JRITM
JOURNAL OF RESEARCH AND INNOVATION IN TECHNOLOGY, COMMERCE AND MANAGEMENT

Journal of Research and Innovation in Technology, Commerce and Management
Vol. 3 Issue 9, September 2026, pp. 39054-39065
ISSN: 3049-3129 (Online)

Decoding Parkinson's Disease from EEG Signals: A Systematic Review of Machine Learning and Deep Learning Frameworks

Dilpal Singh Soni
UIET, Panjab University, Chandigarh
Email: dilpalsoni@gmail.com

Sukesha Sharma
UIET, Panjab University, Chandigarh
Email: sukesha@pu.ac.in

Cite as: Dilpal Singh Soni, & Sukesha Sharma. (2026). Decoding Parkinson's Disease from EEG Signals: A Systematic Review of Machine Learning and Deep Learning Frameworks. Journal of Research and Innovation in Technology, Commerce and Management, Vol. 3(Issue 9), 39054–39065. https://doi.org/10.5281/zenodo.22702786

DOI: https://doi.org/10.5281/zenodo.22702786

Abstract. Parkinson's disease (PD) is the second most common neurodegenerative disorder worldwide, yet it continues to be diagnosed primarily through subjective clinical evaluation - a process that is slow, inconsistent across observers, and largely blind to early-stage disease. This paper reviews some studies published between 2021 and 2026, drawn from IEEE Xplore, IOP Publishing, Elsevier, and Springer, all of which examine automated PD diagnosis using EEG signals combined with machine learning (ML) and deep learning (DL). Across these studies, a regular methodological pattern emerges: Researchers convert raw 1D EEG time-series into 2D time-frequency representations (spectrograms) using transforms such as Gabor, Morlet Wavelet, Wavelet Scattering (WST), Smoothed Pseudo-Wigner-Ville Distribution (SPWVD), and the S-transform. Then classify the resulting images using hybrid deep learning models. Hybrid architectures particularly CNN-Transformer pipelines and Densely Linked Bidirectional LSTM (DLBLSTM) networks consistently achieve accuracies above 95%, sometimes reaching 99.97%. They outperformed standalone CNNs by roughly 4-6% and traditional ML methods by up to 9%. A key finding of this review is that Explainable AI (XAI), when combined with automated channel selection, can reduce the number of EEG electrodes by 50% with almost no loss in diagnostic accuracy, which has significant implications for portable, wearable screening devices. That said, three recurring challenges continue to limit clinical translation: datasets remain small and often homogeneous; the volume conduction effect distorts functional connectivity measurements; and the opaqueness of deep learning models undermines clinician trust. We further examined an emerging body of Graph Neural Network studies which addressed the connectivity-modelling and volume-conduction challenges directly. We concluded this review by identifying Graph Neural Networks for brain-

connectivity modeling, multimodal data fusion, and XAI-driven saliency pipelines as the most promising directions for achieving clinically trustworthy, individualized PD diagnostics.

Keywords: Parkinson's Disease, EEG, CNN-Transformer pipeline, Densely Linked Bidirectional LSTM, Graph Neural Networks, Explainable AI, Time-Frequency Representation

1. INTRODUCTION

Parkinson's disease is the second most prevalent neurodegenerative disorder in the world. It predominantly affects people over the age of 60, and its defining pathological feature is the gradual loss of dopaminergic neurons in the substantia nigra pars compacta region. This region is deep within the basal ganglia and plays a central role in coordinating movement [1]. As these neurons degenerate and dopamine levels fall, the regulatory balance between the brain's direct and indirect motor pathways breaks down [2], giving rise to the characteristic cluster of symptoms associated with Parkinson's Disease.

These symptoms fall into two broad groups. Motor symptoms like resting tremor, muscle rigidity, slowness of movement (bradykinesia), postural instability, and gait disturbance which tend to be the most visible and are usually what prompts an initial clinical evaluation. But PD also carries a substantial non-motor burden like gastrointestinal dysfunction, chronic pain, disrupted sleep, and depression are all well-documented, and can be every bit as debilitating as the motor features [3]. The underlying cause of PD remains incompletely understood. Genetics, advancing age, and prolonged exposure to certain environmental toxins are all thought to contribute, though no single factor explains the full picture [4].

The current clinical standard for diagnosis relies on structured assessments like the Unified Parkinson's Disease Rating Scale (UPDRS). The limitations of this approach are widely acknowledged: it is inherently subjective, sensitive to observer bias, and most importantly ineffective at detecting PD before motor symptoms become apparent. By the time a formal diagnosis is made, substantial and largely irreversible neuronal damage has already occurred. This matters enormously because early intervention be it pharmacological or otherwise is considerably more effective at slowing disease progression than treatment initiated at later stages. Surgical approaches such as Deep Brain Stimulation (DBS) can be highly effective for managing advanced motor symptoms [5] but they require careful pre-operative evaluation of cognitive status as some patients experience neuropsychiatric changes following the procedure [6]. All of this makes objective, early, and reliable PD assessment not just a scientific question but a pressing clinical need [7].

Electroencephalography (EEG) offers a particularly promising path toward meeting that need. It is a non-invasive, relatively affordable and capable of capturing the brain's electrical activity with high temporal resolution, and has long been the diagnostic standard for epilepsy [9]. There is growing evidence that it can detect PD-related neurophysiological changes even in the early stages of the disease. In PD specifically, the brain exhibits what researchers describe as a cortical slowing effect: a measurable shift in oscillatory activity away from the faster alpha frequencies and toward slower theta and delta rhythms, accompanied by reduced beta-band activity [13]. These changes are detectable before motor symptoms are fully noticeable and offer a quantifiable, objective window into PD pathology that conventional clinical tools simply cannot match.

The practical challenge is the interpretation. Analyzing EEG data by hand is time-consuming, technically demanding, and prone to human error. This is precisely what has motivated the growing use of machine learning and deep learning for automated EEG analysis. Deep learning, as a sub-field of machine learning employs hierarchical neural network architectures to extract increasingly abstract features from raw data without relying on manually engineered inputs [11, 12]. This capability is especially valuable for EEG, where disease-related patterns are non-stationary, non-linear, and distributed across many channels simultaneously [8].

