A standard experimental procedure was carefully designed to induce FOG in hospital surroundings. A total of 12 PD patients completed the experiments, producing a total of 3 h and 42 min of valid data. The FOG episodes in the multimodal data were labeled by two qualified physicians.

**Multi-PD-FOG dataset [163]:** This database includes videos, inertial measurement unit data, and clinical scales of freezing of gait in individuals with Parkinson's disease during a turning-in-place task. It encompasses both (a) patients' demography and (b) clinical conditions (PD severity, the number and duration of FOG episodes for each individual, clinical scales, and medication state during testing), as well as kinematics (video, acceleration, and angular velocity) during a turning-in-place task in individuals with PD in the ON medication state. A convenience sample of 35 idiopathic PD patients with FOG (16 females and 19 males) was recruited to participate in this study. The inertial sensors recorded triaxial linear accelerations and triaxial angular velocities at 128 Hz. Turning trials were recorded via a commercial digital camera (Sony, 30 Hz).

**Multi-Symptom dataset [71]:** This database includes videos, inertial measurement unit data, and clinical scales of freezing of gait in individuals with Parkinson's disease during a turning-in-place task. It encompasses both (a) patients' demography and (b) clinical conditions (PD severity, the number and duration of FOG episodes for each individual, clinical scales, and medication state during testing), as well as kinematics (video, acceleration, and angular velocity) during a turning-in-place task in individuals with PD in the ON medication state. A convenience sample of 35 idiopathic PD patients with FOG (16 females and 19 males) was recruited to participate in this study. The inertial sensors recorded triaxial linear accelerations and triaxial angular velocities at 128 Hz. Turning trials were recorded via a commercial digital camera (Sony, 30 Hz).

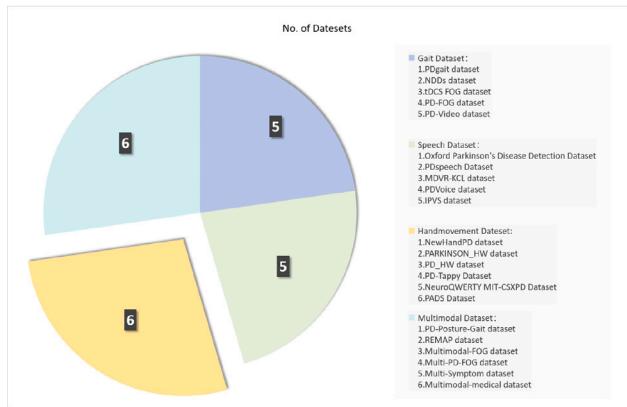


Fig. 9. No. of the datasets introduced in this paper.

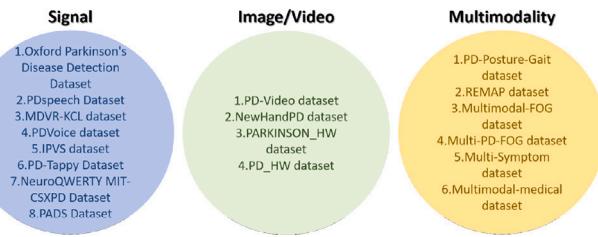


Fig. 10. Partition of datasets with different modalities.

**Multimodal-medical dataset [143]:** This dataset includes individual-level demographics, health history, neurological testing, physical/neurological exams, and multisequence MRI scans. These data sources, when available, were aggregated from nine independent cohorts: 4RTNI, ADNI, AIBL, FHS, LBDSU, NACC, NIFD, OASIS, and PPMI. It collected demographics, personal and family history, laboratory results, findings from physical/neurological exams, medications, neuropsychological tests, and functional assessments, as well as multisequence magnetic resonance imaging (MRI) scans from 9 distinct cohorts, totaling 51,269 participants. All participants or their designated informants provided written informed consent. All protocols received approval from the respective institutional ethical review boards of each cohort. There were 19,849 participants with normal cognition (NC), 9357 participants with mild cognitive impairment (MCI), and 22,063 participants with dementia.

Fig. 9 illustrates our presentation of four types of Parkinson's disease (PD) datasets, which include five speech datasets, six gait datasets, six hand action datasets, and five multimodal datasets. These are predominantly composed of data from three modalities: images, videos, and signals. Despite variations in signal sources, this paper does not further categorize them into subtypes such as physical or biological signals. Interested readers are encouraged to conduct detailed research using the dataset addresses provided in this article.

As depicted in Fig. 10, the datasets are categorized into three groups based on their modality. There are 12 datasets that consist of signal data, while four datasets comprise image and video data. This distribution suggests that sensor-based data collection is more prevalent for Parkinson's disease patients, whereas image and video data collection faces challenges due to privacy concerns, as well as issues of redundancy and noise that necessitate further preprocessing. The limited image data of Parkinson's disease patients in multimodal datasets underscores the need for more diverse non-contact data types, such as images and videos, to foster new advancements in the assessment of PD's external symptoms.

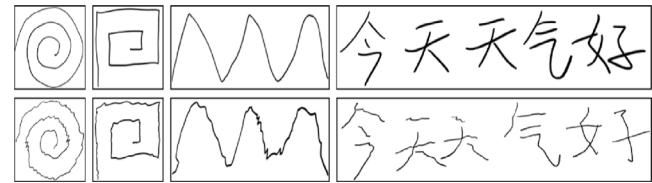


Fig. 11. Handwriting samples.

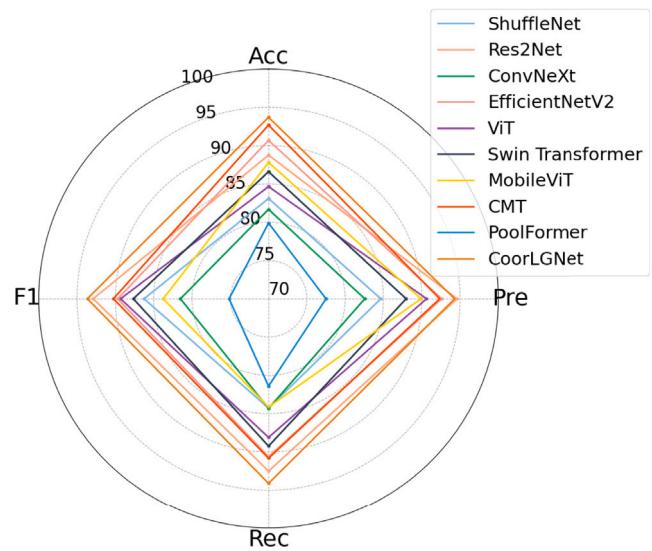


Fig. 12. Spider graph for model comparison.

Table 9

Performance comparison with state-of-the-art methods.

