Exploring the Potential of Deep Learning in the Classification and Early Detection of Parkinson's Disease

V. S. Bakkialakshmi¹, V. Arulalan², Gowdham Chinnaraju³, Hritwik Ghosh⁴*, Irfan Sadiq Rahat⁵, Ankit Saha⁶

¹,²Department of Computing Technologies, SRM Institute of Science and Technology, Kattankulathur, Tamil Nadu
³Department of Computer Science and Engineering, AMET University, Kanathur-603112, India
⁴,⁵,⁶School of Computer Science and Engineering (SCOPE), VIT-AP University, Amaravati, Andhra Pradesh

Abstract

INTRODUCTION: Parkinson's Disease (PD) is a progressive neurological disorder affecting a significant portion of the global population, leading to profound impacts on daily life and imposing substantial burdens on healthcare systems. Early identification and precise classification are crucial for effectively managing this disease. This research investigates the potential of deep learning techniques in facilitating early recognition and accurate classification of PD.

OBJECTIVES: The primary objective of this study is to leverage advanced deep learning techniques for the early detection and precise classification of Parkinson's Disease. By utilizing a rich dataset comprising speech signal features extracted from 3000 PD patients, including Time Frequency Features, Mel Frequency Cepstral Coefficients (MFCCs), Wavelet Transform based Features, Vocal Fold Features, and TWQT features, this research aims to evaluate the performance of various deep learning models in PD classification.

METHODS: The dataset containing diverse speech signal features from PD patients' recordings serves as the foundation for training and evaluating five different deep learning models: ResNet50, VGG16, Inception v2, AlexNet, and VGG19. Each model undergoes training and assessment to determine its capability in accurately classifying PD patients. Performance metrics such as accuracy are employed to evaluate the models' effectiveness.

RESULTS: The results demonstrate promising potential, with overall accuracies ranging from 89% to 95% across the different deep learning models. Notably, AlexNet emerges as the top-performing model, achieving an accuracy of 95% and demonstrating balanced performance in accurately identifying both true and false PD cases.

CONCLUSION: This research highlights the significant potential of deep learning in facilitating the early detection and classification of Parkinson's Disease. Leveraging speech signal features offers a non-invasive and cost-effective approach to PD assessment. The findings contribute to the growing body of evidence supporting the integration of artificial intelligence in healthcare, particularly in the realm of neurodegenerative disorders. Further exploration into the application of deep learning in this domain holds promise for advancing PD diagnosis and management.

Keywords: Parkinson's Disease, ResNet50, VGG16, Inception v2, AlexNet, and VGG19, DL, accuracy, healthcare

Received on 24 December 2023, accepted on 21 March 2024, published on 27 March 2024

Copyright © 2024 V. S. Bakkialakshmi et al., licensed to EAI. This is an open access article distributed under the terms of the CC BY-NC-SA 4.0, which permits copying, redistributing, remixing, transformation, and building upon the material in any medium so long as the original work is properly cited.

doi: 10.4108/eetpht.10.5568

*Corresponding authors Email: me.hritwikghosh@gmail.com

1. Introduction

Parkinson's Disease (PD) is a persistent, advancing neurologic condition that primarily impairs motor abilities, leading to signs such as shaking, inflexibility, and slowed movements. It ranks as the second most frequent neurodegenerative ailment globally, affecting approximately 1% of individuals aged 60 years and above. Despite thorough scientific investigations, the initial detection and precise categorization of PD persist as formidable challenges, largely due to the intricate characteristics of the disease and the wide spectrum of its clinical symptoms. It has shown promising results in various fields of medicine, including neurology. However, its application in the early detection and classification of PD is still in the nascent stages and...
warrants further exploration. Speech impairments, such as hypokinetic dysarthria, are common in PD patients and often precede motor symptoms. These impairments can be quantified using various speech signal processing algorithms, providing a non-invasive and cost-effective method for PD assessment. In this study, we utilized a rich dataset of speech signal features from 3000 PD patients, including Time Frequency Features, Mel Frequency Cepstral Coefficients (MFCCs), Wavelet Transform based Features, Vocal Fold Features, and TWQT features. These features were extracted from the speech recordings of PD patients, providing a comprehensive basis for our deep learning models. We employed five different deep learning models: ResNet50, VGG16, Inception v2, AlexNet, and VGG19. Each model was trained and evaluated on its ability to classify PD patients accurately. The primary objective of this study is to explore the potential of deep learning in the early detection and classification of PD. We aim to contribute to the growing body of evidence supporting the integration of AI in healthcare and open avenues for further exploration into the application of deep learning in neurodegenerative disorders.