Several reviews have examined ML-based PD detection using biomarkers like handwriting, voice,

and gait. A focused and critical synthesis of EEG-specific deep learning frameworks is harder to find. Existing reviews rarely examine why certain pipeline choices consistently outperform others or what the actual implications are for clinical deployment. This review is an attempt to address that gap directly. Its specific contributions are:

- A systematic comparative analysis of eight studies (2021–2026), organized around a new taxonomy of EEG-based PD detection approaches.
- An account of the dominant methodological shift from 1D feature extraction toward 2D time-frequency imaging combined with hybrid deep learning classifiers.
- Quantitative comparison of hybrid model performance against traditional ML approaches.
- A synthesis of persistent challenges and future directions, with particular attention to clinical translation.
- A focused review of the emerging GNN literature for EEG-based PD detection, identifying the volume-conduction, architectural, and explainability gaps that define the field's next steps.

2. Literature Review

The review covered four major academic databases: IEEE Xplore, IOP Publishing, Elsevier, and Springer. The search used combinations of the following keywords and Boolean operators:

- "Parkinson's disease"
- "Machine learning"
- "Deep learning"
- "Artificial intelligence"
- "Parkinson's diagnosis"
- "Parkinson's biomarkers"
- "Parkinson's early diagnosis"
- "Parkinson's early diagnosis EEG Data"

After applying inclusion and exclusion criteria, a final set of eight studies published between 2021 and 2026 were selected. It should be noted that this primary search protocol was designed to capture deep-learning classification frameworks applied to EEG and did not include graph- or connectivity-specific search terms. A supplementary targeted search - adding the terms "graph neural network," "graph convolutional network," "graph attention network," and "functional connectivity" - was subsequently conducted to identify the emerging graph-based literature. Because these studies adopt a fundamentally different methodological paradigm from the image-based classifiers that form the core of this review, they are examined separately as an emerging direction in Section 6 rather than incorporated into the comparative synthesis of the eight selected studies. Table 1 provides a comparative summary covering the signal transformation technique, model architecture, intended application, reported accuracy, and identified limitations of each work.