Method	Acc (%)	Pre (%)	Rec (%)	F1 (%)
ShuffleNet [164]	83.10	84.65	84.27	86.25
Res2Net [165]	88.76	94.60	92.50	92.88
ConvNeXt [166]	81.68	82.56	84.31	81.50
EfficientNetV2 [167]	90.68	92.60	90.61	89.69
ViT [168]	84.65	90.62	88.10	89.31
Swin Transformer [169]	86.59	87.96	89.23	87.65
MobileViT [170]	87.80	89.90	84.14	83.75
CMT [171]	92.68	92.35	90.82	90.25
PoolFormer [172]	79.89	77.50	81.41	75.15
CoorLGNet [173]	93.70	94.30	94.10	93.65

#### 4.5. PD clinical handwriting dataset

PD Clinical Handwriting Dataset was collected at the Affiliated Hospital of Qingdao University, aiming to provide a comprehensive resource for analyzing handwriting patterns to detect early signs of Parkinson's Disease.

The dataset included 70 participants, consisting of 28 Parkinson's disease (PD) patients, 28 healthy controls, and 14 individuals in a false-positive control group. Handwriting data were collected in a quiet, warm indoor setting using a digitizing tablet. As shown in Fig. 11, participants performed a series of writing tasks, such as drawing Archimedean spirals, waves, and writing fixed sentences, to evaluate both short-term and long-term handwriting abilities. The tasks were designed with progressively increasing complexity to comprehensively assess motor control and detect subtle impairments.

The performance of various models on the clinical handwriting dataset for Parkinson's disease is visually illustrated in Fig. 12, which provides a clear comparison of the performance of ShuffleNet [164], Res2Net [165], ConvNeXt [166], EfficientNetV2 [167], ViT [168], Swin Transformer [169], MobileViT [170], CMT [171], PoolFormer [172],

and our proposed CoorLGNNet [173] model. To ensure statistical robustness, all models were evaluated using 5-fold cross-validation, with performance metrics averaged across folds. Each axis within the graph serves as a proxy for an individual assessment criterion, and the polygonal contours attributed to each model mirror their respective manifestations across these diverse indices. The magnitude of the polygon's area is indicative of the model's aggregate proficiency; a larger area suggests a more robust composite performance. It is discernible that the polygon representing the CooLGNNet model is in close proximity to the periphery on all axes, signifying its preeminence over alternative models in terms of accuracy, precision, recall, and F1 score. The graphical representation not only delineates the comparative efficacy of the models but also elucidates the intricate interdependencies among different metrics. A detailed numerical comparison is presented in Table 9, meticulously listing the specific values for each model. Notably, the CooLGNNet model achieves Accuracy of 95.12%, Precision of 95.29%, Recall of 95.12%, and F1 Score of 95.15%, demonstrating its superior performance compared to all other models evaluated. These results unequivocally highlight that CooLGNNet outperforms its counterparts, solidifying its position as the most effective model in this comprehensive comparative analysis.

This dataset provides a valuable resource for the early detection of PD through handwriting analysis. Its multimodal nature and comprehensive task design enable robust model validation and comparison, contributing to advancements in PD diagnosis.

## 5. Conclusions and future directions

This article presents an extensive review of machine learning and deep learning methods used for the assessment and detection of Parkinson's disease (PD) symptoms from 2014 to 2024. It summarizes the profound impact of PD on three major manifestations: speech, gait, and hand movements, and elucidates the progress made in diagnostic accuracy and symptom recognition through the application of neuromodulating algorithms. Initially, we surveyed the literature on PD detection and motor symptom evaluation using artificial intelligence over the past decade, analyzing the recognition accuracy of these methods. Subsequently, we conducted an in-depth investigation into multimodal data analysis methods for each symptom and provided a vast collection of publicly available PD datasets for researchers, facilitating the simplification of the research process. Finally, we introduced the performance of algorithms on our own collected dataset, verifying the effectiveness of some algorithms in real-world scenarios.

The future of machine learning research in Parkinson's disease (PD) continues to face several challenges. Firstly, the scarcity of multimodal datasets is a significant issue. Existing datasets primarily focus on single symptoms, and there is a lack of comprehensive datasets that encompass multiple symptoms. This limitation hinders the development of holistic diagnostic approaches. Secondly, while wearable devices are the primary means of data collection, they may affect patients' emotional states and the reliability of the data. Moreover, data collection using these devices is time-consuming and resource-intensive. Additionally, current algorithms are mostly designed for single symptoms and modalities, making it difficult to handle the complexity and heterogeneity of multimodal, multi-symptom data. As a result, there is still no integrated clinical framework for PD assessment and detection. Furthermore, the use of diverse sensors and data collection methods in research leads to inconsistencies in data formats, scales, and conditions. This variability increases the complexity of data preprocessing and affects diagnostic efficiency. Beyond the Unified Parkinson's Disease Rating Scale (UPDRS), there is a lack of standardized evaluation tools. Therefore, there is an urgent need to develop a new assessment paradigm to address these challenges.

## CRediT authorship contribution statement

**Aite Zhao:** Project administration, Methodology, Funding acquisition, Data curation, Conceptualization. **Yongcan Liu:** Investigation. **Xinglin Yu:** Investigation. **Xinyue Xing:** Investigation. **Huiyu Zhou:** Writing – review & editing.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

This research was supported in part by Natural Science Foundation of Shandong Province under Grant No. ZR2021QF084, National Natural Science Foundation of China under Grant No. 62106117, National College Student Innovation and Entrepreneurship Training Program under Grant No. SA2500000567

## Data availability

Data will be made available on request.

