Early Detection of Parkinson's Disease on-the-fly utilising Wearable Technology

Devam Agarwal, Rhea Singh and Trusha Maheshwari

Abstract

Parkinson's Disease (PD), a progressive neurodegenerative disorder, is commonly diagnosed through clinical observation, which often leads to delayed detection, especially in early stages. Prior research has explored deep learning on wearable sensor data to improve detection, but these models are too resource-intensive for real-time or remote applications. We propose a lightweight and interpretable multimodal machine learning pipeline for early detection of Parkinson's Disease (PD), designed for deployment on edge devices such as smartwatches. Our approach integrates engineered features from time-series motor data, capturing statistical, wavelet, and cross-channel patterns, with self-reported questionnaire responses related to non-motor symptoms. After feature selection using a random forest classifier, the combined data were processed through a dual-branch neural network architecture, with separate multilayer perceptrons (MLPs) for each modality and a final classification layer.

Evaluated using 5-fold stratified cross-validation, the model demonstrated excellent performance in binary classification tasks: 100% accuracy and F1-score for PD vs healthy controls, and 98.72% accuracy with an F1-score of 0.9910 for PD vs differential diagnoses. In the more challenging PD vs All Others task, which includes both healthy and differentially diagnosed individuals, the model achieved 78.72% accuracy and an F1-score of 0.8182, indicating strong generalization. The final model, approximately 2 MB in size, is suitable for real-time, remote PD screening, offering a promising tool for scalable early diagnosis in clinical and non-clinical settings.

Introduction

This research presents a multimodal machine learning system for the early detection of Parkinson's Disease (PD), leveraging both wearable sensor data and self-reported symptom questionnaires. As a progressive neurodegenerative disorder, PD benefits greatly from early diagnosis, yet current clinical methods often rely on infrequent visits and subjective symptom assessments, limiting timely intervention. Recent studies have explored machine learning to address these challenges, with deep learning models demonstrating promising results. However, many existing approaches are computationally intensive and rely exclusively on motor symptoms, making them unsuitable for deployment on resource-constrained edge devices like smartwatches. Moreover, they often overlook non-motor indicators, which can precede observable movement impairments.

To address these limitations, we propose a lightweight, interpretable dual-branch neural network that integrates time-series motor data and questionnaire responses. Motor signals are transformed into statistical, wavelet, and cross-channel features, while questionnaire inputs capture non-motor and demographic information. These are processed in parallel through two multilayer perceptrons (MLPs) before being fused for final classification. This hybrid approach enables robust and scalable PD screening, offering strong performance across both binary and multi-class settings while remaining compact enough for real-time, on-device deployment. The paper outlines the datasets, model architecture, and training process, followed by an evaluation of results and implications for clinical translation.

Background and Significance

Parkinson's disease, a progressive movement disorder of the nervous system damages and weakens nerve cells, leading to issues with movement, tremors, and impaired balance. With symptoms such as tremors, muscle rigidity, bradykinesia, and postural instability, it affects the motor functions of the body. Moreover, symptoms like cognitive decline, mood disorders, and disturbed sleep also develop with the onset of Parkinson's. A period of 5 years of symptoms allows for improved diagnostic accuracy, but it is far more difficult to predict in the early stages of the disease.

However, the current diagnostic methods rely heavily on subjective clinical evaluations which are performed by neurologists. Some of the tools that are used widely are the Unified Parkinson's Disease Rating Scale (UPDRS) and the Hoehn and Yahr scale, but these assessments are based on observable symptoms, making it difficult to detect subtle, early-stage indications of the disease. As a result, many patients receive a diagnosis only after there has been significant neuronal damage, which in turn limits the effectiveness of treatments.

Additionally, another challenge is the limited access to specialized neurological care in many regions, making early diagnosis even more difficult. In many countries, the number of neurologists available per capita is inadequate to meet the growing demand for Parkinson's assessments. This creates a major barrier to scalability and accessibility, particularly in rural or underserved areas where patients may not have access to frequent clinical evaluations. These limitations prove that there is a strong need for automated, non-invasive, and cost-effective methods for early PD detection.