Table 1. Comparative Summary of Reviewed Studies on EEG-Based PD Detection

Study	Summary	Technique Used	Application	Results & Accuracy	Drawbacks & Future Scope
Khare et al. (2021) [15]	PDCNNet framework: 1D EEG converted to 2D images via Smoothed Pseudo Wigner-Ville Distribution (SPWVD), then classified using a custom CNN.	Bandpass filter (0.5-45 Hz), ICA. SPWVD-generated Time Frequency representations (TFR) images (227×227 px) fed into CNN (PDCNNet). Ten-fold cross-validation.	Diagnostic screening for brain dysfunction in PD patients and for assessing DBS eligibility by evaluating pre-operative cognitive state.	Dataset 1: 100% (HC vs. PDSO), 99.84% (HC vs. PDSP). Dataset 2: 99.97% accuracy. Outperforms prior state-of-the-art on both datasets.	Short 2-second EEG segments; small training sample. Future: larger diverse datasets, extension to depression, autism, Alzheimer's, and other disorders.
Obayya et al. (2023) [16]	Densely Linked Bidirectional LSTM (DLBLSTM) with Gabor Transform spectrograms for automated PD detection.	Bandpass filter (0.5-100 Hz), ICA. Gabor Transform generates spectrograms (2D TFR). DLBLSTM models temporal patterns bidirectionally and aggregates hidden states. Six-fold cross-validation.	Automated clinical diagnosis of PD, reducing reliance on subjective symptom evaluation and manual EEG interpretation.	99.6% test accuracy (100% training). PD sensitivity/precision: 99%; specificity: 100%. Outperforms CNN (93.6%), LSTM (95.8%), BiLSTM (96.2%) on same data.	Only 15 PD + 16 HC subjects; 100% training accuracy indicates overfitting risk. Black-box nature limits clinical interpretability. Future: larger datasets, model explainability.
Sharma et al. (2025) [7]	EEG signals were decomposed into frequency bands (alpha, beta, theta and delta) and using Morlet Wavelet Transform time frequency features extracted machine learning classifiers used to identify four cognitive impairment levels in PD patients.	Bandpass filter (0.5-45 Hz), ICA for artifacts. 1D-to-2D via Morlet Wavelet Transform, Power Spectral Density (PSD) features across delta/theta/alpha/beta. Classifiers: Decision Tree, Random Forest, XGBoost. Welch's ANOVA for statistical validation.	Fast, non-invasive cognitive screening for DBS eligibility. Identifies frontopolar, frontal, and parietal regions as key monitoring areas.	XGBoost: 95.43% accuracy, sensitivity 0.94, specificity 0.98. Delta/theta power increases with impairment and alpha/beta decreases. Frontopolar/frontal/parietal 1 feature importance >50%.	Small cohort. Volume conduction effect may produce false connectivity readings. Future: GNNs with deep learning, directional connectivity techniques.
Mukherjee & Halder Roy (2025) [17]	Hybrid CNN-Transformer with WST-derived TFR images. Generative Adversarial Network (GAN) - based augmentation doubles dataset size.	Bandpass (0.1-50 Hz), ICA. 250 ms segments. Wavelet Scattering Transform (Morlet wavelet) to 2D maps. GAN augmentation (synthetic maps). CNN + 8-head attention + Tanh-LSTM (TLSTM) classification.	Early PD screening and remote/telemedicine monitoring, enabling at-home assessment for high-risk patients.	99.52% accuracy, precision 99.58%, recall 99.91%, F1 99.77%, AUC ~99.95% (PD), ~99.99% (HC).	GAN data may not capture real biological variability; high computational overhead. Future: multimodal fusion (EMG, fNIRS, MRI), architectural improvements.
Mollazadeh Azari et al. (2025) [18]	Investigates time-perception distortion in PD via EEG during an interval-timing task; MLP-ANN classifier on nonlinear dynamic features.	64-channel EEG. Bandpass (0.1–60 Hz) + 50 Hz notch filter, ICA. Wavelet decomposition into five rhythms. Features: Correlation Dimension (CD), Fractal Dimension (FD), Lyapunov Exponent (Lyap), Shannon, Approximate, PSD. 57 selected via statistical testing). MLP-ANN, 10-fold CV.	Early detection of non-motor/cognitive PD deficits; identifying interval-timing impairment linked to basal ganglia-frontocortical network disruption.	90% average test accuracy; 89% sensitivity, 87% F1-score; 95% specificity. Confirms disrupted basal ganglia-posterior parietal connectivity underlies temporal processing deficits.	Small sample (80 PD, 40 HC); only ON medication recordings. Future: larger datasets, ON/OFF comparison, combined EEG + fMRI, anima model validation.
Kolla et al. (2025) [19]	CortiNet-Net: hybrid deep learning framework (Stacked Autoencoder + MobileNetV2) with novel SC-Tune hyperparameter optimization for EEG-based PD diagnosis.	Bandpass filter. Time-domain (mean, variance, ZCR, RMS), frequency-domain delta/theta/alpha/beta power), and nonlinear (entropy, kurtosis, Hurst exponent) features. Fused feature vector → CortiNet-Net. SC-Tune auto-optimizes CNN stride/padding.	Real-time clinical decision support for neurological diagnostics; scalable due to lightweight MobileNet and non-invasive EEG.	Accuracy, precision, recall, F1-score all 0.99 (99%). Minimum error rate 0.01; minimum loss 0.05; overall efficiency 0.95.	Extended training time; small dataset. Future: multimodal data fusion.
Saqib et al. (2025) [20]	Intelligent Neuro-Supervised Deep Learning Network (INSDLN) architecture combining RNNs with MLPs; trained on mathematically-simulated PD brain rhythms generated from nonlinear ODEs.	Simulated EEG via nonlinear ODEs → Adams/NDSolve solver. Denoised with MFCA. 64/128/64 recurrent layers + Tanh activation → Softmax output. Trained with Adam optimizer. Compared: LMBPA vs. Bayesian Regularization.	Advances diagnosis, prognostic tracking, and treatment planning for PD. Also applicable to broader bio signal analysis (EEG, EMG) and medical imaging.	Near-perfect regression (MSE: 5.86×10⁻⁸). BRM more stable for noisy data; LMBPA faster but overfits noise. Both outperform large Transformer/LSTM on shorter sequences.	Uses simulated rather than real patient EEG. LMBPA overfits noisy real-world data. Future: multimodal DL (EEG + MRI + records), XAI integration, fractional gradient algorithms.
Afonso et al. (2025) [21]	XAI-driven ARI-ECS framework: automated EEG channel selection using CNN saliency maps reduces channels from 32 to 16 while maintaining near-identical accuracy.	Bandpass (0.05-60 Hz). S-transform → 2D STA grayscale matrices → Deep CNN (S-CovNet). XAI saliency maps via gradient backpropagation. ARI-ECS ranks and selects optimal channel subset. Validated on SD and UNM datasets.	Non-invasive early PD detection with reduced hardware requirements paves way for compact, wearable, and handheld diagnostic tools.	ARI-ECS: 32 → 16 channels (50% reduction). Test accuracy: 98% (SD), 96% (UNM). Comparable to full 32-channel accuracy (98% SD, 97% UNM).	Limited public datasets (42 total PD cases); residual medication effects in 'OFF' EEG. Future: new data integration, extension to other neurodegenerative disorders.

3. General Techniques Used Across the Reviewed Papers

Looking across all eight studies, it is possible to identify common methodological threads that cut through most of the individual approaches. Figure 1 below provides the General EEG-based PD detection pipeline. Further the sections below trace three key dimensions: how signals are cleaned, how features are extracted or transformed, and what classification architectures are applied.

General EEG-Based PD Detection Pipeline

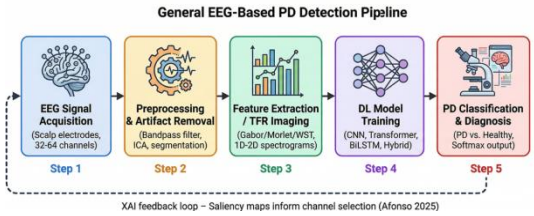

Fig. 1. *General EEG-based Parkinson's Disease detection pipeline, from signal acquisition through preprocessing, feature extraction (1D-to-2D transformation), model training, and final classification, with an XAI feedback loop for channel selection.*

3.1. Signal Cleaning and Artifact Removal

The preprocessing strategies across the reviewed studies can be grouped into three broad categories

ICA (Independent Component Analysis) based Preprocessing: This was the most common approach used by Sharma et al. [7], Obayya et al. [16], Mukherjee and Halder Roy [17], and Mollazadeh Azari et al. [18]. It combines bandpass filtering with Independent Component Analysis (ICA). The bandpass filter have cutoffs generally ranging from 0.1 Hz on the low end to 100 Hz on the high end, removes low-frequency drift and high-frequency noise. ICA then decomposes the EEG into statistically independent components further allowing the removal of specific noise sources such as eye blinks and muscle activity. Sharma et al. [7] specifically targeted channels Fp1 and Fp2 for blink detection, while Mollazadeh Azari et al. [18] additionally addressed cardiac (ECG) artifact removal. This step is often overlooked but can introduce meaningful noise, particularly in older participant groups.

Manual and Hybrid Cleaning: The second approach taken by Khare et al. [15] and Afonso et al. [21], combines manual inspection with standard preprocessing pipelines. Artifacts from muscle activity, electrical interference and movement are identified and removed by hand with high-pass and bandpass filters and epoch segmentation. This

process is applied to ensure only clean segments reach the classifier. This approach can be more thorough in principle but it is also more time-consuming and harder to standardize across clinical sites.