2. Literature Review

Niall P. Quinn (1998) [1] emphasises the intricacy of clinical variations that PD patients experience in his study on swings in the condition. Fluctuations can be short-lived, like freezing and paradoxical kinesia, or they can be medium-lived, like the wearing-off and "on-off" effects of L-dopa medication. Additionally, responses with a long duration of up to two weeks are seen. As effects may take weeks to equilibrate in the brain, these fluctuations need to be carefully taken into account during treatment optimisation. These swings frequently come with dyskinesias, alterations in mood, and discomfort, highlighting the complexity of PD care.

The authors of the research (A. Tsanas et al., 2012) [2] explore the connection between speech impairment and Parkinson's disease (PD). In order to use speech signals to predict the severity of PD, they investigate unique speech signal processing algorithms. The study uses choosing features and analytical classifiers to extract 132 dysphonia metrics from sustained vowels. Surprisingly, these new metrics outperform prior results by diagnosing PD cases and healthy controls with approximately 99% accuracy. The results show how useful these indicators could be in making non-invasive Parkinson's disease diagnosis judgements.

The authors of the study (Pasha & Latha, 2020) [3] use the bio-inspired algorithms Binary Particle Swarm Optimisation (BPSO) and Genetic Algorithm (GA) to optimise feature subsets for effective PD classification. They make use of a dataset that includes 755 attributes and 756 observations from 252 PD patients. They choose the best feature subsets using cross-validation and 11 machine learning classifiers, obtaining up to 90.7% classification accuracy and a dimensionality reduction of up to 52.32%. These results highlight how bio-inspired approaches may improve PD classification accuracy by carefully choosing features.

The authors of the paper (Yuanxi Li & Tucker, 2010) [4] offer a novel method for identifying illness foci by exploiting clinical trial data's pseudo-time-series trajectories. They create pseudo time-series from cross-sectional data by fusing distance measurements, graphing processes, and resampling methods. These techniques are expanded to automatically identify illness zones at pivotal junctures and trajectories' ends. They illustrate the efficacy of this approach in recognising illness states and transitions between them by testing it on several medical datasets, including those for Glaucoma, Parkinson's illness, and breast cancer. The study focuses on automating the detection of illness states as well as building models of time-series from cross-sectional data, offering important insights for disease modelling.

In their study (P. Drotár et al., 2013) [5] the authors develop a unique method for assessing Parkinson's disease (PD) using in-air handwriting movements, focusing on the condition's early symptoms. To accurately identify PD, the study suggests examining in-air trajectories while a person is writing. Their research demonstrates how in-air trajectories can be used to identify modest motor anomalies associated with PD. In-air trajectories and conventional on-surface handwriting analysis are combined, and the result is a predictive model that classifies PD with over 80% accuracy. The authors use svm for binary categorization after selecting characteristics from more than 600 handwriting features using the Mann-Whitney U-test filter and relief procedure.

Rehman et al. (2019) [10] identified critical gait traits for differentiating early Parkinson's disease from normal controls using machine learning approaches. Five essential gait characteristics were identified, and they provided excellent sensitivity and specificity early PD classification accuracy of 73–97%. Their research emphasises the possibility of these characteristics for a precise, early diagnosis of PD.

For the purpose of diagnosing epilepsy, the authors of the study (Acharya et al., 2018) [11] put special emphasis on the characteristics of focal (F) and non-focal (NF) EEG signals. They create a computer-aided detection technique that can accurately locate epileptogenic regions by separating F from NF signals using nonlinear characteristics. The study highlights the possible use of nonlinear characteristics in identifying hidden EEG signal patterns for precise clinical interpretation.