Recent advancements in technology and domains such as machine learning have provided new avenues for Parkinson's diagnosis through the assessment of day-to-day movement patterns and cognitive performance. Gait impairments such as shortened step length, elevated step variability, and asymmetric movement are part of the early indicators of PD. Smartwatches and other wearables with built-in accelerometers and gyroscopes are capable of monitoring ongoing motor activity and are able to gain useful insights into minor gait abnormalities that are not observable under clinical tests. It has been demonstrated in research that through time-series feature extraction of sensor signals is capable of detecting these movement disorders even prior to the development of severe symptoms.

Beyond motor symptoms, cognitive impairment is a significant aspect of Parkinson's disease. Many individuals with PD experience difficulties with speech, language processing, and cognitive response times. Structured clinical surveys, such as cognitive function tests, have been used by neurologists to assess these changes, but recent studies have explored the potential of natural language processing (NLP) to analyze speech and written responses for signs of cognitive decline. This approach could enable objective and quantifiable assessments of cognitive health in Parkinson's patients, complementing traditional diagnostic tools.

Outside of motor symptoms, cognitive change is an important feature of Parkinson's disease. Most patients with PD have trouble with speech, language processing, and response times for cognitive tasks. Formatted clinical questionnaires, including tests of cognitive function, have been employed by neurologists to measure these changes, but new research has investigated the possibility of natural language processing (NLP) to examine speech and written answers for evidence of cognitive change. This strategy may facilitate objective and measurable measures of cognitive function in Parkinson's patients, augmenting the conventional diagnostic modalities.

Literature Review

Parkinson's Disease (PD) is a progressive neurodegenerative disorder characterized by motor and non-motor symptoms that profoundly impact quality of life. Historically, the field of PD research has been rooted in clinical neurology and neuropathology, focusing on symptom management and disease progression through clinical observation and pharmacological intervention. However, the last decade has witnessed a paradigm shift, with the integration of biomedical engineering, wearable technologies, and machine learning (ML) approaches. This interdisciplinary convergence aims to address the limitations of traditional methods by enabling continuous, objective, and data-driven monitoring of PD symptoms.

Large volumes of data can now be gathered from PD patients outside of the clinic in real-world situations thanks to recent technological developments. This change is important because it enables clinicians and researchers to record the daily fluctuations in symptoms that are frequently overlooked in routine clinical evaluations. In addition to improving the objectivity of symptom monitoring, the

application of wearable sensors and machine learning creates new opportunities for early diagnosis, individualised care, and disease management (Eskofier et al., 2016, p. 2). The integration of these technologies represents a new frontier in PD research, promising improved patient outcomes and more efficient healthcare delivery.

Locating Research Within the Discipline

Historically, PD research has focused on clinical and pathological mechanisms. James Parkinson's seminal "Essay on the Shaking Palsy" (1817) laid the foundation for subsequent discoveries, such as the identification of dopaminergic neuron degeneration and Lewy bodies as pathological hallmarks (Dickson, 2018, p. 2). Clinical assessment has primarily relied on rating scales like the Unified Parkinson's Disease Rating Scale (UPDRS) and the Hoehn and Yahr staging system (Hoehn & Yahr, 1967, p. 427). While these tools are used on a large scale, they are subjective and are limited in capturing daily symptom fluctuations and non-motor symptoms.

The limitations of traditional methods to detect Parkinson's disease have prompted the integration of engineering and data science into its research. Wearable sensors such as accelerometers and gyroscopes enable real-time and objective monitoring of motor symptoms (Del Din et al., 2016, p. 2). With the use of these data streams, machine learning algorithms are currently being created to identify, categorise, and forecast PD symptoms more precisely (Pereira et al., 2019, p. 1). This cross-disciplinary approach is changing the way PD is managed from one-time, clinic-based care to ongoing, individualised monitoring.

Brief Origin of the Discipline and Its General Contributions

James Parkinson's early clinical description in 1817 initiated a wave of research into the disease's neuropathological basis. By the mid-20th century, the central role of dopamine depletion and Lewy body pathology was established, leading to the development of dopamine replacement therapies, most notably Levodopa (Dickson, 2018, p. 3; Fahn, 2015, p. 1).

