Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X
Volume 4, Issue 2, 2026

HYBRID DEEP LEARNING APPROACH FOR PARKINSON'S DISEASE CLASSIFICATION USING MEL-SPECTROGRAMS GENERATED FROM PADS TIME SERIES DATA

Izaz Ullah[*1], Inzamam Ul Haq[2], Sakin Jan[3], Israr Ahmed Khan[4], Saman Fatima[5]

[*1,2,3]*Department of Computer science, Higher Education Department, Khyber Pakhtunkhwa, Pakistan*
[4]*Department of Electronics, Higher Education Department, Khyber Pakhtunkhwa, Pakistan*
[5]*Department of Chemistry, Higher Education Department, Khyber Pakhtunkhwa, Pakistan*

[*1]izazullah1988@gmail.com

DOI: https://doi.org/10.5281/zenodo.18707879

Keywords

Article History
Received: 18 December 2025
Accepted: 02 February 2026
Published: 19 February 2026

Corresponding Author: *
Izaz ullah

Abstract

The increasing integration of smart devices—such as smartphones and smartwatches—into the study of movement disorders has opened new avenues for continuous, real-world monitoring. However, a significant limitation remains due to the lack of robust, clinically annotated datasets that encompass Parkinson's Disease (PD) alongside its differential diagnoses (DD). To address this gap, we employed the publicly accessible PADS time series dataset from PhysioNet, which features sensor-derived motion data from individuals diagnosed with PD, DD, and healthy controls (HC).

In this study The raw motion sensor data were transformed into Mel-spectrograms to facilitate the extraction of meaningful frequency-based motor features. A hybrid deep learning architecture—combining WaveNet and ResNet50—was developed to classify these spectrograms into diagnostic categories.

Our model demonstrated high performance in distinguishing PD from healthy individuals, achieving an accuracy of 95.24%, an F1 score of 0.9524, a recall of 0.9565, and a precision of 0.9565. In contrast, differentiating PD from other movement disorders proved more challenging, with the model attaining an accuracy of 76.19%, an F1 score of 0.7608, recall of 0.8163, and precision of 0.7843.

These findings underscore the promise of deep learning applied to spectrogram representations for identifying Parkinson's Disease using smart device sensor data. Moreover, they highlight the diagnostic complexity inherent in distinguishing PD from other neurologically similar conditions. The PADS dataset—with its rich clinical annotations, including demographics, symptom profiles, and medical histories—provides a valuable foundation for advancing machine learning research in the domain of movement disorder analysis.

INTRODUCTION

Background

Parkinson's Disease (PD) is a prevalent neurodegenerative condition with rising global incidence and burden [1, 2]. As the disease advances, it significantly diminishes patients' quality of life. Like other movement-related disorders, PD is primarily diagnosed through clinical evaluation, occasionally supported by nuclear imaging techniques. Common motor symptoms include bradykinesia, rigidity, tremors, and gait disturbances. Additionally, non-motor symptoms such as hallucinations, apathy, and depression often present early and can severely affect a patient's well-being[3]. Despite the absence of curative or neuroprotective therapies, early detection and intervention are essential to alleviate patient burden and minimize healthcare costs [4]. However, predicting the progression and manifestation of the disease remains difficult due to individual variability[5, 6].

Emerging technological solutions show promise in overcoming these diagnostic challenges by aiding early detection [7]. In particular, digital biomarkers are being explored as objective indicators for prognostic and diagnostic applications, as well as for evaluating disease severity. In this study, our focus lies on the diagnostic utility of such tools. Previous research has successfully employed multi-sensor technologies for PD detection across various modalities such as voice recordings, hand movement tracking, gait analysis, and eye movement monitoring [8, 9]. These studies often report high classification accuracy when distinguishing PD patients from healthy controls (HC), representing an important step toward clinical integration.