A DL approach for automated Parkinson's disease (PD) diagnosis utilising EEG signals is presented by Oh et al. (2020) [12]. They successfully identify between PD and normal participants in their study using a thirteen-layer convolutional neural network (CNN), obtaining excellent precision of 88.25%, a sensitivity of 84.71%, and specificity of 91.77%. This method does away with the necessity for
traditional feature extraction and has the potential for extensive clinical use.

Complex Parkinson's disease (PD) manifests in a variety of ways clinically. Early diagnosis is difficult because of symptoms that overlap and because of unusual motor and cognitive outcomes. Misdiagnosis is common in PD patients due to the prevalence of low mood, anhedonia, motivation loss, and psychomotor slowness. According to Madabushi et al. (2023) [13], making distinctions between apathy, anhedonia, and alexithymia is essential for a precise diagnosis. Parkinson's disease, a complex illness associated with motivation loss, is explored by Béreau et al. (2023) [14] by examining the significance of apathy. The article talks about the stages of apathy and how it changes from a dopa-responsive motivation deficit to a cognitive decline sign in advanced Parkinson's disease.

Jiang-ting Li et al. (2023) [15] examined demographic data, Apathy Scale ratings, and serum iron metabolism indicators in 201 Parkinson's disease patients. They looked the connections among male and female iron metabolism characteristics and Apathy Scale scores using Spearman correlations. They created a nomogram utilising the findings from logistic regression to help distinguish PD patients who are disinterested. Emmie Cohen et al. (2022) [16] looked on the symptoms of apathy in various phases of Parkinson's disease (PD). Their research showed that mild PD patients exhibit apathy-related symptoms, such as a loss of pleasure and energy, early on. The correlation between these emotions and disease severity emphasises the significance of understanding how indifference affects patients' quality of life for prompt diagnosis and care.

Using rs-fMRI data, Xu et al. (2022) [17] examined the brain activity of PD-A individuals with Parkinson's disease. They discovered decreased activity in particular brain regions in PD-A compared to those with Parkinson's without apathy (PD-NA), as well as changed activity in regions including the centre of the occipital gyrus and insula. Notably, aberrant connection patterns between PD-A and healthy controls were found in the hippocampus, parahippocampus, and cerebellum.

In a randomised clinical experiment, Kulisevsky et al. (2022) [18] investigated the effects of safinamide on Parkinson's disease (PD) patients who were not yet demented and showed apathy. Safinamide's effects on apathy, cognition, everyday activities, motor scores, and safety were examined during the 24-week trial. A post-hoc analysis revealed that the Apathy Scale (AS) score significantly improved in the safinamide group, despite the fact that the change in AS score only became towards significance (p = 0.059). Secondary outcome variables generally stayed the same. Both treatment groups experienced only minor adverse outcomes.

In their discussion of managing neuropsychiatric signs in Parkinson's disease, Daniel Weintraub et al. (2022) [20] emphasise the symptoms' frequent incidence, lack of diagnosis, and likely neurobiological foundation. They promote a thorough strategy combining a range of actions for successful treatment and enhanced care for patients [Table.1].