Simulated or High-Level Filtering: The third category is quite distinct. Saqib et al. [20] generated their training data mathematically using nonlinear ordinary differential equations to simulate brain activity, so conventional biological artifact removal was unnecessary. Stochastic solvers and model optimization handled signal variability instead. Kolla et al. [19] also sit apart from the ICA-based approaches by focusing on high-level filtering, normalization, and peak detection rather than component-level biological decomposition.

3.2. Feature Engineering and Transformations

Advanced Time-Frequency Imaging (2D Representations): The clearest trend across the reviewed papers is the near-universal preference for converting 1D EEG time-series into 2D time-frequency representations (TFRs). Once EEG data takes the form of an image, it can be processed by convolutional neural network architectures that have been extensively developed and refined for visual pattern recognition and which outclass at detecting spatially distributed features. This 1D-to-2D conversion has become the dominant paradigm in EEG-based PD research, and for good reason.

Several papers converge on this strategy while employing different transform methods. Khare et al. [15] used the Smoothed Pseudo-Wigner-Ville Distribution (SPWVD) which generates high-resolution time-frequency plots with reduced cross-term interference compared to standard Wigner-Ville methods. Obayya et al. [16] applied the Gabor transform, producing spectrograms that capture how brain activity's frequency content evolves over time. Afonso et al. [21] used the Stockwell transform (S-transform) which converted the resulting amplitude matrices to grayscale images for CNN input. Mukherjee and Halder Roy [17] took the Wavelet Scattering Transform (WST) route and then augmented the resulting image dataset using a Generative Adversarial Network, effectively doubling their training data through synthetic generation.

Signal-to-Feature Decomposition (Manual/Mathematical): Not every study took the imaging route. Sharma et al. [7], Mollazadeh Azari et al. [18], and Kolla et al. [19] instead decomposed the EEG into standard brainwave frequency bands and extracted mathematical features directly. Sharma et al. used Morlet wavelet decomposition to estimate average spectral power across specific brain regions. Mollazadeh Azari et al. went considerably further by computing a suite of nonlinear dynamic features including Fractal Dimension, Lyapunov Exponents, Shannon Entropy, and Approximate Entropy. Kolla et al. took a mixed approach by combining statistical time-domain features with frequency-domain band power measures and nonlinear indices such as entropy and the Hurst exponent.

Automated Temporal Modeling (End-to-End): Saqib et al. [20] took yet another direction entirely. Rather than transforming signals into images or hand-crafting features, they used an end-to-end architecture. This architecture learns directly from raw simulated time-series data through recurrent layers. This is a reminder that even within a field that has largely converged on 2D imaging, there are still meaningful alternative approaches being explored.

3.3. Algorithm Architectures

Traditional ML and shallow networks: Two of the eight studies rely on established supervised learning algorithms applied to pre-extracted features. Sharma et al. [7] compared Decision Trees, Random Forests, and XGBoost for classifying cognitive states in PD patients. Mollazadeh Azari et al. [18] used a Multilayer Perceptron (MLP) trained on their suite of nonlinear features. These approaches are computationally efficient and more interpretable than deep learning alternatives but they do not match the accuracy of the more complex architectures reviewed.

Deep learning and hybrid frameworks: The majority of the reviewed papers deploy sophisticated deep learning architectures and it is this category that consistently achieves the highest reported accuracies. Obayya et al. [16] introduced the Densely Linked Bidirectional LSTM (DLBLSTM), which processes sequential EEG data in both forward and backward directions and aggregates hidden states across all layers allowing the model to capture temporal dependencies that a standard unidirectional LSTM would miss. Saqib et al. [20] combined a Recurrent Neural Network backbone with a Multi-Layer Perceptron for pattern recognition in simulated signals.

On the CNN side, Khare et al. [15] trained a custom convolutional network directly on SPWVD-transformed images. Mukherjee and Halder Roy [17] built a more involved pipeline: a CNN that first extracts spatial features from the WST images which are then passed to an 8-head self-attention Transformer module for capturing global temporal context. This is followed by a tanh-LSTM for final classification. Kolla et al. [19] combined MobileNetV2 which is a lightweight architecture originally designed for resource-constrained mobile applications combined with stacked autoencoders and introduced the Snow Shepherd Stride Configuration Tuning (S^3C-Tune) method to automatically optimize the CNN's stride and padding hyperparameters.

Explainable AI (XAI) and automated optimization: The most clinically oriented paper in the set is Afonso et al. [21], which focuses not just on accuracy but on transparency and interpretability. Their model generates saliency maps through gradient backpropagation, visually indicating which parts of the input spectrogram most influenced each classification decision. An Average Region Intensity (ARI) scoring method then ranks EEG channels by their diagnostic significance and selects an optimal subset. The result is that input dimensionality drops from 32 channels to 16 which is a 50% reduction with accuracy that is almost indistinguishable from using the full channel set. This is a genuinely useful finding for anyone thinking about portable diagnostic hardware.

Figure 2 summarizes this landscape as a taxonomy, grouping methods into traditional ML, standalone deep learning, and hybrid/XAI frameworks.

Taxonomy of ML/DL Approaches for EEG-Based PD Detection

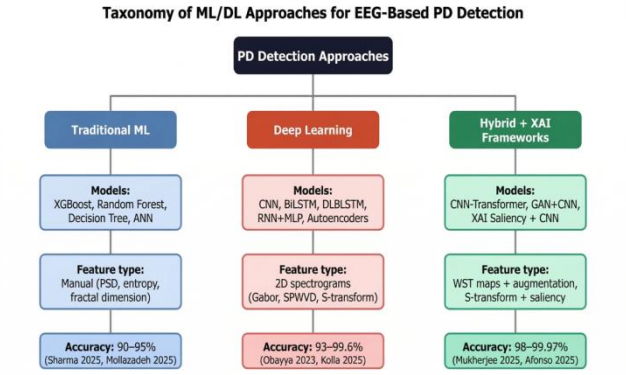

Fig. 2. *Taxonomy of ML/DL approaches used for EEG-based PD detection, categorized into Traditional ML, standalone Deep Learning, and Hybrid/XAI frameworks, with representative models and accuracy ranges*

4. Key Learnings and Comparative Insights

The most consistent theme running through these studies is a methodological shift away from manually designed feature extraction and toward automated, image-based processing using hybrid deep learning. Traditional approaches like XGBoost, Random Forest, and MLP classifiers still deliver competitive results as XGBoost reaches 95.43% in Sharma et al. [7] but they trail the deep learning architectures across the board. Figure 3 makes this performance gap visually clear.