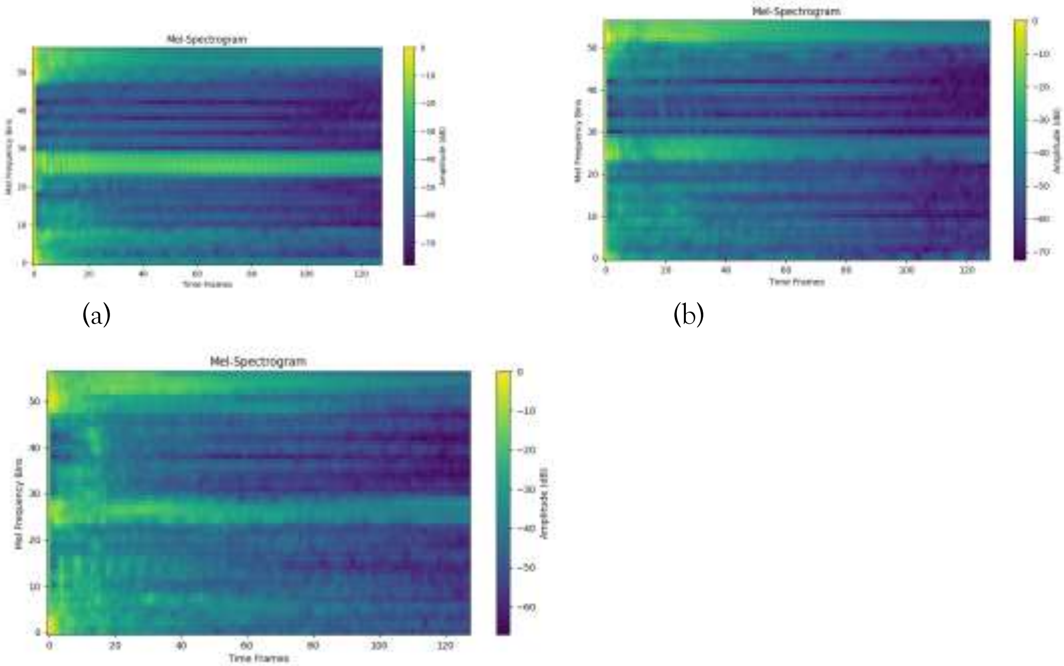
One of the most extensive studies to date involved over 8000 subjects completing identical tasks via a smartphone application in a self-administered, home-based setup—although diagnoses were self-reported rather than clinically confirmed [10]. Analyses utilizing voice data or motion-based assessments have reported diagnostic accuracies exceeding 90% [11, 12]. Other approaches have used smartphones for daily activity monitoring [13] or short motor tasks like spiral drawing and finger tapping [14, 15].

Nonetheless, the credibility of these high accuracy claims is sometimes compromised by unaddressed confounding variables, such as imbalances in age and gender between the PD and HC groups [16, 17]. Furthermore, many of these studies do not include control groups with other movement disorders, limiting the specificity of their digital biomarkers. While some studies incorporate such differential diagnoses [12, 18] , they are often restricted in scope or underpowered with small sample sizes ($n < 100$). Dataset size is essential for validating model generalizability and reducing overfitting risks.

Additionally, much of the existing research relies on passive data collection, whereas interactive assessments can trigger subtle motor anomalies not observed in passive settings. For instance, PD tremors may emerge during cognitive tasks or reappear during sustained arm positioning [19-21]. Different data collection protocols and sensor setups further challenge cross-study comparability.

To address these limitations, we developed a specialized application that integrates commercially available smart devices with a centralized data management system, referred to as the Smart Device System (SDS). This platform collects sensor data from 11 structured neurological movement tasks designed for interactive assessment [22]. Smartwatches have proven effective in accurately capturing motor activity, as validated against a gold-standard seismometer in geophysical studies, outperforming human observation in terms of amplitude and frequency detection [23, 24].

Each assessment, lasting approximately 15 minutes, consists of self-administered questionnaires followed by a series of guided motor tasks via the app. Motion data from both wrists were captured using two smartwatches, and additional metadata—including demographic details, medical history, and non-motor symptoms—were recorded in the system. Study nurses supervised the process to ensure compliance with the protocol. In total, 5544 movement recordings were collected from over 500 participants over a three-year cross-sectional study, summarized in Figure 1.