<table>
<thead>
<tr>
<th>Reference</th>
<th>Focus of Study</th>
<th>Techniques Used</th>
<th>Key Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tsanas et al. (2012) [2]</td>
<td>Connection between speech impairment and PD</td>
<td>Speech signal processing algorithms</td>
<td>New dysphonia metrics outperform prior results, diagnosing PD cases and healthy controls with approximately 99% accuracy.</td>
</tr>
<tr>
<td>Reference</td>
<td>Methodology</td>
<td>Techniques</td>
<td>Findings</td>
</tr>
<tr>
<td>-----------</td>
<td>-------------</td>
<td>------------</td>
<td>----------</td>
</tr>
<tr>
<td>Drotár et al. (2013) [5]</td>
<td>Assessing PD using in-air handwriting movements</td>
<td>SVM for binary categorization</td>
<td>In-air trajectories combined with on-surface handwriting analysis result in a predictive model with over 80% accuracy.</td>
</tr>
<tr>
<td>Madabushi et al. (2023) [13]</td>
<td>Distinguishing apathy, anhedonia, and alexithymia in PD</td>
<td>Not specified</td>
<td>Emphasizes the importance of making distinctions between apathy, anhedonia, and alexithymia for a precise PD diagnosis.</td>
</tr>
<tr>
<td>Béreau et al. (2023) [14]</td>
<td>Exploring the significance of apathy in PD</td>
<td>Not specified</td>
<td>Examines the stages of apathy and its changes in advanced Parkinson's disease.</td>
</tr>
<tr>
<td>Jiang-ting Li et al. (2023) [15]</td>
<td>Examining connections among iron metabolism and apathy</td>
<td>Spearman correlations, logistic regression</td>
<td>Nomogram created to help distinguish PD patients with apathy based on iron metabolism characteristics.</td>
</tr>
<tr>
<td>Emmie Cohen et al. (2022) [16]</td>
<td>Symptoms of apathy in various phases of PD</td>
<td>Not specified</td>
<td>Mild PD patients exhibit apathy-related symptoms early on, emphasizing the significance of understanding how indifference affects patients' quality of life.</td>
</tr>
<tr>
<td>Xu et al. (2022) [17]</td>
<td>Examining brain activity in PD-A individuals</td>
<td>rs-fMRI data analysis</td>
<td>Identifies decreased activity in specific brain regions in PD-A compared to PD-NA, highlighting aberrant connection patterns.</td>
</tr>
<tr>
<td>Kulisevsky et al. (2022) [18]</td>
<td>Effects of safinamide on apathy in PD patients</td>
<td>Randomized clinical experiment</td>
<td>Safinamide group shows improvement in Apathy Scale (AS) score, suggesting potential benefits for apathy in PD patients.</td>
</tr>
</tbody>
</table>
3. Dataset Overview

The dataset employed in this research is a comprehensive collection of speech signal features from 3000 patients diagnosed with Parkinson's Disease (PD). This dataset is unique in its breadth and depth, encompassing a wide demographic range with ages spanning from 20 to 85, with an average age of 65.1±10.9. The gender distribution includes 2000 men and 1000 women, providing a diverse sample for our analysis. The speech signal features in this dataset are derived from various signal processing algorithms, each providing distinct insights into the speech characteristics of PD patients. Time Frequency Features provide information about the frequency content of the speech signal over time, capturing the dynamic changes in speech that are often observed in PD patients. MFCCs are commonly used in speech and audio processing and can capture the unique vocal characteristics of PD patients. Wavelet Transform based Features allow for the analysis of non-stationary signals, such as speech, and can reveal subtle changes in speech patterns. Vocal Fold Features provide information about the functioning of the vocal folds, which can be affected in PD. Lastly, TWQT features are a novel set of features that have shown promise in the analysis of PD speech. This extensive and varied dataset serves as a solid base for our deep learning algorithms. By utilizing these all-encompassing speech signal attributes, our objective is to delve into the capabilities of deep learning in facilitating the early recognition and precise categorization of PD. The employment of this dataset highlights our dedication to devising non-intrusive, economically viable techniques for PD evaluation, contributing to the overarching aim of enhancing the life quality for individuals afflicted with this neurodegenerative condition.

3.1 Methodology

The crux of this research lies in the application of deep learning methodologies to a robust dataset of speech signal attributes from 3000 individuals diagnosed with Parkinson's Disease (PD). The dataset encapsulates a wide demographic spectrum, with ages ranging from 20 to 85 and a gender distribution of 2000 males and 1000 females. The speech signal attributes were derived using a variety of signal processing algorithms, including Time Frequency Features, Mel Frequency Cepstral Coefficients (MFCCs), Wavelet Transform based Features, Vocal Fold Features, and TWQT features. These attributes offer a comprehensive portrayal of the speech characteristics of PD patients, capturing the unique vocal alterations associated with this disease. In this study, we employed five distinct deep learning models: ResNet50, VGG16, Inception v2, AlexNet, and VGG19. Each model was trained on the dataset with the objective of learning to classify PD patients accurately based on their speech signal attributes. The models were subsequently evaluated on their classification performance, providing a comparative analysis of their efficacy. The process of model training necessitated the partitioning of the dataset into two subsets: a training subset for model training and a testing subset for performance evaluation. A suite of metrics, including accuracy, precision, recall, and F1 score, were employed to assess the performance of the models. This methodology offers a robust approach to investigating the potential of deep learning in the early identification and accurate classification of PD. By leveraging a comprehensive dataset and a variety of deep learning models, this research contributes to the ongoing efforts to devise non-invasive, cost-effective methods for PD assessment.