Classification Accuracy Comparison Across Reviewed Studies

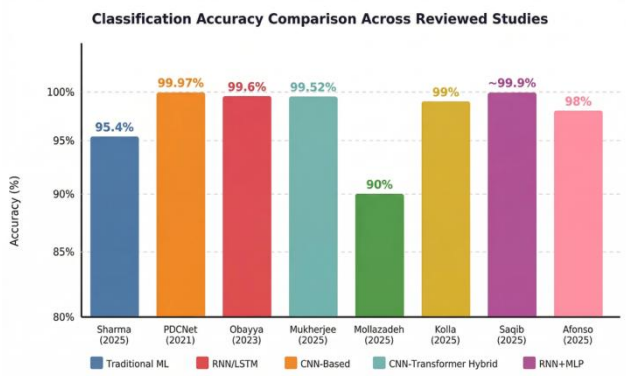

Fig. 3. *Comparison of classification accuracies reported across the reviewed studies, color-coded by model family. Hybrid CNN-Transformer and LSTM-based models consistently achieve the highest accuracies, while traditional ML models perform comparatively lower.*

What stands out is the consistency with which hybrid architectures outperform their simpler counterparts. CNN-Transformer pipelines and Densely Linked Bidirectional LSTMs routinely achieve between 99% and 99.97% which is roughly 4 to 6 % points better than standalone CNNs and up to 9 points better than traditional ML on comparable tasks. The explanation is that the CNNs are effective at capturing spatial patterns within a spectrogram image while attention mechanisms and LSTMs are better suited to the temporal dynamics that unfold across the EEG recording. Combining both in a single architecture gives the model access to both kinds of information simultaneously.

The most technically sound pipeline to emerge from this comparison has three core stages: ICA-based preprocessing for precise artifact removal; transformation of 1D EEG data into 2D spectrograms using Gabor, S-transform, or Wavelet Scattering methods; and classification using a hybrid CNN-Transformer or Densely Linked BiLSTM that handles spatial and temporal feature extraction together. This combination consistently outperforms any single-stage or single-modality approach.

The results from Afonso et al. [21] deserve particular attention. Showing that 16 carefully selected EEG channels can nearly match the performance of 32 is not just academically interesting but it significantly reduces the hardware requirements for practical screening, opening the door to compact wearable devices. That said, the recurring caveat across all eight studies is that the datasets are small and often from a single center, rarely validated on external cohorts. The risk of overfitting is real and the field will need substantially larger and more diverse datasets before these accuracy figures can be trusted as reliable estimates of real-world performance.

5. Critical Challenges and Limitations

The most persistent challenge identified across the reviewed literature is data. Nearly every study here was limited by small sample sizes. One used as few as 42 total PD cases across two publicly available datasets [21], and several others trained on fewer than 35 subjects. EEG signals are inherently noisy, non-linear and highly variable between individuals. Models trained on small, homogeneous samples may end up learning the features of a particular group rather than generalizing anything about PD pathology.

Two studies present related but distinct versions of this problem. Saqib et al. [20] trained their model entirely on mathematically simulated brain activity rather than actual patient recordings. The simulation is rigorous but it remains genuinely unclear whether the patterns learned from theoretical models will transfer to the trickier reality of clinical EEG. Mukherjee and Halder Roy [17] acknowledge a similar concern about GAN-generated data: synthetic augmentation helps with the sample size problem, but the generated examples may not capture the full biological variability of real PD patients.

The second major concern is overfitting and the way accuracy is being reported across the field. Classification accuracies exceeding 99% are common in this literature and they are often accompanied by 100% training accuracy which is a strong indicator of overfitting. Most studies use cross-validation on a single dataset rather than truly independent external test sets and a very few validate across multiple sites or populations. This matters considerably as a model achieving 99.97%

on one small, homogeneous dataset may perform substantially worse on a different population, a different scanner, or a different clinical protocol.

The third major issue is interpretability - the so called black-box problem of deep learning. Even a model with 99% accuracy is of limited clinical value if the clinician cannot understand why, it is making a specific decision. For a diagnosis as consequential as PD, the clinicians need to be able to verify the reasoning behind the model's output, not just accept its statistics on faith. This is one of the most significant remaining barriers to real clinical adoption of DL-based diagnostic tools and it is only partially addressed in the literature reviewed here.

Finally, there is the volume conduction problem, which is specific to scalp EEG. Electrical activity from a single brain source does not stay localized it rather spreads passively across the scalp, causing multiple electrodes to pick up the same underlying signal. This creates apparent functional connectivity between brain regions that may not actually be interacting. For PD research where understanding inter-regional brain network dynamics is central, this is a meaningful confound that current approaches handle only imperfectly.

6. EMERGING GRAPH-BASED APPROACHES TO EEG-BASED PD DETECTION

Section 5 identified three barriers to clinical translation: small and homogeneous datasets, the volume-conduction confound, and the opacity of deep-learning classifiers. A promising but methodologically distinct line of work has recently begun to address the latter two directly by abandoning the image-based paradigm and modelling EEG as a graph. The electrodes act as nodes and functional connectivity as edges - so that the relational structure of the brain, discarded by spectrogram-based CNNs, is preserved.