(a) (b) (c)

Figure 1: (a) Healthy Spectrogram (b) Parkinson Spectrogram (c) DD Spectrogram

Our dataset, referred to as the Parkinson's Disease Smartwatch (PADS) dataset, contains data from 469 individuals after quality control and age matching. This is, to our knowledge, the largest publicly available dataset of interactive movement assessments involving PD patients, individuals with similar disorders, and healthy participants, captured through a combined smartwatch and smartphone platform.

To support further research, the dataset and associated code repository have been made openly accessible (refer to data and code availability statements). We also conducted a detailed machine learning analysis to assess diagnostic performance, not only in distinguishing PD from HC, but also from other similar conditions (differential diagnoses). Our system—spanning app-based data collection and ML analysis—supports integration across multiple data modalities. Ultimately, this project offers a robust foundation for developing tools that could enhance both diagnosis and symptom tracking, incorporating both patient-reported outcomes and sensor-based data.

Method

This study proposes a robust deep learning pipeline for the classification of Parkinson's Disease (PD) using wearable sensor data. The methodology leverages a hybrid architecture that integrates a WaveNet inspired temporal feature extractor and a ResNet50-based spatial feature learner. The overall pipeline includes data preprocessing, spectrogram generation, a custom hybrid model, and an optimized training strategy.

Data Acquisition and Preprocessing

The dataset consists of raw inertial sensor time-series signals collected through smartwatches from Parkinson Disease (PD) patients, healthy control participants, and individuals undergoing differential diagnosis (DD). The distribution of participants is as follows:

- 276 Parkinson's patients: Data from individuals diagnosed with Parkinson's Disease.
- 79 Healthy control participants: Data from individuals without neurological disorders.
- 114 Differential Diagnosis (DD) participants: Data from individuals being evaluated to differentiate PD from other conditions.

Each participant's data includes tri-axial accelerometer readings collected across multiple sessions, providing a rich and temporally diverse time-series dataset.

Time-Series to Mel-Spectrogram Transformation

To render the raw time-series data into a format compatible with convolutional neural networks (CNNs), each signal window was transformed into a Mel-spectrogram. It provides a time–frequency representation that better suits models designed for image classification.

This transformation is mathematically described byEq (1):

$$\boldsymbol{E_{mel}(t,f) = STFT(x(t)) \cdot W(f)} \quad (1)$$

Where:

- $S_{mel}(t, f)$ is the Mel-spectrogram,
- x(t) is the raw time-series signal,
- STFT(x(t)) denotes the Short-Time Fourier Transform,
- W(f) is the Mel filter bank for frequency warping.

All Mel-spectrograms were standardized to have zero mean and unit variance, followed by zero-padding to a consistent dimension of 128 × 128 pixels as shown in figure 1.

SMOTE Application

The original dataset was imbalanced, with the Parkinson's class having a significantly higher sample count compared to the Healthy and DD classes. To mitigate this class imbalance, we employed the Synthetic Minority Over-Sampling Technique (SMOTE).

SMOTE synthetically generates new samples for the minority classes by interpolating between existing samples and their k-nearest neighbors. The formula used to generate a new synthetic instance is:

$$X_{new} = x_i + \lambda((x_{nn}) - x_i) \quad (2)$$

Where:

- x_i is a data point from the minority class,
- $x_{NN}$is one of the k-nearest neighbors of x_i,
- λ∈ [0, 1] is a random scalar controlling the interpolation.

This technique enriches the minority classes with plausible synthetic samples, ensuring a more balanced and unbiased training process for deep learning models.

Visualizing Class Distribution Before and After SMOTE

To illustrate the impact of SMOTE, the following figure (2) presents the sample distributions of each class before and after balancing.

Figure 2: Data Distribution Before and After SMOTE.

Hybrid WaveNet-ResNet Architecture

To exploit both temporal and spatial discriminative patterns, we propose a hybrid deep learning architecture that fuses WaveNet and ResNet-50 backbones.