## References

- [1] Y. Yuzhe, Y. Yuan, Z. Guo, W. Hao, C. Ying-Cong, L. Yingcheng, T. Christopher, C. Daniel, B. Jan, J. Mithri, V. Aleksandar, E. Terry, L. Melissa, D. Ray, K. Dina, Artificial intelligence-enabled detection and assessment of Parkinson's disease using nocturnal breathing signals, *Nature Med.* 10 (28) (2022) 2207–2215.
- [2] A. Caroppo, A. Manni, G. Rescio, A.M. Carluccio, P.A. Siciliano, A. Leone, Movement disorders and smart wrist devices: A comprehensive study, *Sensors* 25 (1) (2025) 266.
- [3] J. Xu, X. Xu, X. Guo, Z. Li, B. Dong, C. Qi, C. Yang, D. Zhou, J. Wang, L. Song, et al., Improving reliability of movement assessment in Parkinson's disease using computer vision-based automated severity estimation, *J. Parkinson's Dis.* (2025) 1877718X241312605.
- [4] T. He, J. Chen, M.S. Hossain, Z. Lyu, Enhanced detection of early Parkinson's disease through multi-sensor fusion on smartphone-based IoT platforms, *Inf. Fusion* 117 (2025) 102889.
- [5] S. Niu, J. Ma, L. Bai, Z. Wang, L. Guo, X. Yang, EHR-KnowGen: Knowledge-enhanced multimodal learning for disease diagnosis generation, *Inf. Fusion* 102 (2024) 102069.
- [6] X. Zhang, Z. Wan, J. Zhao, X. Li, A. Liu, X. Fan, W. Sun, F. Tian, D. Wang, Doctorpupil: A virtual reality system for Parkinson's diagnosis through task-evoked pupil response, *IEEE J. Biomed. Heal. Inform.* (2025).
- [7] Y. Guo, J. Yang, Y. Liu, X. Chen, G.-Z. Yang, Detection and assessment of parkinson's disease based on gait analysis: A survey, *Front. Aging Neurosci.* 14 (2022) 916971.
- [8] N. Kour, S. Arora, et al., Computer-vision based diagnosis of Parkinson's disease via gait: A survey, *Ieee Access* 7 (2019) 156620–156645.
- [9] L. van Gelderen, C. Tejedor-García, Innovative speech-based deep learning approaches for parkinson's disease classification: A systematic review, 2024, arXiv preprint arXiv:2407.17844.
- [10] Q.C. Ngo, M.A. Motin, N.D. Pah, P. Drotár, P. Kempster, D. Kumar, Computerized analysis of speech and voice for Parkinson's disease: A systematic review, *Comput. Methods Programs Biomed.* 226 (2022) 107133.
- [11] M.A. Islam, M.Z.H. Majumder, M.A. Hussein, K.M. Hossain, M.S. Miah, A review of machine learning and deep learning algorithms for parkinson's disease detection using handwriting and voice datasets, *Heliyon* (2024).
- [12] H. Huang, W. Pedrycz, K. Hirota, F. Yan, A multiview-slice feature fusion network for early diagnosis of Alzheimer's disease with structural MRI images, *Inf. Fusion* (2025) 103010.
- [13] P. Yan, W. Gong, M. Li, J. Zhang, X. Li, Y. Jiang, H. Luo, H. Zhou, TDF-Net: Trusted dynamic feature fusion network for breast cancer diagnosis using incomplete multimodal ultrasound, *Inf. Fusion* 112 (2024) 102592.
- [14] J. Lv, B.-G. Kim, B. Parameshachari, A. Slowik, K. Li, Large model-driven hyperscale healthcare data fusion analysis in complex multi-sensors, *Inf. Fusion* 115 (2025) 102780.
- [15] W. Zheng, L. Yan, C. Gou, Z.-C. Zhang, J. Jason Zhang, M. Hu, F.-Y. Wang, Pay attention to doctor-patient dialogues: Multi-modal knowledge graph attention image-text embedding for COVID-19 diagnosis, *Inf. Fusion* 75 (2021) 168–185.