As noted in Section 2, these studies fall outside the primary search protocol and adopt a different paradigm from the eight classifiers reviewed above. They are therefore treated here as an emerging direction rather than part of the comparative synthesis. To our knowledge, only three studies have so far applied graph neural networks (GNNs) to scalp EEG for PD classification - a scarcity that both confirms how underexplored the direction remains and how much methodological space is still open. Table 2 summarizes them, and the discussion below examines each before drawing out the common gaps.

The earliest, by Chang et al. [22], introduced an attention-based sparse graph convolutional network (ASGCNN). EEG channels are treated as graph nodes, and the model couples a trunk branch of two graph convolution layers, which extracts global inter-electrode features. An attention branch produces a per-channel weighting, while a Ll penalty on the attention vector enforces channel sparsity so that only the most informative electrodes are retained. A spatial LSTM then captures the topological arrangement of the selected channels. Notably, the adjacency matrix is not derived from any connectivity metric but is randomly initialized and learned through backpropagation. It is evaluated on the UNM auditory-oddball dataset (24 PD recorded ON and OFF medication, 24 matched controls, 60 channels) with subject-independent leave-pair-out cross-validation. The model reached 87.67% accuracy (90.36% recall, 88.41% Fl-score). Its learned channel attention consistently emphasized the frontal and temporal lobes and revealed a contralateral frontal-lobe lateralization between patient groups, consistent with dopaminergic asymmetry. The authors identified data scarcity as the principal limitation.

Zhao et al. [23] departed from the resting-state paradigm entirely, applying their Graph-Signal-Processing Graph Convolutional Network (GSP-GCN) to task-related EEG recorded during a vocal pitch-regulation (frequency-altered-feedback) task. Functional connectivity between the 64 electrodes was computed using either the Pearson correlation coefficient or the phase-locking value. A graph-signal-processing module (incomplete network alignment plus a sparse-graph operation) reorganized the resulting network to balance local single-hop and global multi-hop information, and a Chebyshev spectral GCN performed classification. Interpretability was obtained from a guided-backpropagation saliency map cross-validated against an independent EEG microstate analysis. On

52 PD and 48 HC subjects under five-fold cross-validation, the best configuration achieved 90.2% accuracy and 89.1% AUC, with the most discriminative network localized to the left ventral premotor cortex, superior temporal gyrus, and Broca's area. The authors acknowledged that their electrode-level analysis cannot localize cortical sources and explicitly recommended graph attention networks as a future architecture.

The most methodologically rigorous of the three - and the most directly comparable to a resting-state pipeline is the work of Neves et al. [24]. It was evaluated on the same UC San Diego dataset (15 PD, 16 HC, 32 channels) that supports much of the benchmark literature. Their framework pairs a structured-global-convolution feature encoder, pretrained with self-supervised contrastive learning to compensate for the small cohort, with a multi-head graph-structure learner that learns the adjacency matrix through multi-head self-attention rather than imposing a fixed connectivity metric. A single-layer Chebyshev GNN and a head-wise gradient-weighted graph-attention explainer complete the model. Under strict subject-wise leave-one-subject-out cross-validation the model reached 69.40% accuracy (with AUC as high as 0.715 for the graph-structure-learning variant). The study's most consequential contribution is methodological: the identical model attains roughly 98% accuracy when evaluated with sample-wise splitting, a direct demonstration of the data-leakage problem that inflates much of the field's reported performance. The gradient-weighted explainer produced group-averaged connectivity maps that emphasized global and posterior connections, in contrast to the neighbor-dominated graphs yielded by Pearson correlation.

Considered together, these three studies map out both the promise and the open frontier of graph-based PD detection. Three observations stand out. First, the dominant graph-construction strategies are either volume-conduction-sensitive - the Pearson and phase-locking measures used by Zhao et al. inflate zero-lag coupling between distant electrodes or sidestep the connectivity-metric choice altogether by learning the adjacency from data. Although Neves et al. explicitly motivate their learned graph by the volume-conduction problem, no study has yet adopted a volume-conduction-robust connectivity measure, such as the phase-lag index or weighted phase-lag index, as the explicit basis for graph construction.

Second, despite the natural fit of attention-based message passing for this problem, the graph attention network has not been employed as a classifier for PD EEG. Chang et al. weight node features with learned coefficients, Zhao et al. employ a spectral GCN, and Neves et al. use multi-head attention to learn the graph rather than to classify.

Third, the wide gap between sample-wise (>95%) and subject-wise (69-90%) accuracy underscores that only rigorous subject-independent evaluation yields trustworthy estimates, and that realistic, clinically honest performance for this task lies well below the inflated figures common elsewhere. While each study incorporates some form of interpretability, none unifies individual-level and population-level explanation within a single connectivity-aware framework. These gaps - volume-conduction-robust graph construction, graph-attention classification, and multi-level explainability under rigorous evaluation - are the most immediate and tractable opportunities for the field, and the following section develops them into concrete future directions.

Table 2. Graph neural network studies applied to EEG-based PD detection.

Study	Architecture	Dataset (PD + HC, channels)	Graph Construction	Accuracy	Evaluation
Chang et al. (2023) [22]	Attention-based sparseGCN + spatial LSTM	UNM auditory oddball(24+24, 60 ch)	Learned adjacency (Backpropagation)	87.67%	Leave-pair-out CV (Subject independent)
Zhao et al. (2024) [23]	GSPmodule+ Chebyshev spectral GCN	Voice-task EEG (52+ 48,64ch)	PCCorPLV+ iNEAT sparse graph	90.2% (AUC 89.1%)	5-fold CV
Neves et al. (2024) [24]	Contrastive SGConv encoder + multi-head graph structure learner + Chebyshev GNN	UCSD resting state(15 + 16,32 ch)	Learned (multi head self-attention)	69.40% (AUCupto 0.715)	Subject-wise LOSO

7. FUTURE SCOPE

Building on the emerging graph-based studies surveyed in Section 6, several existing directions would move the field toward clinically trustworthy, individualized PD diagnostics. The most immediate concerns the graph-construction step itself. The existing GNN studies rely on either volume-conduction-sensitive connectivity metrics or purely learned adjacency matrices, leaving a clear opening:

adopting volume-conduction-robust measures such as the phase-lag index (PLI) or the weighted phase-lag index (wPLI), which discard the zero-lag coupling that volume conduction produces. This produces the explicit basis for graph construction would yield brain networks that reflect genuine neural coordination rather than passive signal spread.