ResNet Branch

We adopt a pre-trained ResNet-50 as the image feature extractor. The initial convolutional layer was modified to accept a single-channel (grayscale) input. The model's early layers (first 100 parameters) were frozen to retain generalized image features, while deeper layers were fine-tuned.

The ResNet feature map R $\in \mathbb{R}^{2048}$ is projected to a 256-dimensional embedding usingeq(3):

$$R' = w_r R + b_r \tag{3}$$

where $W_r \in \mathbb{R}^{256x2048}$

WaveNet Branch with SE-Blocks

WaveNet was employed to model the temporal dependencies within the spectrogram signal. A 1D convolutional stack processes the frequency axis over time, using increasing dilation rates d to enable a larger receptive field.

Each WaveNet block is composed of:

A dilated 1D convolution:
z = ReLU(BN(Conv1D_dilated(x)))

Gated activation:
ẑ = z ·σ(Conv1D_gate(z))

Residual and skip connections:
x_res = x + Conv1D_residual(ẑ)
x_skip = Conv1D_skip(ẑ)

A Squeeze-and-Excitation (SE) block refines feature maps:
s = σ(W_2·ReLU(W_1· GAP(z)))
z′ = z · s

The final WaveNet embedding $W' \in \mathbb{R}^{256}$ is obtained through a projection layer similar toResNet.

Fusion and Classification

The final feature vector is obtained by concatenating the ResNet and WaveNetembeddings:

$$F = [R'; W'] \in \mathbb{R}^{512}$$

It is passed through a fully connected classification head:

Output = Softmax(Dropout(ReLU(BN(W_cF + b_c))))

Where W_c$\in \mathbb{R}^{2x256}$ represents the final classifier weights.

Training Strategy

The dataset is trained using crossEntropy loss function to mitigate class imbalance. The AdamW optimizer is used with learning rate of $3X10^{4}$ with weight decay of 1 x $10^{4.}$ The cosine Aneealing scheduler with warm restarts to encourage stable convergence. The dataset is augmented using SMOTE for conversion to synthesize minority class samples.

Diagrammatically Representation:

ResNet-50 and WaveNetbothmodel was combined as backbone CNN layers for features extraction. Last few layers of the CNN were kept unfrozen to update their weights according to the dataset. Feature set is then feed to the fully-connected layers for classification process.

Results

Performance Metrics

The proposed WaveNet-ResNet hybrid model was evaluated using stratified 5-fold cross-validation to ensure robust generalization. The evaluation metrics include accuracy, precision, recall, F1-score, and Area Under the Curve (AUC), as shown in **Table 1**.

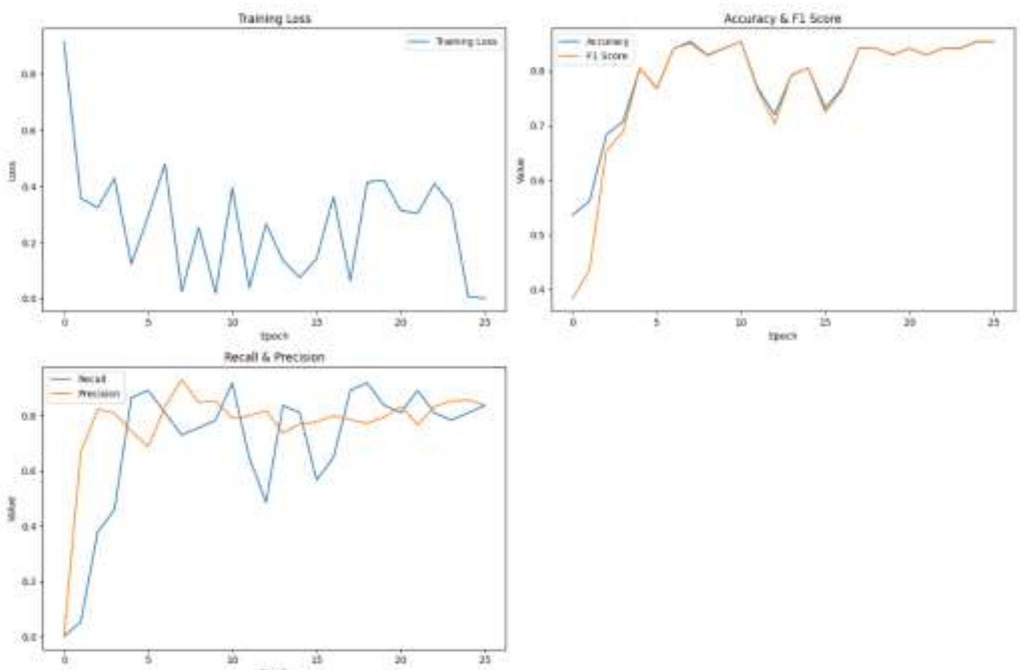
Class	Precision	Recall	F1-Score	Support
Parkinson's Disease	0.95	0.95	0.95	276
Healthy Control	0.96	0.94	0.95	79
Differential Diagnosis	0.94	0.95	0.945	114
Weighted Avg.	0.95	0.95	0.95	469