- [16] F.A. Torghabeh, E.A. Moghadam, S.A. Hosseini, Simultaneous time-frequency analysis of gait signals of both legs in classifying neurodegenerative diseases, *Gait Posture* 113 (2024) 443–451.
- [17] F. Abedinzaheh Torghabeh, S.A. Hosseini, Y. Modaresnia, Potential biomarker for early detection of ADHD using phase-based brain connectivity and graph theory, *Phys. Eng. Sci. Med.* 46 (4) (2023) 1447–1465.
- [18] M. Shahbakhi, D.T. Far, E. Tahami, Speech analysis for diagnosis of Parkinson's disease using genetic algorithm and support vector machine, *J. Biomed. Sci. Eng.* 2014 (2014).
- [19] A. Benba, A. Jilbab, A. Hammouch, S. Sandabad, Voiceprints analysis using MFCC and SVM for detecting patients with parkinson's disease, in: 2015 International Conference on Electrical and Information Technologies, ICEIT, IEEE, 2015, pp. 300–304.
- [20] A. Benba, A. Jilbab, A. Hammouch, Detecting patients with Parkinson's disease using mel frequency cepstral coefficients and support vector machines, *Int. J. Electr. Eng. Inform.* 7 (2) (2015) 297.
- [21] Z. Soumaya, B.D. Taoufiq, B. Nsiri, A. Abdelkrim, Diagnosis of parkinson disease using the wavelet transform and MFCC and SVM classifier, in: 2019 4th World Conference on Complex Systems, WCCS, IEEE, 2019, pp. 1–6.
- [22] G. Solana-Lavalle, J.-C. Galán-Hernández, R. Rosas-Romero, Automatic parkinson disease detection at early stages as a pre-diagnosis tool by using classifiers and a small set of vocal features, *Biocybern. Biomed. Eng.* 40 (1) (2020) 505–516.
- [23] Z. Soumaya, B.D. Taoufiq, N. Benayad, K. Yunus, A. Abdelkrim, The detection of parkinson disease using the genetic algorithm and SVM classifier, *Appl. Acoust.* 171 (2021) 107528.
- [24] A. Rahman, S.S. Rizvi, A. Khan, A. Afzaal Abbasi, S.U. Khan, T.-S. Chung, Parkinson's disease diagnosis in cepstral domain using MFCC and dimensionality reduction with SVM classifier, *Mob. Inf. Syst.* 2021 (1) (2021) 8822069.
- [25] S. Sharanya, P.N. Renjith, K. Ramesh, Classification of parkinson's disease using speech attributes with parametric and nonparametric machine learning techniques, in: 2020 3rd International Conference on Intelligent Sustainable Systems, ICIS, 2020, pp. 437–442.
- [26] S. Hawi, J. Alhozami, R. AlQahtani, D. AlSafran, M. Alqarni, L. El Sahmarany, Automatic Parkinson's disease detection based on the combination of long-term acoustic features and mel frequency cepstral coefficients (MFCC), *Biomed. Signal Process. Control.* 78 (2022) 104013.
- [27] R. Indu, S.C. Dimri, P. Malik, A modified kNN algorithm to detect Parkinson's disease, *Netw. Model. Anal. Heal. Inform. Bioinform.* 12 (1) (2023) 24.
- [28] M. Pramanik, R. Pradhan, P. Nandy, A.K. Bhoi, P. Barsocchi, Machine learning methods with decision forests for parkinson's detection, *Appl. Sci.* 11 (2) (2021) 581.
- [29] A. Ouhmida, A. Raihani, B. Cherradi, Y. Lamale, Parkinson's disease classification using machine learning algorithms: performance analysis and comparison, in: 2022 2nd International Conference on Innovative Research in Applied Science, Engineering and Technology, IRASET, IEEE, 2022, pp. 1–6.
- [30] H.-H. Zhang, L. Yang, Y. Liu, P. Wang, J. Yin, Y. Li, M. Qiu, X. Zhu, F. Yan, Classification of Parkinson's disease utilizing multi-edit nearest-neighbor and ensemble learning algorithms with speech samples, *Biomed. Eng. Online* 15 (2016) 1–22.
- [31] Y. Liu, Y. Li, X. Tan, P. Wang, Y. Zhang, Local discriminant preservation projection embedded ensemble learning based dimensionality reduction of speech data of Parkinson's disease, *Biomed. Signal Process. Control.* 63 (2021) 102165.
- [32] O. Barukab, A. Ahmad, T. Khan, M.R. Thayyil Kunhumuhammed, Analysis of Parkinson's disease using an imbalanced-speech dataset by employing decision tree ensemble methods, *Diagn.* 12 (12) (2022) 3000.
- [33] Q.W. Oung, S.N. Basah, H. Muthusamy, V. Vijean, H. Lee, Evaluation of short-term cepstral based features for detection of parkinson's disease severity levels through speech signals, in: IOP Conference Series: Materials Science and Engineering, Vol. 318, IOP Publishing, 2018, pp. 012–039.
- [34] H. Kuresan, D. Samiappan, S. Masunda, Fusion of WPT and MFCC feature extraction in Parkinson's disease diagnosis, *Technol. Health Care* 27 (4) (2019) 363–372.
- [35] N. Boualoulou, T. Belhoussine Drissi, B. Nsiri, Comparison of feature extraction methods between MFCC, BFCC, and GFCC with SVM classifier for Parkinson's disease diagnosis, in: International Conference on IoT Based Control Networks and Intelligent Systems, Springer, 2023, pp. 231–247.
- [36] I. Nissar, D.R. Rizvi, S. Masood, A.N. Mir, Voice-based detection of Parkinson's disease through ensemble machine learning approach: A performance study, *EAI Endorsed Trans. Pervasive Heal. Technol.* 5 (19) (2019) e2–e2.
- [37] G. Tallapareddy, D. Radha, Analysis of ensemble of machine learning algorithms for detection of Parkinson's disease, in: 2022 International Conference on Applied Artificial Intelligence and Computing, ICAIAC, 2022, pp. 354–361.
- [38] K.A. Shastry, An ensemble nearest neighbor boosting technique for prediction of Parkinson's disease, *Heal. Anal.* 3 (2023) 100181.
- [39] G. Celik, E. Başaran, 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 (2023) 109476.
- [40] A. Zlotnik, J.M. Montero, R. San-Segundo, A. Gallardo-Antolín, Random forest-based prediction of Parkinson's disease progression using acoustic, ASR and intelligibility features, in: Sixteenth Annual Conference of the International Speech Communication Association, 2015.
- [41] F.A. Torghabeh, S.A. Hosseini, E.A. Moghadam, Enhancing Parkinson's disease severity assessment through voice-based wavelet scattering, optimized model selection, and weighted majority voting, *Med. Nov. Technol. Dev.* 20 (2023) 100266.
- [42] W. Kim, Y. Kim, Abnormal gait recognition based on integrated gait features in machine learning, in: 2021 IEEE 45th Annual Computers, Software, and Applications Conference, COMPSAC, IEEE, 2021, pp. 1683–1688.
- [43] A. Zhao, L. Qi, J. Li, J. Dong, H. Yu, A hybrid spatio-temporal model for detection and severity rating of Parkinson's disease from gait data, *Neurocomputing* 315 (2018) 1–8.
- [44] T. Khan, A. Zeeshan, M. Dougherty, A novel method for automatic classification of Parkinson gait severity using front-view video analysis, *Technol. Health Care* 29 (4) (2021) 643–653.
- [45] F. Wahid, R.K. Begg, C.J. Hass, S. Halgamuge, D.C. Ackland, Classification of Parkinson's disease gait using spatial-temporal gait features, *IEEE J. Biomed. Heal. Inform.* 19 (6) (2015) 1794–1802.
- [46] S. Shetty, Y. Rao, SVM based machine learning approach to identify Parkinson's disease using gait analysis, in: 2016 International Conference on Inventive Computation Technologies, ICICT, Vol. 2, IEEE, 2016, pp. 1–5.
- [47] E. Abdulhay, N. Arunkumar, K. Narasimhan, E. Vellaiappan, V. Venkatraman, Gait and tremor investigation using machine learning techniques for the diagnosis of Parkinson disease, *Future Gener. Comput. Syst.* 83 (2018) 366–373.
- [48] R. Alkhatab, M.O. Diab, C. Corbier, M. El Badaoui, Machine learning algorithm for gait analysis and classification on early detection of Parkinson, *IEEE Sensors Lett.* 4 (6) (2020) 1–4.
- [49] E. Balaji, D. Brindha, R. Balakrishnan, Supervised machine learning based gait classification system for early detection and stage classification of Parkinson's disease, *Appl. Soft Comput.* 94 (2020) 106494.
- [50] J.M. Chavez, W. Tang, A vision-based system for stage classification of parkinsonian gait using machine learning and synthetic data, *Sensors* 22 (12) (2022) 4463.
- [51] B. Chen, C. Chen, J. Hu, Z. Sayeed, J. Qi, H.F. Darwiche, B.E. Little, S. Lou, M. Darwish, C. Foote, et al., Computer vision and machine learning-based gait pattern recognition for flat fall prediction, *Sensors* 22 (20) (2022) 7960.
- [52] J. Loureiro, P.L. Correia, Using a skeleton gait energy image for pathological gait classification, in: 2020 15th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2020, IEEE, 2020, pp. 503–507.
- [53] B. Muñoz-Ospina, D. Alvarez-García, H.J.C. Clavijo-Moran, J.A. Valderrama-Chaparro, M. García-Peña, C.A. Herrán, C.C. Urcuquí, A. Navarro-Cadavid, J. Orozco, Machine learning classifiers to evaluate data from gait analysis with depth cameras in patients with parkinson's disease, *Front. Hum. Neurosci.* 16 (2022) 826376.
- [54] N.K. Orphanidou, A. Hussain, R. Keight, P. Lishoa, J. Hind, H. Al-Askar, Predicting freezing of gait in parkinsons disease patients using machine learning, in: 2018 IEEE Congress on Evolutionary Computation, CEC, IEEE, 2018, pp. 1–8.
- [55] L. Gong, J. Li, M. Yu, M. Zhu, R. Clifford, A novel computer vision based gait analysis technique for normal and Parkinson's gaits classification, in: 2020 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress, DASC/PiCom/CBDCom/CyberSciTech, IEEE, 2020, pp. 209–215.
- [56] L. Borzl, L. Sigcha, G. Olmo, Context recognition algorithms for energy-efficient freezing-of-gait detection in Parkinson's disease, *Sensors* 23 (9) (2023) 4426.
- [57] N. Kleanthous, A.J. Hussain, W. Khan, P. Liatsis, A new machine learning based approach to predict freezing of gait, *Pattern Recognit. Lett.* 140 (2020) 119–126.
- [58] D. Impedovo, G. Pirlo, G. Vessio, Dynamic handwriting analysis for supporting earlier Parkinson's disease diagnosis, *Inf.* 9 (10) (2018) 247.
- [59] C.R. Pereira, D.R. Pereira, F.A. Da Silva, C. Hook, S.A. Weber, L.A. Pereira, J.P. Papa, A step towards the automated diagnosis of parkinson's disease: Analyzing handwriting movements, in: 2015 IEEE 28th International Symposium on Computer-Based Medical Systems, Ieee, 2015, pp. 171–176.
- [60] N.M. Ranjan, G. Mate, M. Bembde, Detection of Parkinson's disease using machine learning algorithms and handwriting analysis, *J. Data Min. Manag.* 8 (1) (2023) 21–29.
- [61] C. Kotsavasiloglou, N. Kostikis, D. Hristu-Varsakelis, M. Arnaoutoglou, Machine learning-based classification of simple drawing movements in Parkinson's disease, *Biomed. Signal Process. Control.* 31 (2017) 174–180.
- [62] R. Castrillón, A. Acién, J.R. Orozco-Arroyave, A. Morales, J. Vargas, R. Vera-Rodríguez, J. Fíerrez, J. Ortega-García, A. Villegas, Characterization of the handwriting skills as a biomarker for Parkinson's disease, in: 2019 14th IEEE International Conference on Automatic Face & Gesture Recognition, FG 2019, IEEE, 2019, pp. 1–5.
- [63] D.L. Guarín, J.K. Wong, N.R. McFarland, A. Ramirez-Zamora, Characterizing disease progression in Parkinson's disease from videos of the finger tapping test, *IEEE Trans. Neural Syst. Rehabil. Eng.* (2024).