A further step would model directional connectivity using techniques such as Granger causality or phase-transfer entropy, which can distinguish genuine directional interactions between brain regions from the symmetric, artificially shared activity introduced by passive signal spread, making EEG-based PD models not only more accurate but more neurobiologically meaningful. On the architectural side, attention-based graph networks and graph attention networks (GATs) remain almost entirely unexplored for PD EEG, despite being explicitly recommended by prior graph-based work [23]. Because a GAT learns how strongly each inter-channel connection contributes to its decision, its attention weights double as a built-in explanation, aligning the choice of classifier with the field's need for transparency.

The black-box problem also needs to be addressed more systematically. Afonso et al. [21] took an important first step with their XAI-based channel selection framework, but this approach has not yet been widely adopted. Building explainability modules like gradient-weighted class activation maps (Grad-CAM), SHAP-based explanations, or attention-based saliency into the deep learning architectures used for PD diagnosis would give clinicians a way to verify which brain regions, connections, and frequency bands drove a particular decision. A specific and largely unmet need is explanation at two complementary scales: no existing graph-based study reports both individual-patient attributions and population-level connectivity biomarkers. A framework that pairs per-subject explanations with group-averaged connectivity maps would offer clinicians both case-specific transparency and a population fingerprint that can be checked against established PD pathology. Such transparency is arguably a prerequisite for clinical deployment, not an optional enhancement.

Equally important, evaluation practice must mature alongside the models. Much of the field's headline accuracy is inflated by sample-wise cross-validation, in which epochs from the same subject appear in both the training and test sets; the resulting subject-level leakage can lift reported accuracy by thirty percentage points or more, as recent graph-based work has directly demonstrated. Subject-wise protocols such as leave-one-subject-out cross-validation, together with external validation on entirely independent cohorts, should become the reporting standard before any accuracy figure is treated as a reliable estimate of real-world clinical performance.

The data problem is the hardest to solve quickly, but multimodal data fusion offers a promising partial solution. Designing architectures that combine EEG with complementary modalities - voice recordings for dysphonia metrics, accelerometer data for tremor and gait quantification. Functional near-infrared spectroscopy (fNIRS), or structural MRI would both increase the effective volume of training information and make models more robust to the noise characteristics of any single modality. Future studies should also ensure that EEG is recorded from patients in both medicated (ON) and unmedicated (OFF) states, since dopaminergic medication substantially alters brain oscillatory patterns and must be accounted for in any model with genuine clinical validity.

8. CONCLUSION

This paper has reviewed eight studies, spanning 2021 to 2026, on the application of machine learning and deep learning to EEG-based Parkinson's disease detection, and additionally examined an emerging cluster of graph neural network studies as a distinct, forward-looking direction. What emerges from the primary comparison is a clear and increasingly consistent methodological pattern: converting 1D EEG time-series into 2D time-frequency spectrograms and classifying them using hybrid deep learning architectures represents the current best practice in the field. Hybrid CNN-Transformer and Densely Linked Bidirectional LSTM frameworks routinely

achieve classification accuracies between 99% and 99.97%, substantially outperforming both standalone CNNs and traditional ML models.

These are impressive numbers. But three challenges stand between the current state of the research and genuine clinical adoption. First, datasets are too small and too homogeneous for confident generalization, and evaluation practices are often insufficiently rigorous. Second, the volume conduction effect introduces systematic noise into connectivity-based analyses that current EEG-based approaches cannot fully remove. Third, the opacity of deep learning models - their inability to explain their own reasoning in terms clinicians can evaluate - makes them difficult to trust in high-stakes diagnostic settings.

The path forward builds directly on the nascent graph-based work surveyed here by addressing its specific gaps: volume-conduction-robust graph construction that models real neural coordination, attention-based architectures that are interpretable by design, and explainability delivered at both the individual and population levels. Combined with multimodal data fusion for greater robustness and with rigorous subject-wise and cross-dataset evaluation, these directions would place brain-network modeling on a principled and clinically credible footing. As datasets grow larger and more diverse, and as these technical challenges are progressively addressed, EEG-based deep learning holds genuine promise for enabling earlier, more objective, and more individualized PD diagnostics at scale - improving quality of life and therapeutic outcomes for the many millions of people living with this disease worldwide.

REFERENCES

[1] K. R. Chaudhuri and A. H. Schapira, "Non-motor symptoms of Parkinson's disease: dopaminergic pathophysiology and treatment," Lancet Neurol., vol. 8, no. 5, pp. 464–474, May 2009, doi: 10.1016/S1474-4422(09)70068-7.

[2] C. D. Marsden and J. A. Obeso, "The functions of the basal ganglia and the paradox of stereotaxic surgery in Parkinson's disease," Brain, vol. 117, no. 4, pp. 877–897, 1994, doi: 10.1093/brain/117.4.877.

[3] L. T. Grinberg, U. Rueb, A. T. di L. Alho, and H. Heinsen, "Brainstem pathology and non-motor symptoms in PD," J. Neurol. Sci., vol. 289, no. 1–2, pp. 81–88, Feb. 2010, doi: 10.1016/j.jns.2009.08.021.

[4] N. Schiess et al., "Six Action Steps to Address Global Disparities in Parkinson Disease: A World Health Organization Priority," JAMA Neurol., vol. 79, no. 9, pp. 929–936, Sep. 2022, doi: 10.1001/jamaneurol.2022.1783.

[5] J. E. Ahlskog and M. D. Muenter, "Frequency of levodopa-related dyskinesias and motor fluctuations as estimated from the cumulative literature," Movement Disorders, vol. 16, no. 3, pp. 448–458, May 2001, doi: 10.1002/mds.1090.