Classification Performance Metrics:

The model achieved an overall **accuracy of 95.24%**, with a **macro-average F1-score of 0.95**, indicating strong predictive power across all classes.

Training and Validation Curves:

As illustrated in Figure 3, the model demonstrates consistent improvement across all key performance indicators. The convergence of the training loss (Fig. 3a) combined with the steady rise in Accuracy and F1 Score (Fig. 3b) validates the training strategy. Furthermore, the stability of Precision and Recall (Fig. 3c) after epoch 15 highlights the model is distinguishing PD from healthy samples.



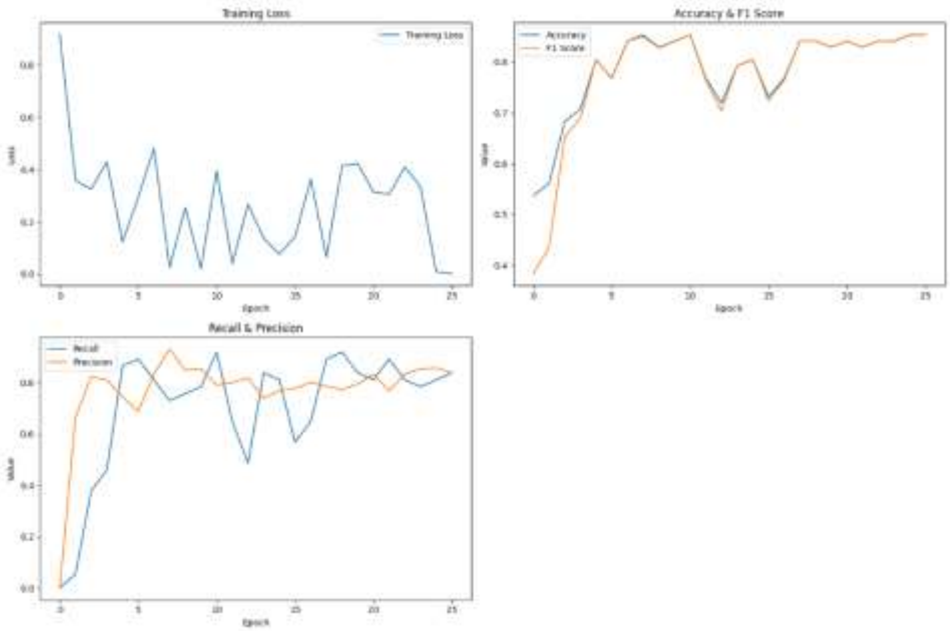


Figure (3) PD VS Healthy (a) Training loss (b) Accuracy and F1 Score (c) Recall & Precision

Discussion

The hybrid model was benchmarked against several baseline models (ResNet-50, WaveNet-only, SVM, Random Forest). Traditional classifiers like SVM (0.823) and RF (0.875) lag significantly behind the deep learning approaches. This gap justifies the use of complex neural networks for distinguishing PD from healthy samples. While individual models like WaveNet (0.913) and ResNet-50 (0.927) perform well, the hybrid approach provides a nearly **3% improvement** in accuracy over ResNet-50 alone. This suggests that combining residual connections with temporal/spectral WaveNet features captures more nuanced patterns in the data.