- [64] N. Yang, D.-F. Liu, T. Liu, T. Han, P. Zhang, X. Xu, S. Lou, H.-G. Liu, A.-C. Yang, C. Dong, et al., Automatic detection pipeline for assessing the motor severity of Parkinson's disease in finger tapping and postural stability, *IEEE Access* 10 (2022) 66961–66973.
- [65] T. Yu, K.W. Park, M.J. McKeown, Z.J. Wang, Clinically informed automated assessment of finger tapping videos in Parkinson's disease, *Sensors* 23 (22) (2023) 9149.
- [66] T. Khan, D. Nyholm, J. Westin, M. Dougherty, A computer vision framework for finger-tapping evaluation in Parkinson's disease, *Artif. Intell. Med.* 60 (1) (2014) 27–40.
- [67] Q. He, H. Gao, An effective hand pose estimation based evaluation method in assessing Parkinson's finger tap movements, in: *International Symposium on Artificial Intelligence and Robotics*, Springer, 2023, pp. 511–518.
- [68] J.I. Serrano, J.P. Romero, A. Arroyo-Ferrer, M.D. del Castillo, Estimation of motor severity scales in Parkinson's disease by linear models of bimanual non-alternating index finger tapping features, *Expert Syst. Appl.* 246 (2024) 123077.
- [69] D. Todorov, A. Schnitzler, J. Hirschmann, Parkinsonian rest tremor can be distinguished from voluntary hand movements based on subthalamic and cortical activity, *Clin. Neurophysiol.* 157 (2024) 146–155.
- [70] M. Nahiduzzaman, M. Tasnim, N.T. Newaz, M.S. Kaiser, M. Mahmud, Machine learning based early fall detection for elderly people with neurological disorder using multimodal data fusion, in: *International Conference on Brain Informatics*, Springer, 2020, pp. 204–214.
- [71] J. Li, Y. Zhao, H. Zhang, W.J. LiMember, C. Fu, C. Lian, P. Shan, Image encoding and fusion of multi-modal data enhance depression diagnosis in Parkinson's disease patients, *IEEE Trans. Affect. Comput.* (2024).
- [72] X.-a. Bi, X. Hu, H. Wu, Y. Wang, Multimodal data analysis of Alzheimer's disease based on clustering evolutionary random forest, *IEEE J. Biomed. Heal. Inform.* 24 (10) (2020) 2973–2983.
- [73] M. Diwakar, P. Singh, A. Shankar, Multi-modal medical image fusion framework using co-occurrence filter and local extrema in NSST domain, *Biomed. Signal Process. Control.* 68 (2021) 102788.
- [74] Y. Jie, Y. Xu, X. Li, H. Tan, Tsjnet: A multi-modality target and semantic awareness joint-driven image fusion network, 2024, arXiv preprint [arXiv:2402.01212](https://arxiv.org/abs/2402.01212).
- [75] M. Diwakar, P. Singh, A. Shankar, S.R. Nayak, J. Nayak, S. Vimal, R. Singh, D. Sisodia, Directive clustering contrast-based multi-modality medical image fusion for smart healthcare system, *Netw. Model. Anal. Heal. Inform. Bioinform.* 11 (1) (2022) 15.
- [76] R. Dhaundiyal, A. Tripathi, K. Joshi, M. Diwakar, P. Singh, Clustering based multi-modality medical image fusion, in: *Journal of Physics: Conference Series*, Vol. 1478, IOP Publishing, 2020, 012024.
- [77] A. Zhao, J. Li, J. Dong, L. Qi, Q. Zhang, N. Li, X. Wang, H. Zhou, Multimodal gait recognition for neurodegenerative diseases, *IEEE Trans. Cybern.* 52 (9) (2022) 9439–9453.
- [78] R.M. Al-Tam, F.A. Hashim, S. Maqsood, L. Abualigah, R.M. Alwhaibi, Enhancing Parkinson's disease diagnosis through stacking ensemble-based machine learning approach, *IEEE Access* 12 (2024) 79549–79567.
- [79] M.B. Makarious, H.L. Leonard, D. Vitale, H. Iwaki, L. Sargent, A. Dadu, I. Violich, E. Hutchins, D. Saffo, S. Bandres-Ciga, et al., Multi-modality machine learning predicting Parkinson's disease, *Npj Parkinson's Dis.* 8 (1) (2022) 35.
- [80] W. Huo, P. Angeles, Y.F. Tai, N. Pavese, S. Wilson, M.T. Hu, R. Vaidyanathan, A heterogeneous sensing suite for multisymptom quantification of Parkinson's disease, *IEEE Trans. Neural Syst. Rehabil. Eng.* 28 (6) (2020) 1397–1406.
- [81] J.C. Vásquez-Correa, N. García-Ospina, J.R. Orozco-Arroyave, M. Cernak, E. Nöth, Phonological posteriors and GRU recurrent units to assess speech impairments of patients with Parkinson's disease, in: P. Sojka, A. Horák, I. Kopeček, K. Pala (Eds.), *Text, Speech, and Dialogue*, Springer International Publishing, 2018, pp. 453–461.
- [82] A. Hernandez, P.A. Pérez-Toro, E. Nöth, J.R. Orozco-Arroyave, A.K. Maier, S.H. Yang, Cross-lingual self-supervised speech representations for improved dysarthric speech recognition, 2022, ArXiv [abs/2204.01670](https://arxiv.org/abs/2204.01670).
- [83] S. Bhati, L.M. Velazquez, J. Villalba, N. Dehak, LSTM siamese network for Parkinson's disease detection from speech, in: *2019 IEEE Global Conference on Signal and Information Processing*, Globalsip, IEEE, 2019, pp. 1–5.
- [84] H. Gunduz, Deep learning-based Parkinson's disease classification using vocal feature sets, *IEEE Access* 7 (2019) 115540–115551.
- [85] J.C. Vásquez-Correa, J.R. Orozco-Arroyave, E. Nöth, Convolutional neural network to model articulation impairments in patients with Parkinson's disease, in: *Interspeech*, Stockholm, 2017, pp. 314–318.
- [86] H. Zhang, A. Wang, D. Li, W. Xu, Deepvoice: A voiceprint-based mobile health framework for parkinson's disease identification, in: *2018 IEEE EMBS International Conference on Biomedical & Health Informatics*, BHI, IEEE, 2018, pp. 214–217.
- [87] A. Zhao, N. Wang, X. Niu, M. Chen, H. Wu, A triplet multimodel transfer learning network for speech disorder screening of Parkinson's disease, *Int. J. Intell. Syst.* 2024 (1) (2024) 8890592.