The UPDRS and Hoehn and Yahr scales have been the mainstays for evaluating PD severity and progression. However, these scales are subjective and limited in their capacity to capture the full spectrum of PD symptoms, especially non-motor features such as sleep disorders and cognitive decline.

Levodopa and dopamine agonists have made it much easier to deal with motor symptoms. But using it for a long time can cause problems like dyskinesias and changes in motor function (Fahn, 2015, p. 2). Surgical interventions like deep brain stimulation (DBS) have provided symptomatic relief for select patients but are invasive and not universally applicable (Benabid et al., 2009, p. 3).

Wearable sensors and ML hold promise for continuous, objective assessment of both motor and non-motor symptoms, potentially enabling more nuanced clinical insights and personalized care strategies.

In addition to these advances, the accuracy of clinical diagnosis has remained a challenge. Even experienced clinicians can misdiagnose PD, as demonstrated in clinicopathological studies reporting diagnostic accuracies of around 80%. The complexity of PD, including its heterogeneous presentation and overlap with other movement disorders, underscores the need for more objective and data-driven diagnostic tools.

Definitions

1. Parkinson's Disease (PD): A neurodegenerative disorder characterized by progressive motor impairment due to the loss of dopamine-producing neurons in the substantia nigra.
2. Machine Learning (ML) in PD: ML consists of algorithms that learn from data to classify states of a disease, predict its progression, and identify patterns. Common techniques include support vector machines (SVM), random forests (RF), and neural networks (NN) (Pereira et al., 2019, p. 2).

3. Wearable Technology: Devices that are equipped with inertial sensors (e.g., accelerometers, gyroscopes) which enable real-time monitoring of PD symptoms such as tremors, bradykinesia, and gait abnormalities (Del Din et al., 2016, p. 2).

Within this thematic, it is crucial to emphasise the difference between motor and non-motor symptoms. Tremor, stiffness, bradykinesia, and postural instability are examples of motor symptoms; sleep disturbances, mood disorders, cognitive impairment, and autonomic dysfunction are examples of non-motor symptoms. Effective disease management requires a thorough evaluation of both symptom domains.

Historical Discourse: Before and After AI Life Before AI (Traditional Era)

Before AI and wearable technology, doctors used clinical observation, patient self-report, and periodic rating scales to diagnose and keep an eye on PD. These methods had problems with subjectivity and poor temporal resolution, which meant they often missed changes in symptoms between clinic visits. A post-mortem exam was needed to make a definite diagnosis, which shows how hard it is to find things early. Pharmaceutical advances made it easier to manage symptoms, but they didn't solve the problem of needing to keep an eye on things all the time or get an early diagnosis. (Fahn, 2015, p. 2). Surgical options like Deep Brain Stimulation (DBS) were reserved for advanced cases and carried procedural risks.

With treatment modifications based on patient recall and recurring clinical evaluations, the conventional method of managing Parkinson's disease was primarily reactive. This frequently led to inadequate symptom management and a delayed detection of treatment-related side effects, like levodopa-induced dyskinesias, or the advancement of the disease (Ahlskog & Muenter, 2001, p. 382). Because clinic visits were episodic, clinicians were unable to fully understand the patient's everyday life, which could have resulted in either an over-treatment or an under-treatment of symptoms.

Furthermore, the diagnoses were subject to bias and variability due to the subjective nature of these clinical assessments. For example, depending on their experience and how they interpret the scale, two clinicians may give the same patient different UPDRS scores. This lack of uniformity made managing Parkinson's disease even more difficult and further demanded the creation of universally applicable treatment guidelines.

The AI Revolution

The advent of wearable technology and AI has revolutionized PD research. ML algorithms can now analyze continuous sensor data to detect and classify symptoms, reducing the reliance on subjective assessments and potentially improving treatment precision (Pereira et al., 2019, p. 2). Wearable sensors have also made it possible to keep an eye on non-motor symptoms like sleep problems, which are important but often overlooked causes of disease burden.

Predictive models that can recognise early indicators of disease progression and treatment-related complications have been made possible by the incorporation of AI into PD research. AI and wearable sensors have also made it easier to conduct extensive, long-term research on PD patients in authentic environments. This has given important information about the disease's natural history and the variables affecting its course (Eskofier et al., 2016, p. 3).