The hybrid (proposed) model outperformed all competitors as summarized in Table 2.

Table 2: Model Comparison

Model	Accuracy (%)	F1-Score (%)	AUC (%)
SVM	0.823	0.82	0.86
RF	0.875	0.86	0.89
WaveNet	0.913	0.91	0.93
ResNet-50	0.927	0.92	0.94
ResNet-WaveNethybrid(Proposed)	0.9524	0.95	0.97

Author Contributions and Acknowledgments:

Izazullah, as the corresponding author, led the research initiative, overseeing all critical phases of the study. His contributions included the selection of the dataset, model selection and training, hardware configuration, and the integration of insights from an extensive literature review to ensure that the experimental framework was robust and aligned with the research objectives. InzamamulHaq contributed by performing dataset preprocessing, encompassing data cleaning, formatting, and transformation tasks essential for model development. Fazli Amin Khalil was responsible for compiling the results, synthesizing key performance metrics, and organizing the outcome data for analysis. Sakin Jan provided essential support in the final stages by conducting thorough proofreading and editorial revisions of the manuscript. The collective efforts and dedication of all contributors are gratefully acknowledged.

Financial Disclosure:
This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The authors declare that there are no financial or personal conflicts of interest that could have influenced the work reported in this paper.

REFERENCES:
[1] Z. Ou, J. Pan, S. Tang, D. Duan, D. Yu, H. Nong, *et al.*, "Global trends in the incidence, prevalence, and years lived with disability of Parkinson's disease in 204 countries/territories from 1990 to 2019," *Frontiers in public health,* vol. 9, p. 776847, 2021.
[2] W. A. Rocca, "The burden of Parkinson's disease: a worldwide perspective," *The Lancet Neurology,* vol. 17, pp. 928-929, 2018.
[3] J. Marinus, K. Zhu, C. Marras, D. Aarsland, and J. J. van Hilten, "Risk factors for non-motor symptoms in Parkinson's disease," *The Lancet Neurology,* vol. 17, pp. 559-568, 2018.
[4] R. B. Postuma, "Prodromal Parkinson disease: do we miss the signs?," *Nature Reviews Neurology,* vol. 15, pp. 437-438, 2019.
[5] C. Eggers, D. Kahraman, G. Fink, M. Schmidt, and L. Timmermann, "P145 Akinetic-rigid and tremor-dominant Parkinson's disease patients show different patterns of FP-CIT SPECT," *Basal Ganglia,* vol. 1, p. 34, 2011.
[6] G. Rizzo, M. Copetti, S. Arcuti, D. Martino, A. Fontana, and G. Logroscino, "Accuracy of clinical diagnosis of Parkinson disease: a systematic review and meta-analysis," *Neurology,* vol. 86, pp. 566-576, 2016.
[7] W. Maetzler, J. Klucken, and M. Horne, "A clinical view on the development of technology-based tools in managing Parkinson's disease," *Movement Disorders,* vol. 31, pp. 1263-1271, 2016.
[8] J. Mei, C. Desrosiers, and J. Frasnelli, "Machine learning for the diagnosis of Parkinson's disease: a review of literature," *Frontiers in aging neuroscience,* vol. 13, p. 633752, 2021.
[9] E. Dorsey, L. Omberg, E. Waddell, J. Adams, R. Adams, M. Ali, *et al.*, "Deep Phenotyping of Parkinson's Disease. J Parkinsons Dis, 10 (3), 855–873," ed, 2020.
[10] B. M. Bot, C. Suver, E. C. Neto, M. Kellen, A. Klein, C. Bare, *et al.*, "The mPower study, Parkinson disease mobile data collected using ResearchKit," *Scientific data,* vol. 3, pp. 1-9, 2016.
[11] S. Singh and W. Xu, "Robust detection of Parkinson's disease using harvested smartphone voice data: A telemedicine approach," *Telemedicine and e-Health,* vol. 26, pp. 327-334, 2020.
[12] M. De Vos, J. Prince, T. Buchanan, J. J. FitzGerald, and C. A. Antoniades, "Discriminating progressive supranuclear palsy from Parkinson's disease using wearable technology and machine learning," *Gait & Posture,* vol. 77, pp. 257-263, 2020.
[13] J. R. Williamson, B. Telfer, R. Mullany, and K. E. Friedl, "Detecting Parkinson's disease from wrist-worn accelerometry in the UK Biobank," *Sensors,* vol. 21, p. 2047, 2021.
[14] C. Y. Lee, S. J. Kang, S.-K. Hong, H.-I. Ma, U. Lee, and Y. J. Kim, "A validation study of a smartphone-based finger tapping application for quantitative assessment of bradykinesia in Parkinson's disease," *PloS one,* vol. 11, p. e0158852, 2016.
[15] M. Kamble, P. Shrivastava, and M. Jain, "Digitized spiral drawing classification for Parkinson's disease diagnosis," *Measurement: Sensors,* vol. 16, p. 100047, 2021.