- [88] D. Hemmerling, M. Wodzinski, J.R. Orozco-Arroyave, D. Sztaho, M. Daniol, P. Jemiolo, M. Wojcik-Pedziwiatr, Vision transformer for Parkinson's disease classification using multilingual sustained vowel recordings, in: *2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society*, EMBC, IEEE, 2023, pp. 1–4.
- [89] R. Nijhawan, M. Kumar, S. Arya, N. Mendiratta, S. Kumar, S. Towfek, D.S. Khafaga, H.K. Alkahtani, A.A. Abdelhamid, A novel artificial-intelligence-based approach for classification of Parkinson's disease using complex and large vocal features, *Biomimetics* 8 (4) (2023) 351.
- [90] H. Fang, C. Gong, C. Zhang, Y. Sui, L. Li, Parkinsonian Chinese speech analysis towards automatic classification of Parkinson's disease, in: *Machine Learning for Health*, PMLR, 2020, pp. 114–125.
- [91] S. Mehra, V. Ranga, R. Agarwal, A deep learning approach to dysarthric utterance classification with BiLSTM-GRU, speech cue filtering, and log mel spectrograms, *J. Supercomput.* (2024) 1–28.
- [92] S.-M. Jeong, S. Kim, E.C. Lee, H.J. Kim, Exploring spectrogram-based audio classification for Parkinson's disease: A study on speech classification and qualitative reliability verification, *Sensors* 24 (14) (2024) 4625.
- [93] O. Klempíř, D. Příhoda, R. Krupička, Evaluating the performance of wav2vec embedding for Parkinson's disease detection, *Meas. Sci. Rev.* 23 (6) (2023) 260–267.
- [94] M. Dhakal, A. Chhetri, A.K. Gupta, P. Lamichhane, S. Pandey, S. Shakya, Automatic speech recognition for the Nepali language using CNN, bidirectional LSTM and ResNet, in: *2022 International Conference on Inventive Computation Technologies*, ICICT, 2022, pp. 515–521.
- [95] O. Klempíř, R. Krupička, Analyzing wav2vec embedding in Parkinson's disease speech: A study on cross-database classification and regression tasks, *MedRxiv* (2024) 1–15, 2024-2004.
- [96] P.N. Chowdary, M.S. Akshay, V.S. Aravind, M.S. Aashish, G.V.V. Vardhan, G.J. Lal, A few-shot approach to dysarthric speech intelligibility level classification using transformers, in: *2023 14th International Conference on Computing Communication and Networking Technologies*, ICCCNT, IEEE, 2023, pp. 1–6.
- [97] H. Jiang, Self-Supervised Learning for Early Detection of Neurodegenerative Diseases with Small Data (Doctoral thesis), 2023.
- [98] F. Zhou, S. Hu, X. Wan, Z. Lu, J. Wu, Risevi: A disease risk prediction model based on vision transformer applied to nursing homes, *Electronics* 12 (15) (2023) 3206.
- [99] A. Zhao, L. Qi, J. Dong, H. Yu, Dual channel LSTM based multi-feature extraction in gait for diagnosis of neurodegenerative diseases, *Knowl.-Based Syst.* 145 (2018) 91–97.
- [100] H. Tian, X. Ma, H. Wu, Y. Li, Skeleton-based abnormal gait recognition with spatio-temporal attention enhanced gait-structural graph convolutional networks, *Neurocomputing* 473 (2022) 116–126.
- [101] J. Ajay, C. Song, A. Wang, J. Langan, Z. Li, W. Xu, A pervasive and sensor-free deep learning system for parkinsonian gait analysis, in: *2018 IEEE EMBS International Conference on Biomedical & Health Informatics*, BHI, IEEE, 2018, pp. 108–111.
- [102] I. El Maachi, G.-A. Bilodeau, W. Bouachir, Deep 1D-convnet for accurate Parkinson disease detection and severity prediction from gait, *Expert Syst. Appl.* 143 (2020) 113075.
- [103] Q. Zou, Y. Wang, Q. Wang, Y. Zhao, Q. Li, Deep learning-based gait recognition using smartphones in the wild, *IEEE Trans. Inf. Forensics Secur.* 15 (2020) 3197–3212.
- [104] A. Zhao, J. Li, J. Dong, L. Qi, Q. Zhang, N. Li, X. Wang, H. Zhou, Multimodal gait recognition for neurodegenerative diseases, *IEEE Trans. Cybern.* 52 (9) (2021) 9439–9453.
- [105] E. Balaji, D. Brindha, V.K. Elumalai, R. Vikrama, Automatic and non-invasive Parkinson's disease diagnosis and severity rating using LSTM network, *Appl. Soft Comput.* 108 (2021) 107463.
- [106] K.V.S.S. Kumar, I. Sirisha, K. Vathsalya, K.K.V.V. Vamsi, Parkinson disease diagnosis and severity rating prediction based on gait analysis using deep learning, *Int. Res. J. Adv. Sci. Hub* 5 (05) (2023) 418–425.
- [107] F.A. Torghabeh, Y. Modaresnia, S.A. Hosseini, An efficient tool for parkinson's disease detection and severity grading based on time-frequency and fuzzy features of cumulative gait signals through improved LSTM networks, *Med. Nov. Technol. Dev.* 22 (2024) 100297.
- [108] Y. Modaresnia, F.A. Torghabeh, S.A. Hosseini, A deep time-frequency approach in automated diagnosis of neurodegenerative diseases using gait signals, *Basic Clin. Neurosci.* 15 (6) (2024).
- [109] M. Chen, T. Ren, P. Sun, J. Wu, J. Zhang, A. Zhao, Fuselnet: Fusion of local and global information for detection of parkinson's disease, *Information* 14 (2) (2023).
- [110] D.M.D. Nguyen, M. Miah, G.-A. Bilodeau, W. Bouachir, Transformers for 1D signals in Parkinson's disease detection from gait, in: *2022 26th International Conference on Pattern Recognition*, ICPR, 2022, pp. 5089–5095.
- [111] M. Cheriet, V. Dentamaro, M. Hamdan, D. Impedovo, G. Pirlo, Multi-speed transformer network for neurodegenerative disease assessment and activity recognition, *Comput. Methods Programs Biomed.* 230 (2023) 107344.