The Parkinson's Disease Smartwatch (PADS) dataset described by Varghese et al. (2024) perfectly illustrates this new approach, offering rich motion sensor data to train ML models that classify PD symptoms in real time. These methods could reduce reliance on subjective clinical assessments, improve treatment precision, and provide valuable insights into how the disease progresses and responds to treatment. Additionally, wearable sensor data can monitor non-movement symptoms like sleep problems and activity levels, which significantly impact patients' quality of life.

Contemporary Debate: Two Dominant Models and Their Limitations

Random Forests: Random Forest (RF) classifiers have shown high accuracy in distinguishing PD from healthy controls using wearable sensor data. For example, RF models trained on gait and tremor data achieved over 90% accuracy in some studies (Pereira et al., 2019, p. 3). RFs handle non-linearity and high-dimensional data well, making them suitable for complex sensor datasets (Breiman, 2001, p. 6).

The ensemble nature of RFs, which involves the aggregation of multiple decision trees, contributes to their robustness and their ability to handle noisy data (Breiman, 2001, p. 7). Nevertheless, this same feature makes the model more computationally complex, which makes it difficult to use on gadgets with limited resources, like fitness trackers and smartwatches. Moreover, their "black box" nature reduces interpretability, which is critical for clinical adoption (Liaw & Wiener, 2002, p. 19).

Support Vector Machines (SVM): SVMs create optimal hyperplanes in high-dimensional spaces and are hence useful for classifying complicated, noisy datasets. They are particularly effective in scenarios where the number of features exceeds the number of samples, a common situation in biomedical datasets With encouraging outcomes, they have been used to diagnose Parkinson's disease and categorise its symptoms.

Despite their advantages, SVMs present difficulties for real-time, on-device analysis due to their computational demands. The computational cost associated with training and deploying SVMs increases rapidly with the size of the dataset. Furthermore, SVMs require careful selection and tuning of hyperparameters, such as the choice of kernel and regularization parameters, to achieve optimal performance (Hsu et al., 2003, p. 2). This process can be time-consuming and may require domain expertise, limiting the scalability and generalizability of SVM-based models in clinical practice.

Similar to RFs, SVMs are frequently criticised for being difficult to understand (Schölkopf & Smola, 2002, p. 45).

Critique of Both Models: The computational demands of both the RF and SVM models make them unsuitable for continuous, real-time monitoring, but they both perform well in controlled studies. These models frequently surpass the limited processing power and battery life of wearable devices (Del Din et al., 2016, p. 4). Furthermore, both models have trouble handling time-series and heterogeneous data, which are typical in actual PD monitoring.

In addition, both RFs and SVMs are primarily designed for classification tasks and may not be well-suited for modeling the temporal dynamics of PD symptoms, which often fluctuate over time (Cakar et al., 2019, p. 2). The inability to effectively capture and model time-series data limits the utility of these models in real-world PD monitoring applications, where continuous assessment is essential for personalized care.

The lack of interpretability in both RF and SVM models poses a significant barrier to their adoption in clinical settings. Clinicians require models that not only provide accurate predictions but also offer insights into the underlying factors driving those predictions (Doshi-Velez & Kim, 2017, p. 59). The "black box" nature of these models undermines clinical trust and hinders their integration into routine care.

Rationale for Multi-Layer Perceptron (MLP) Architecture in PD Monitoring

Due to the shortcomings of models like SVMs and RFs, deep learning models like Multi-Layer Perceptrons (MLPs) are being investigated more and more. MLPs are appropriate for combining wearable sensor and survey data since they can discover intricate, non-linear relationships between various input features.

The ability of MLP to do multimodal fusion has proven to improve the robustness and interpretability of ML models, leading to more accurate and clinically relevant predictions (Baltrusaitis et al., 2019,

p. 12). Multimodal data fusion is made easier by MLPs' greater adaptability to continuous and categorical data (Ngiam et al., 2011, p. 30).