3.2 Data Preprocessing

The initial step in our methodology involved the preprocessing of our dataset, a crucial phase that ensures the quality and reliability of the data before it is input into our deep learning models. The dataset, comprised of speech signal features from 3000 PD patients, underwent a series of preprocessing steps to ensure it was suitable for analysis. Firstly, we performed a thorough cleaning of the data. This involved checking for any inconsistencies, errors, or missing values within the dataset. Any missing values were handled appropriately to ensure they did not impact the performance of our models. In cases where data was missing, we employed imputation techniques suitable for the nature of the data. Next, we conducted a normalization process on the numerical data. Given the diverse range of speech signal features, normalization was crucial to ensure all features had the same scale. This prevents any one feature from dominating others in the model due to differences in their scale. Additionally, we performed a feature selection process. While all features in our dataset provide valuable information, not all were equally relevant to the classification of PD. Feature selection allowed us to reduce the dimensionality of our data and focus on the most informative features, improving the efficiency and performance of our models. Finally, we partitioned our dataset into a training set and a testing set. This is a standard practice in ML and DL studies, allowing us to train our models on one subset of the data (the training set) and then evaluate their performance on a separate subset (the testing set). Through these preprocessing steps, we ensured our dataset was clean, reliable, and ready for analysis with our DL models. This rigorous approach to data preprocessing underscores our commitment to conducting high-quality, reliable research.

4. Performance of All Models

In this study, we employed five different deep learning models: ResNet50, VGG16, Inception v2, AlexNet, and VGG19. Each model was evaluated based on its classification performance, using metrics such as accuracy, precision, recall, F1 score, and support.

- **ResNet50:** This model achieved an overall accuracy of 91%. The precision was 80% for false cases and 94% for true cases.

EAI Endorsed Transactions on Pervasive Health and Technology | Volume 10 | 2024 |
for true cases. The recall was 77% for false cases and 95% for true cases. The F1 score, which is the harmonic mean of precision and recall, was 79% for false cases and 95% for true cases. The support, which is the number of actual occurrences of the class in the dataset, was 31 for false cases and 121 for true cases [Fig.1,2].

![Confusion Matrix for ResNet50 Model](image1)

**Fig 1: Confusion Matrix for ResNet50 Model**

![TPR vs. FPR: Discriminative Performance Analysis for ResNet50 Model](image2)

**Fig 2: TPR vs. FPR: Discriminative Performance Analysis for ResNet50 Model**

- **VGG16:** This model achieved an overall accuracy of 89%. The precision was 83% for false cases and 91% for true cases. The recall was 61% for false cases and 97% for true cases. The F1 score was 70% for false cases and 94% for true cases. The support was 31 for false cases and 121 for true cases [Fig.3,4].

![Confusion Matrix for VGG16 Model](image3)

**Fig 3: Confusion Matrix for VGG16 Model**

![TPR vs. FPR: Discriminative Performance Analysis for VGG16 Model](image4)

**Fig 4: TPR vs. FPR: Discriminative Performance Analysis for VGG16 Model**

- **Inception v2:** This model achieved an overall accuracy of 92%. The precision was 88% for false cases and 93% for true cases. The recall was 71% for false cases and 98% for true cases. The F1 score was 79% for false cases and 95% for true cases. The support was 31 for false cases and 121 for true cases [Fig.5,6].

![Confusion Matrix for Inception v2 Model](image5)

**Fig 5: Confusion Matrix for Inception v2 Model**
Exploring the Potential of Deep Learning in the Classification and Early Detection of Parkinson's Disease

Fig 6: TPR vs. FPR: Discriminative Performance Analysis
Inception v2 Model

- **AlexNet**: This model achieved the highest overall accuracy of 95%. The precision was 87% for false cases and 97% for true cases. The recall was 87% for false cases and 97% for true cases. The F1 score was 87% for false cases and 97% for true cases. The support was 31 for false cases and 121 for true cases [Fig.7,8].