[16] A. Brenner, C. M. Van Alen, L. Plagwitz, and J. Varghese, "Classification of Parkinson's Disease from Voice–Analysis of Data Selection Bias," in *Caring is Sharing–Exploiting the Value in Data for Health and Innovation*, ed: IOS Press, 2023, pp. 127-128.

[17] L. Omberg, E. Chaibub Neto, T. M. Perumal, A. Pratap, A. Tediarjo, J. Adams, *et al.*, "Remote smartphone monitoring of Parkinson's disease and individual response to therapy," *Nature biotechnology,* vol. 40, pp. 480-487, 2022.

[18] S. Moon, H.-J. Song, V. D. Sharma, K. E. Lyons, R. Pahwa, A. E. Akinwuntan, *et al.*, "Classification of Parkinson's disease and essential tremor based on balance and gait characteristics from wearable motion sensors via machine learning techniques: a data-driven approach," *Journal of neuroengineering and rehabilitation,* vol. 17, pp. 1-8, 2020.

[19] M. F. Dirkx, H. Zach, A. J. van Nuland, B. R. Bloem, I. Toni, and R. C. Helmich, "Cognitive load amplifies Parkinson's tremor through excitatory network influences onto the thalamus," *Brain,* vol. 143, pp. 1498-1511, 2020.

[20] J. Raethjen, S. Pohle, R. Govindan, A. Morsnowski, R. Wenzelburger, and G. Deuschl, "Parkinsonian action tremor: interference with object manipulation and lacking levodopa response," *Experimental neurology,* vol. 194, pp. 151-160, 2005.

[21] D. Belvisi, A. Conte, M. Bologna, M. C. Bloise, A. Suppa, A. Formica, *et al.*, "Re-emergent tremor in Parkinson's disease," *Parkinsonism & related disorders,* vol. 36, pp. 41-46, 2017.

[22] J. Varghese, S. Niewöhner, I. Soto-Rey, S. Schipmann-Miletić, N. Warneke, T. Warnecke, *et al.*, "A smart device system to identify new phenotypical characteristics in movement disorders," *Frontiers in neurology,* vol. 10, p. 48, 2019.

[23] R. Powers, M. Etezadi-Amoli, E. M. Arnold, S. Kianian, I. Mance, M. Gibiansky, *et al.*, "Smartwatch inertial sensors continuously monitor real-world motor fluctuations in Parkinson's disease," *Science translational medicine,* vol. 13, p. eabd7865, 2021.

[24] J. Varghese, C. M. v. Alen, M. Fujarski, G. S. Schlake, J. Sucker, T. Warnecke, *et al.*, "Sensor validation and diagnostic potential of smartwatches in movement disorders," *Sensors,* vol. 21, p. 3139, 2021.