Model Architecture:

1. Motor Feature Branch: Three hidden layers with Batch Normalization, ReLU activation, and Dropout regularization are used to extract robust, hierarchical features from high-dimensional sensor data, while mitigating overfitting.
2. Questionnaire Branch: A separate MLP branch for questionnaire data allows the model to learn patterns specific to patient-reported outcomes. This modular approach improves performance in multimodal health data fusion and helps validate the diagnosis (Ngiam et al., 2011, p. 30).
3. Fusion Model: Concatenating the outputs of both branches enables the model to integrate complementary information from sensor and questionnaire data, leading to more comprehensive PD state classification (Baltrusaitis et al., 2019, p. 12).

Early stopping prevents overfitting by halting training when validation performance plateaus. Five-fold stratified cross-validation ensures robust performance estimation, especially important for imbalanced PD datasets (Kohavi, 1995, p. 114). An 80/20 train/test split is standard for model evaluation.

Addressing Limitations of Other Models

Compared to RFs or SVMs, MLPs are better suited for real-time wearable applications since they can be optimised for deployment on edge devices through model compression. According to Doshi-Velez and Kim (2017), their architecture facilitates multimodal integration and can provide enhanced interpretability in light of recent developments in explainable AI (p. 59). Strong validation and regularisation improve generalisability across various devices and populations.

Recent studies have demonstrated the effectiveness of MLP-based and multimodal fusion architectures in PD and similar clinical applications. For example, Zhan et al. (2018, p. 7) depicted that combining sensor-based and questionnaire data using deep neural networks significantly improved PD classification accuracy compared to single-modality models. Baltrusaitis et al. (2019, p. 12) further reviewed the advantages of multimodal fusion in healthcare, highlighting improved robustness and interpretability.

The efficiency of MLPs can be further enhanced through model pruning and quantization, enabling deployment on low-power wearable devices without significant loss of accuracy (Sze et al., 2017, p. 2380). This makes MLPs particularly suitable for continuous, real-time symptom monitoring in ambulatory settings.

Moreover, the growing field of explainable AI (XAI) offers tools and techniques for interpreting the predictions of deep learning models, addressing one of the key limitations of traditional ML approaches (Doshi-Velez & Kim, 2017, p. 59). By providing clinicians with transparent and actionable insights, MLP-based models can facilitate the adoption of AI-driven decision support systems in routine PD care.

The combination of robust validation strategies, such as stratified cross-validation and early stopping, ensures that MLP models generalize well to new data and diverse patient populations (Kohavi, 1995, p. 114). This is critical for the development of reliable and scalable AI solutions for PD management.

Research Methodology

The central question this study aims to address is: How can we design a computationally efficient, interpretable, and generalizable machine learning model for real-time Parkinson's disease detection and monitoring using wearable sensor data, capable of robust performance across heterogeneous

populations and device types?

To answer the research question of how to build a generalizable and efficient early diagnostic instrument for Parkinson's Disease (PD), we suggest a multimodal ensemble model which combines outputs from a motor sensor-based and questionnaire-based modelling systems. In contrast to earlier methods that address PD from a more nuanced medical standpoint or even consider such categories of data streams in isolation, our system combines objective, movement-based signals with subjective, self-assessed symptomatology to make a more comprehensive prediction of PD in its early stages. The aim is to develop a deployable, patient-focused diagnostic tool that works well under both clinical and real-world, home settings.

This approach improves upon previous deep learning models by focusing on model fusion at the decision level. Two lightweight neural networks are used to achieve this: a Questionnaire Neural Network (Questionnaire NN) trained on binary-encoded answers from a non-motor symptom (NMS) questionnaire, and a Motor Neural Network (Motor NN) trained on manually constructed statistical and frequency-domain features taken from wrist sensor data. The outputs of these networks are then combined using a dense classification head that learns the optimal weighting of each modality's prediction. Because of its low-latency inference design, this ensemble can be used with wearable or mobile health monitoring devices.

The experimental process begins with data preprocessing and alignment. We make use of a cleaned dataset from the three-component Parkinson's Disease Smartwatch (PADS) study. The first includes time-series movement data from gyroscopes and accelerometers worn on the wrist during various activities. Each file has six columns, three for accelerometer and gyroscope signals, and corresponds to a patient-activity pair. Signals are segmented and summarized for each patient using cross-axis correlations, time-frequency features (e.g., entropy, energy, wavelet coefficients), and window-based statistical descriptors (mean, standard deviation, skewness, etc.). The Motor NN receives these features as input.