- **VGG19**: This model achieved an overall accuracy of 92%. The precision was 91% for false cases and 92% for true cases. The recall was 74% for false cases and 98% for true cases. The F1 score was 81% for false cases and 95% for true cases. The support was 53 for false cases and 174 for true cases [Fig.9,10].

Fig 7: Confusion Matrix for AlexNet Model

Fig 8: TPR vs. FPR: Discriminative Performance Analysis for AlexNet Model

Fig 9: Confusion Matrix for VGG19 Model

Fig 10: TPR vs. FPR: Discriminative Performance Analysis for VGG19 Model

From these results, it is evident that AlexNet performed the best overall, with the highest accuracy and balanced performance across both classes. However, the choice of the best model may also depend on the specific requirements of the study. For instance, if it's more important to correctly identify all true cases, then a model like Inception v2 or VGG16, which have higher recall for the true class, might be preferred.

5. Results and Discussion

The application of five different DL models to our dataset yielded promising results, demonstrating the potential of these techniques in the early detection and accurate classification of Parkinson's Disease (PD). ResNet50, VGG16, Inception v2, AlexNet, and VGG19 were the models employed in this study. Each model was trained on the dataset and evaluated based on its classification performance. The performance metrics used for evaluation...
included accuracy, precision, recall, and F1 score. ResNet50 achieved an overall accuracy of 91%, with a precision of 80% and 94% for false and true cases respectively. The recall for this model was 77% for false cases and 95% for true cases. VGG16, on the other hand, achieved an overall accuracy of 89%, with a precision of 83% for false cases and 91% for true cases. The recall for this model was 61% for false cases and 97% for true cases. Inception v2 achieved an overall accuracy of 92%, with a precision of 88% for false cases and 93% for true cases. The recall for this model was 71% for false cases and 98% for true cases. AlexNet achieved the highest overall accuracy of 95%, with a precision and recall of 87% for false cases and 97% for true cases. Finally, VGG19 achieved an overall accuracy of 92%, with a precision of 91% for false cases and 92% for true cases. For the recall this model was 74% for false cases and 98% for true cases. From these results, it is evident that AlexNet performed the best overall, with the highest accuracy and balanced performance across both classes. However, the choice of the best model may also depend on the specific requirements of the study. For instance, if it’s more important to correctly identify all true cases, then a model like Inception v2 or VGG16, which have higher recall for the true class, might be preferred. These findings underscore the potential of deep learning in the early detection and accurate classification of PD. The use of speech signal features provides a non-invasive and cost-effective method for PD assessment. This research contributes to the growing body of evidence supporting the integration of artificial intelligence in healthcare and opens avenues for further exploration into the application of deep learning in neurodegenerative disorders.

6. Conclusion and Future Work

This research has demonstrated the potential of deep learning techniques in the early detection and accurate classification of Parkinson’s Disease (PD). The application of five different deep learning models, namely ResNet50, VGG16, Inception v2, AlexNet, and VGG19, to a comprehensive dataset of speech signal features from 3000 PD patients, yielded promising results. Among the models, AlexNet emerged as the top-performing model, achieving the highest overall accuracy of 95% and balanced performance in identifying both true and false cases. However, the choice of the best model may also depend on the specific requirements of the study. For instance, if it’s more important to correctly identify all true cases, then a model like Inception v2 or VGG16, which have higher recall for the true class, might be preferred. These findings underscore the potential of deep learning in the early detection and accurate classification of PD. The use of speech signal features provides a non-invasive and cost-effective method for PD assessment. This research contributes to the growing body of evidence supporting the integration of artificial intelligence in healthcare and opens avenues for further exploration into the application of deep learning in neurodegenerative disorders. Looking forward, there is a need for further research to refine these models and explore other deep learning techniques. Future work could also involve the integration of other types of data, such as genetic or imaging data, to improve the performance of the models. Additionally, the development of a user-friendly interface for these models could facilitate their use in clinical settings, contributing to the early detection and management of PD.

References

[13] Haikun Xu, Mengchao Zhang, Ziju Wang, Yanyan Yang, Ying Chang, Lin Liu, Abnormal brain activities in multiple