- [112] H.-J. Sun, Z.-G. Zhang, Transformer-based severity detection of Parkinson's symptoms from gait, in: 2022 15th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics, CISB-BMEI, IEEE, 2022, pp. 1–5.
- [113] S. Naimi, W. Bouachir, G.-A. Bilodeau, HCT: Hybrid convnet-transformer for Parkinson's disease detection and severity prediction from gait, in: 2023 International Conference on Machine Learning and Applications, ICMLA, IEEE, 2023, pp. 814–819.
- [114] H. Wu, A. Zhao, Multi-level fine-tuned transformer for gait recognition, in: 2022 International Conference on Virtual Reality, Human-Computer Interaction and Artificial Intelligence, VRHICAI, IEEE, 2022, pp. 83–89.
- [115] H. Wu, Y. Liu, H. Yang, Z. Xie, X. Chen, M. Wen, A. Zhao, An attention-based temporal network for Parkinson's disease severity rating using gait signals, KSII Trans. Internet Inf. Syst. (TIIS) 17 (10) (2023) 2627–2642.
- [116] N. Kour, S. Gupta, S. Arora, Sensor technology with gait as a diagnostic tool for assessment of Parkinson's disease: a survey, Multimedia Tools Appl. (2022) 1–37.
- [117] S. Naimi, W. Bouachir, G.-A. Bilodeau, 1D-convolutional transformer for Parkinson disease diagnosis from gait, Neural Comput. Appl. 36 (4) (2024) 1947–1957.
- [118] M. Chen, T. Ren, P. Sun, J. Wu, J. Zhang, A. Zhao, Fuselnet: Fusion of local and global information for detection of Parkinson's disease, Information 14 (2) (2023) 119.
- [119] A. Moshkova, A. Samorodov, N. Voinova, A. Volkov, E. Ivanova, E. Fedotova, Parkinson's disease detection by using machine learning algorithms and hand movement signal from LeapMotion sensor, in: 2020 26th Conference of Open Innovations Association, FRUCT, Vol. 1, 2020, pp. 321–327.
- [120] L. Tong, J. He, L. Peng, CNN-based PD hand tremor detection using inertial sensors, IEEE Sensors Lett. 5 (7) (2021) 1–4.
- [121] A. Zhao, J. Li, Two-channel lstm for severity rating of parkinson's disease using 3d trajectory of hand motion, Multimedia Tools Appl. 81 (23) (2022) 33851–33866.
- [122] B. Ma, F. Zhang, B. Ma, [Retracted] self-attention-guided recurrent neural network and motion perception for intelligent prediction of chronic diseases, J. Heal. Eng. 2021 (1) (2021) 6382619.
- [123] R. Guo, H. Li, C. Zhang, X. Qian, A tree-structure-guided graph convolutional network with contrastive learning for the assessment of parkinsonian hand movements, Med. Image Anal. 81 (2022) 102560.
- [124] X. Peng, Y. Zhao, Z. Li, X. Wang, F. Nan, Z. Zhao, Y. Yang, P. Yang, Multi-scale and multi-level feature assessment framework for classification of Parkinson's disease state from short-term motor tasks, IEEE Trans. Biomed. Eng. (2024).
- [125] Y. Zhao, X. Wang, X. Peng, Z. Li, F. Nan, M. Zhou, J. Qi, Y. Yang, Z. Zhao, L. Xu, et al., Selecting and evaluating key MDS-UPDRS activities using wearable devices for Parkinson's disease self-assessment, IEEE J. Sel. Areas Sensors (2024).
- [126] M. Gazda, M. Hirš, P. Drotár, Multiple-fine-tuned convolutional neural networks for Parkinson's disease diagnosis from offline handwriting, IEEE Trans. Syst. Man Cybern.: Syst. 52 (1) (2021) 78–89.
- [127] N. Wang, X. Niu, Y. Yuan, Y. Sun, R. Li, G. You, A. Zhao, A coordinate attention enhanced swin transformer for handwriting recognition of Parkinson's disease, IET Image Process. 17 (9) (2023) 2686–2697.
- [128] C. Ma, P. Zhang, L. Pan, X. Li, C. Yin, A. Li, R. Zong, Z. Zhang, A feature fusion sequence learning approach for quantitative analysis of tremor symptoms based on digital handwriting, Expert Syst. Appl. 203 (2022) 117400.
- [129] S. Sabapathy, S. Maruthu, S.K. Krishnadhas, A.K. Tamilarasan, N. Raghavan, Competent and affordable rehabilitation robots for nervous system disorders powered with dynamic CNN and HMM, Intell. Syst. Rehabil. Eng. (2022) 57–93.
- [130] S. Sadhu, D. Solanki, N. Constant, V. Ravichandran, G. Cay, M.J. Saikia, U. Akbar, K. Mankodiya, Towards a telehealth infrastructure supported by machine learning on edge/fog for Parkinson's movement screening, Smart Heal. 26 (2022) 100351.
- [131] M. Gil-Martín, J.M. Montero, R. San-Segundo, Parkinson's disease detection from drawing movements using convolutional neural networks, Electron. 8 (8) (2019).
- [132] A. Zhao, J. Dong, J. Li, L. Qi, H. Zhou, Associated spatio-temporal capsule network for gait recognition, IEEE Trans. Multimed. 24 (2022) 846–860.
- [133] M. Mohaghegh, J. Gascon, Identifying parkinson's disease using multimodal approach and deep learning, in: 2021 6th International Conference on Innovative Technology in Intelligent System and Industrial Applications, CITISIA, 2021, pp. 1–6.
- [134] J. Ma, Y. Zhang, Y. Li, L. Zhou, L. Qin, Y. Zeng, P. Wang, Y. Lei, Deep dual-side learning ensemble model for Parkinson speech recognition, Biomed. Signal Process. Control. 69 (2021) 102849.
- [135] A. Zhao, Y. Wang, L. Jianbo, Transferable self-supervised instance learning for sleep recognition, IEEE Trans. Multimed. PP (2023) 1–1.
- [136] C. Wang, L. Peng, Z.-G. Hou, Y. Li, Y. Tan, H. Hao, A hierarchical architecture for multisymptom assessment of early parkinson's disease via wearable sensors, IEEE Trans. Cogn. Dev. Syst. 14 (4) (2021) 1553–1563.
- [137] M. Junaid, S. Ali, F. Eid, S. El-Sappagh, T. Abuhmed, Explainable machine learning models based on multimodal time-series data for the early detection of Parkinson's disease, Comput. Methods Programs Biomed. 234 (2023) 107495.
- [138] G. Pahuja, B. Prasad, Deep learning architectures for Parkinson's disease detection by using multi-modal features, Comput. Biol. Med. 146 (2022) 105610.
- [139] J. Li, W. Liang, X. Yin, J. Li, W. Guan, Multimodal gait abnormality recognition using a convolutional neural network-bidirectional long short-term memory (CNN-BiLSTM) network based on multi-sensor data fusion, Sensors 23 (22) (2023) 9101.
- [140] N. Faiem, T. Asuroglu, K. Acici, A. Kallonen, M. Van Gils, Assessment of Parkinson's disease severity using gait data: A deep learning-based multimodal approach, in: Nordic Conference on Digital Health and Wireless Solutions, Springer, 2024, pp. 29–48.
- [141] B. Wang, X. Hu, R. Ge, C. Xu, J. Zhang, Z. Gao, S. Zhao, K. Polat, Prediction of freezing of Gait in Parkinson's disease based on multi-channel time-series neural network, Artif. Intell. Med. 154 (2024) 102932.
- [142] A. Zhao, H. Wu, M. Chen, N. Wang, A spatio-temporal siamese neural network for multimodal handwriting abnormality screening of Parkinson's disease, Int. J. Intell. Syst. 2023 (1) (2023) 9921809.
- [143] C. Xue, S.S. Kowshik, D. Lteif, S. Puducher, V.H. Jasadanand, O.T. Zhou, A.S. Walia, O.B. Guney, J.D. Zhang, S.T. Pham, et al., AI-based differential diagnosis of dementia etiologies on multimodal data, Nature Med. (2024) 1–13.
- [144] M. Das, D. Gupta, A. Bakde, An end-to-end content-aware generative adversarial network based method for multimodal medical image fusion, in: Data Analytics for Intelligent Systems: Techniques and Solutions, IOP Publishing Bristol, UK, 2024, pp. 7–71.
- [145] M. Little, Parkinsons, UCI machine learning repository, 2008, <http://dx.doi.org/10.24432/C59C74>.
- [146] K. Olcay, S. Betul, I. M., S. C., S. Ahmet, G. Fikret, Parkinson's speech with multiple types of sound recordings, UCI machine learning repository, 2014, <http://dx.doi.org/10.24432/C5NC8M>.
- [147] H. Jaeger, D. Trivedi, M. Stadtschnitzer, Mobile device voice recordings at king's college London (MDVR-KCL) from both early and advanced parkinson's disease patients and healthy controls, 2019.
- [148] G. Dimauro, F. Girardi, Italian Parkinson's voice and speech, 2019.
- [149] S. C., S. Gorkem, G. Aysegul, N. Hatice, S. Betul, Parkinson's disease classification, UCI machine learning repository, 2018, <http://dx.doi.org/10.24432/C5MS4X>.
- [150] C.R. Pereira, S.A.T. Weber, C. Hook, G.H. Rosa, J.P. Papa, Deep learning-aided parkinson's disease diagnosis from handwritten dynamics, in: 2016 29th SIBGRAPI Conference on Graphics, Patterns and Images, SIBGRAPI, 2016, pp. 340–346.
- [151] M.E. Isenkul, B.E. Sakar, O. Kursun, Improved spiral test using digitized graphics tablet for monitoring parkinson's disease, in: The 2nd International Conference on E-Health and TeleMedicine-ICEHTM 2014, 2014, 1–1.
- [152] K. Olcay, Improved spiral test using digitized graphics tablet for monitoring Parkinson's disease, UCI machine learning repository, 2016, <http://dx.doi.org/10.24432/C5HW3N>.
- [153] W.R. Adams, High-accuracy detection of early Parkinson's disease using multiple characteristics of finger movement while typing, PLOS ONE 12 (11) (2017) 1–20.
- [154] L. Giancardo, Computer keyboard interaction as an indicator of early Parkinson's disease, Sci. Rep. 6 (10) (2016) 34468.
- [155] V. J., B. A., F. M., C. van Alen, L. Plagwitz, W. T., Machine learning in the Parkinson's disease smartwatch (PADS) dataset, Npj Park. 9 (215–220) (2024) 6.
- [156] NIGMS, physionet PDgait database.
- [157] NIGMS, physionet ndds database.
- [158] A. Salomon, E. Gazit, P. Ginis, B. Urazalinov, H. Takoi, T. Yamaguchi, S. Goda, D. Lander, J. Lacombe, A. Sinha, A. Nieuwboer, L. Kirsch, R. Holbrook, B. Manor, J. Hausdorff, A machine learning contest enhances automated freezing of gait detection and reveals time-of-day effects, Nat. Commun. 15 (2024).
- [159] R. Daniel, P. Meir, H. Jeff, Daphnet freezing of gait , UCI machine learning repository, 2013, <http://dx.doi.org/10.24432/C56K78>.
- [160] NIGMS, physionet multimodal database.
- [161] C. Morgan, E.L. Tonkin, A. Masullo, F. Jovan, A. Sikdar, P. Khaire, M. Mirmehd, R. McConville, G.J. Tourte, A. Whone, et al., A multimodal dataset of real world mobility activities in Parkinson's disease, Sci. Data 10 (1) (2023) 918.
- [162] H. Li, Multimodal dataset of freezing of gait in parkinson's disease, Mendeley Data 3 (2021) 2021.
- [163] C. Ribeiro De Souza, R. Miao, J. Ávila De Oliveira, A. Cristina De Lima-Pardini, D. Fragoso De Campos, C. Silva-Batista, L. Teixeira, S. Shokur, B. Mohamed, D.B. Coelho, A public data set of videos, inertial measurement unit, and clinical scales of freezing of gait in individuals with parkinson's disease during a turning-in-place task, Front. Neurosci. 16 (2022) 832463.
- [164] X. Zhang, X. Zhou, M. Lin, J. Sun, Shufflenet: An extremely efficient convolutional neural network for mobile devices, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 6848–6856.
- [165] S.-H. Gao, M.-M. Cheng, K. Zhao, X.-Y. Zhang, M.-H. Yang, P. Torr, Res2net: A new multi-scale backbone architecture, IEEE Trans. Pattern Anal. Mach. Intell. 43 (2) (2019) 652–662.