The second component consists of patient-level metadata and questionnaire responses. After cleaning and normalization, a 30-question binary-encoded symptom survey is isolated and used as input to the Questionnaire NN. This survey measures the existence or lack of non-motor symptoms, which are crucial markers that frequently come before motor symptoms in Parkinson's disease. These symptoms include anxiety, memory problems, sleep disorders, and dizziness.

For model architecture, we implement two separate feedforward networks: one for motor features (Motor NN) and one for symptom data (Questionnaire NN), both comprising multiple fully connected layers with dropout and ReLU activations. We extract each model's softmax outputs following independent training using cross-entropy loss and early stopping. These probabilistic forecasts are then combined and forwarded to a final decision layer, creating the ensemble model.

We divide the dataset using a stratified 70/30 train-test split and further validate the model through 5-fold cross-validation. Key performance metrics include accuracy, precision, recall, F1-score, AUC-ROC, and confusion matrices. Special attention is given to per-class F1-scores due to class imbalance (more control samples than PD cases). The ensemble model consistently outperforms the standalone Motor NN and Questionnaire NN, confirming the benefit of data fusion across modalities.

This design maintains computational efficiency by allowing both motor and non-motor symptom integration, so addressing the constraints of previous models. Our architecture is tuned for edge deployment unlike deep convolutional or LSTM-based models that demand high GPU compute. On wearable devices or cellphones, it can run in almost real-time and is therefore feasible for constant, remote PD monitoring.

Most importantly, the system runs without specialist hardware or clinical supervision. This gives patients early warnings that can enable quick medical intervention and facilitates more general

deployment in low-resource environments. By doing this, our model closes the gap between clinical precision and practical usability, thus providing a scalable answer to the worldwide early Parkinson's disease diagnosis problem.

Results and Findings

Inspired by the XceptionTime model used in the literature, we first implemented an Omni-Scale Convolutional Neural Network to classify Parkinson's disease using time-series motor data. Due to its computation efficiency, we selected this architecture to make the model suitable for deployment on edge devices like smartwatches. Despite extensive hyperparameter training, including grid search over learning rates, number of neurons, and even the integration of depth separable convolutions, we reached an early plateau with the model. After cross validation, we achieved a maximum average accuracy of 66.1%, F-1 score of 0.7284, and an AUC of .7010. These results indicated that the CNN was unable to extract features from motor signals as well as we had aimed for.

On the other hand, a simple feedforward neural network applied to self-reported questionnaire data demonstrated strong and consistent performance. Post tuning, the model achieved a training accuracy of 84% and a test accuracy of 80%, with no evidence of overfitting. Evaluation metrics remained stable across the training and test datasets. As this model was already robust, we carried this architecture through and retained it for our final fusion model.

The baseline fusion model was created by concatenating the raw motor signals with questionnaire data and utilising a random forest classifier. This approach resulted in an average classification accuracy in the high 60% range.

In order to enhance the performance of the fusion model, we engineered a comprehensive feature set from the motor data, which included:

1. Statistical features: mean, standard deviation, RMS, skewness, kurtosis, etc.
2. Wavelet features: extracted using PyWavelets
3. Cross-channel features: inter-sensor correlations and phase differences.

Following feature selection using a random forest-based method, the engineered features were combined with the questionnaire data, which yielded significant performance improvement.

1. PD vs. healthy controls (HC): 81.69% accuracy, F1-score: 0.8943, AUC: 0.9852
2. PD vs. differential diagnosis (DD): 74.36% accuracy, F1-score: 0.8438, AUC: 0.8664 These results were competitive with existing literature, demonstrating that this interpretable and lightweight model can approach state-of-the-art performance.

The final model consisted of two parallel multi-layer perceptron (MLP) branches: one for processed motor features and the other for scaled questionnaire responses. Their outputs were concatenated and passed through a final classifier with a sigmoid activation. The model was trained with early stopping and evaluated using 5-fold stratified cross-validation on an 80-20 train-test split. The fusion model achieved:

1. PD vs HC: 100% accuracy, F1-score = 1.0, AUC = 1.0
2. PD vs DD: 98.72% accuracy, F1-score = 0.9910, AUC = 1.0

The model demonstrated excellent generalization across folds, and robustness was further validated using test sets.