[6] M. F. Contarino et al., "Cognitive outcome 5 years after bilateral chronic stimulation of subthalamic nucleus in patients with Parkinson's disease," J. Neurol. Neurosurg. Psychiatry, vol. 78, no. 3, pp. 248–252, 2007, doi: 10.1136/jnnp.2005.086660.

[7] A. Sharma, A. Gupta, and S. Sharma, "Advanced cognitive classification in Parkinson's disease through Morlet wavelet analysis of EEG-derived key brain regions," Int. J. Adv. Eng. Sci. Appl. Math., Jan. 2026, doi: 10.1007/s12572-025-00420-y.

[8] F. A. Jibon et al., "Parkinson's disease detection from EEG signal employing autoencoder and RBFNN-based hybrid deep learning framework utilizing power spectral density," Digit. Health, vol. 10, p. 20552076241297355, Nov. 2024, doi: 10.1177/20552076241297355.

[9] F. A. Jibon et al., "Epileptic seizure detection from electroencephalogram (EEG) signals using linear graph convolutional network and DenseNet based hybrid framework," J. Radiat. Res. Appl. Sci., vol. 16, no. 3, p. 100607, Sep. 2023, doi: 10.1016/j.jrras.2023.100607.

[10] T. Wen and Z. Zhang, "Deep Convolution Neural Network and Autoencoders-Based Unsupervised Feature Learning of EEG Signals," IEEE Access, vol. 6, pp. 25399–25410, 2018, doi: 10.1109/ACCESS.2018.2833746.

[11] A. Gupta, Ed., Next Generation Computing and Information Systems: Proceedings of the 2nd International Conference on Next Generation Computing and Information Systems (ICNGCIS 2023). London: CRC Press, 2024, doi: 10.1201/9781003466383.

[12] B. Debus, H. Parastar, P. Harrington, and D. Kirsanov, "Deep learning in analytical chemistry," TrAC Trends in Analytical Chemistry, vol. 145, p. 116459, Dec. 2021, doi: 10.1016/j.trac.2021.116459.

[13] G. Gimenez-Aparisi, E. Guijarro-Estelles, A. Chornet-Lurbe, S. Ballesta-Martinez, M. Pardo-Hernandez, and Y. Ye-Lin, "Early detection of Parkinson's disease: Systematic analysis of the influence of the eyes on quantitative biomarkers in resting state electroencephalography," Heliyon, vol. 9, no. 10, p. e20625, Oct. 2023, doi: 10.1016/j.heliyon.2023.e20625.

[14] A. Delis, G. Tsavdaridis, and P. Tsanakas, "A Novel Battery-Supplied AFE EEG Circuit Capable of Muscle Movement Artifact Suppression," Applied Sciences, vol. 14, no. 16, p. 6886, Aug. 2024, doi: 10.3390/app14166886.

[15] S. K. Khare, V. Bajaj, and U. R. Acharya, "PDCNNet: An Automatic Framework for the Detection of Parkinson's Disease Using EEG Signals," IEEE Sens. J., vol. 21, no. 15, pp. 17017–17024, Aug. 2021, doi: 10.1109/JSEN.2021.3080135.

[16] M. Obayya, M. K. Saeed, M. Maashi, S. S. Alotaibi, A. S. Salama, and M. A. Hamza, "A novel automated Parkinson's disease identification approach using deep learning and EEG," PeerJ Comput. Sci., vol. 9, p. e1663, 2023, doi: 10.7717/peerj-cs.1663.

[17] P. Mukherjee and A. Halder Roy, "A robust deep learning-driven framework for detecting Parkinson's disease using EEG," Comput. Methods Biomech. Biomed. Engin., published online Sep. 2025, doi: 10.1080/10255842.2025.2556310.

[18] M. Mollazadeh Azari, Y. Sarbaz, and S. Meshgini, "Investigating time distortion in Parkinson's disease considering impaired frontoparietal network and changes in the brain dynamic," Biomed. Phys. Eng. Express, vol. 12, no. 2, p. 025001, Feb. 2026, doi: 10.1088/2057-1976/ae4107.

[19] M. Kolla, R. K. Madapuri, P. Kandukuri, S. Salvadi, S. Tadepalli, and R. Gajula, "Optimized cortical EEG modeling for Parkinson disease diagnosis with snow Shepherd Stride tuning mechanism," Cogn. Neurodyn., vol. 20, no. 1, p. 47, Feb. 2026, doi: 10.1007/s11571-025-10406-y.

[20] S. U. Saqib, S. H. Fang, M. A. Z. Raja, K. S. Nisar, and M. Shoaib, "Design of intelligent neuro-supervised deep learning networks to analyze brain electrical activity rhythms of Parkinson's disease model," Cogn. Neurodyn., vol. 20, no. 1, p. 32, Feb. 2026, doi: 10.1007/s11571-025-10404-0.

[21] M. M. Afonso, D. R. Edla, and R. R. Reddy, "XAI-driven average region intensity-based EEG channel selection (ARI-ECS) for optimizing Parkinson's disease detection," Biomed. Signal Process. Control, vol. 113, p. 108899, Mar. 2026, doi: 10.1016/j.bspc.2025.108899.

[22] H. Chang, B. Liu, Y. Zong, C. Lu, and X. Wang, "EEG-Based Parkinson's Disease Recognition via Attention-Based Sparse Graph Convolutional Neural Network," IEEE J. Biomed. Health Inform., vol. 27, no. 11, pp. 5216–5224, Nov. 2023, doi: 10.1109/JBHI.2023.3292452.

[23] S. Zhao et al., "An interpretable model based on graph learning for diagnosis of Parkinson's disease with voice-related EEG," npj Digit. Med., vol. 7, no. 1, p. 3, Jan. 2024, doi: 10.1038/s41746-023-00983-9.

[24] C. Neves, Y. Zeng, and Y. Xiao, "Parkinson's Disease Detection from Resting State EEG using Multi-Head Graph Structure Learning with Gradient Weighted Graph Attention Explanations," arXiv preprint arXiv:2408.00906, Aug. 2024.