- [166] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, S. Xie, A convnet for the 2020s, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 11976–11986.
- [167] M. Tan, Q. Le, Efficientnetv2: Smaller models and faster training, in: International Conference on Machine Learning, PMLR, 2021, pp. 10096–10106.
- [168] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, et al., An image is worth 16x16 words: Transformers for image recognition at scale, 2020, arXiv preprint arXiv:2010.11929.
- [169] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, B. Guo, Swin transformer: Hierarchical vision transformer using shifted windows, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 10012–10022.
- [170] S. Mehta, M. Rastegari, Mobilevit: light-weight, general-purpose, and mobile-friendly vision transformer, 2021, arXiv preprint arXiv:2110.02178.
- [171] J. Guo, K. Han, H. Wu, Y. Tang, X. Chen, Y. Wang, C. Xu, Cmt: Convolutional neural networks meet vision transformers, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 12175–12185.
- [172] W. Yu, M. Luo, P. Zhou, C. Si, Y. Zhou, X. Wang, J. Feng, S. Yan, Metaformer is actually what you need for vision, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 10819–10829.
- [173] X. Niu, N. Wang, Y. Liu, A. Zhao, A coordinate enhanced local and global attention network for handwriting detection of parkinson's disease, in: 2023 International Conference on Computer, Internet of Things and Smart City, CloTSC, IEEE, 2023, pp. 72–79.