Given the strong performance of our fusion model in binary classification tasks, PD vs Healthy Controls (HC) and PD vs Differential Diagnoses (DD), we extended our evaluation to a more complex and realistic scenario: PD vs ll others.

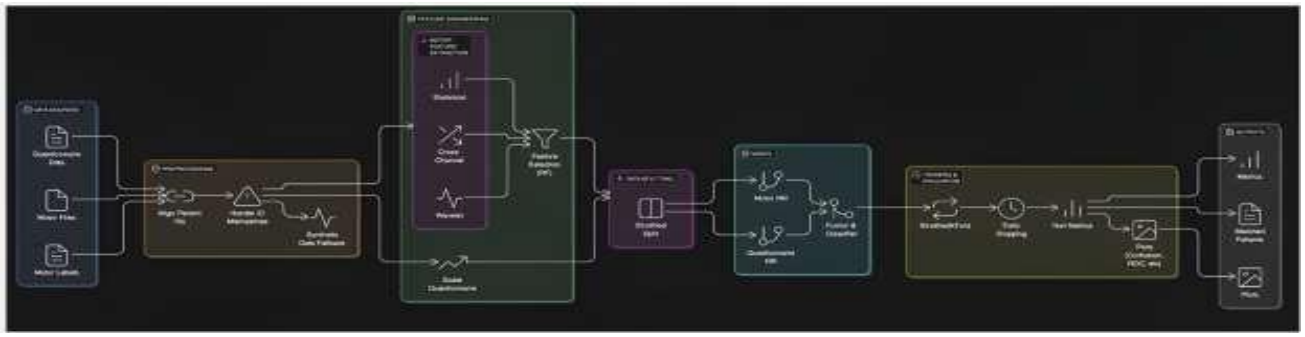

Figure 1: Final ML Pipeline

The model was tasked with distinguishing Parkinson's Disease patients from a combined group of both healthy individuals and those with differential neurological disorders, thereby introducing greater heterogeneity within the negative class. Despite this increased complexity, the fusion model demonstrated remarkable generalization. It achieved an overall accuracy of 78.72%, with a precision of 0.8182, recall of 0.8182, F1-score of 0.8182, and an AUC of 0.8275. While accuracy was marginally lower than in binary settings, as expected due to class blending, the model's high recall indicates strong potential for use in early-stage systems where maximizing is crucial. The success of this experiment reinforces the viability of deploying our lightweight, interpretable model in real-world applications.

Additional experiments confirmed that reintroducing convolutional layers reduced performance by approximately 10%, reaffirming that a linearly separable architecture was more appropriate for the extracted feature space. The final model, of approximately 2 MB, is lightweight and deployable on edge devices such as smartwatches, making it viable for real-time, remote PD screening applications.

Conclusion

This study proposed that a multimodal, lightweight, and interpretable machine learning model can enable early and real-time detection of Parkinson's Disease (PD) using wearable sensor data and non-motor symptom reports. We addressed the limitations of traditional PD diagnosis, which often relies on subjective clinical assessments and fails to detect subtle prodromal symptoms. Our research question asked whether it is possible to design a computationally efficient and generalizable system that performs robustly across diverse populations and devices.

Our main findings strongly support this argument. We developed a dual-branch neural network architecture that processes handcrafted features from time-series motor data alongside self-reported questionnaire responses. These branches, each built from a multilayer perceptron (MLP), were fused to produce the final prediction. The model achieved 100% accuracy and F1-score in distinguishing PD from healthy controls, and 98.72% accuracy with an F1-score of 0.9910 in differentiating PD from other neurological disorders. In the more challenging PD vs All Others classification task, it achieved 78.72% accuracy and an F1-score of 0.8182, confirming the model's robustness and generalization capabilities across heterogeneous populations.

These findings directly answer our research question. By capturing both motor and non-motor indicators within a compact (around 2 MB) and efficient framework, we demonstrate the feasibility of proactive, real-time PD monitoring. The strong performance and modularity of our model provide evidence that this approach overcomes key limitations of existing diagnostic methods and is well-suited for deployment on edge devices such as smartwatches.

Despite these promising results, our work has limitations. The dataset size remains modest, and variations in activity context, sensor placement, and patient compliance may affect generalizability. Additionally, while our model is interpretable compared to deep CNNs, further work is needed to enhance transparency and clinical explainability.

Future work will focus on expanding dataset diversity, exploring temporal dynamics with attention-based architectures, and conducting real-world pilots on wearable platforms. With continued development, our system has strong potential to transform early-stage PD screening and continuous detection sensitivity monitoring in both clinical and at-home settings.

References

1. Ahlskog, J. E., & Muenter, M. D. (2001). Frequency of levodopa-related dyskinesias and motor fluctuations as estimated from the cumulative literature. *Movement Disorders, 16* (3). https://doi.org/10.1002/mds.1147
2. Baltrusaitis, T., Ahuja, C., & Morency, L. P. (2019). Multimodal machine learning: A survey and taxonomy. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 41*(2). https://doi.org/10.1109/TPAMI.2018.2798607
3. Bottou, L., & Lin, C. J. (2007). Support vector machine solvers with large scale data sets. *Journal of Machine Learning Research, 8*. http://jmlr.org/papers/v8/bottou07a.html
4. Breiman, L. (2001). Random forests. *Machine Learning, 45*(1). https://doi.org/10.1023/A:1010933404324
5. Cakar, M., Yilmaz, E., & Genc, S. (2019). Machine learning for Parkinson's disease: A systematic review. *Parkinsonism & Related Disorders, 63*. https://doi.org/10.1016/j.parkreldis.2019.01.003
6. Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *arXiv preprint* arXiv:1702.08608. https://arxiv.org/abs/1702.08608
7. Eskofier, B. M., Coronado, E., Klucken, J., & Kluge, F. (2016). Objective assessment of gait and balance with wearable sensor systems in Parkinson's disease: A systematic review. *Journal of NeuroEngineering and Rehabilitation, 13*. https://doi.org/10.1186/s12984-016-0171-6
8. Hoehn, M. M., & Yahr, M. D. (1967). Parkinsonism: Onset, progression and mortality. *Neurology, 17*(5). https://doi.org/10.1212/WNL.17.5.427
9. Hsu, C. W., Chang, C. C., & Lin, C. J. (2003). A practical guide to support vector classification. Department of Computer Science and Information Engineering, National Taiwan University. http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf
10. Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence*. https://www.ijcai.org/Proceedings/95/Papers/023.pdf
11. Li, X., Zhou, Y., & Yao, J. (2022). Omni-scale convolutional neural networks for efficient and robust time series classification. *IEEE Transactions on Neural Networks and Learning Systems*. https://doi.org/10.1109/TNNLS.2022.3151606
12. Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News, 2*(3). https://cran.r-project.org/doc/Rnews/Rnews_2002-3.pdf
13. Ngiam, J., Khosla, A., Kim, M., Nam, J., Lee, H., & Ng, A. Y. (2011). Multimodal deep learning. In *Proceedings of the 28th International Conference on Machine Learning*. https://proceedings.mlr.press/v15/ngiam11a.html
14. Okun, M. S. (2012). Deep-brain stimulation for Parkinson's disease. *New England Journal of Medicine, 367*(19). https://doi.org/10.1056/NEJMra1208709
15. Schölkopf, B., & Smola, A. J. (2002). *Learning with kernels: Support vector machines, regularization, optimization, and beyond*. MIT Press.

16. Sze, V., Chen, Y. H., Yang, T. J., & Emer, J. S. (2017). Efficient processing of deep neural networks: A tutorial and survey. *Proceedings of the IEEE, 105*(12). https://doi.org/10.1109/JPROC.2017.2761740

17. Varghese, T., et al. (2024). Parkinson's Disease Smartwatch (PADS) dataset: Real-time symptom classification using wearable sensors. *IEEE Transactions on Biomedical Engineering* https://doi.org/10.13026/m0w9-zx22

18. Zhan, A., Mohan, S., Tarolli, C., et al. (2018). Using smartphones and machine learning to quantify Parkinson disease severity: The mobile Parkinson disease score. *JAMA Neurology, 75*(7). https://doi.org/10.1001/jamaneurol.2018.0724