Automated Detection of Parkinson's Disease Patterns in Spiral Drawings Using Convolutional Neural Networks

Ishan Dave*
University of California, Berkeley
San Jose, USA
idave0425@gmail.com
ORCID: 0009-0009-4266-5610

***Abstract*—Parkinson's disease ranks as the second most common neurodegenerative disorder, affecting approximately one million people in the United States and resulting in approximately $52 billion in annual healthcare costs. Conventional diagnostic methods for Parkinson's, including magnetic resonance imaging and dopamine transporter scans, are both invasive and expensive. This paper presents an automated approach for the diagnosis of Parkinson's based on the analysis of spiral drawing patterns using deep learning. We utilized a publicly accessible dataset that contains spiral drawings from both healthy subjects and Parkinson's patients. Classification models were developed using the MobileNetV2 and ResNet50 architectures. Our experimental results demonstrated that MobileNetV2, with its lighter framework, achieved a test accuracy of 91%, outperforming a more complex ResNet50 model. These findings indicate that CNNs, particularly MobileNetV2, provide a viable and less burdensome method for early detection of Parkinson's disease, potentially alleviating the economic and physical impacts of traditional diagnostic techniques. To our knowledge, this is the first study to apply MobileNetV2 to coordinate-rendered spiral images derived from digitized handwriting data for Parkinson's detection.**

***Index Terms*—Parkinsons, convolutional neural networks, transfer learning, image classification**

I. Introduction

Parkinson's disease, the second-most common neurodegenerative disorder after Alzheimer's, affects approximately one million Americans and ten million individuals worldwide [1]. This disease results from the degeneration of dopamine-producing neurons in the brain, leading to significant motor symptoms such as tremors, rigidity, bradykinesia, and postural instability, alongside non-motor symptoms including cognitive decline and mood disorders [2]. The condition predominantly impacts individuals over the age of 60, necessitating early detection to mitigate its progression [3].

The economic burden of Parkinson's is substantial, with annual healthcare costs reaching nearly $52 billion, influenced by both direct and indirect costs [4]. The prevalence of Parkinson's varies significantly with population density; states with larger populations like California and Texas report over 27,300 cases, whereas less populated states such as Wyoming and North Dakota record fewer than 5,200 cases. The incidence rate shows a gender disparity with a 3:2 ratio between biological males and females, with about 90,000 new diagnoses each year [5]. Statistical forecasts suggest an increase in Parkinson's cases to approximately 1.2 million by 2030 in the US alone, underscoring the growing need for effective diagnostic and treatment strategies [6].

The diagnosis of PD is primarily clinical, relying on the identification of its cardinal motor symptoms: tremor, rigidity, bradykinesia, and postural instability. However, these symptoms often manifest after substantial neurodegeneration has already occurred, making early diagnosis difficult but crucial for managing the disease effectively [7].

The diagnostic process for Parkinson's disease is both comprehensive and invasive, involving several high-tech and costly procedures. Patients typically undergo MRI brain scans to rule out other conditions with similar symptoms, as well as Dopamine

Transporter (DaT) Scans that assess dopamine levels in the brain, crucial for diagnosing Parkinson's due to its association with dopamine-producing neuron degeneration [8]. These diagnostic tools, while effective, are also expensive. MRI scans can cost several thousand dollars, and DaT Scans are similarly priced, often adding financial strain to the emotional and physical toll on patients. Additionally, other assessments such as motor function tests and physical exams further evaluate symptoms like tremors and stiffness but also contribute to the overall cost and duration of the diagnostic process. Similarly, another tool for medical imaging (particularly for Parkinson's disease) is the PET scan, otherwise known as the Positron Emission Tomography scan. This process involves a radioactive drug that internally detects disease in patients which has proven to be extremely invasive and costly over time [9].

Despite the availability of advanced diagnostic tools like PET and MRI scans, their routine use in diagnosing Parkinson's disease (PD) is limited due to high costs and accessibility issues. Diagnosing PD continues to be challenging as it presents with a wide array of symptoms that overlap with other neurological disorders and progresses at varying rates among individuals. Additionally, early non-motor symptoms such as loss of smell, constipation, and sleep disturbances are frequently overlooked since they can be easily attributed to aging or other common health issues.

In this study, we explore the potential of automated machine learning techniques to predict PD based on handwriting analysis. Utilizing a specialized dataset obtained from the Department of Neurology at Cerrahpasa Faculty of Medicine, Istanbul University, this dataset includes handwriting samples from 62 people with Parkinson's and 15 healthy individuals [10]. By integrating these tests with machine learning analysis, we aim to develop a more accessible and cost-effective method for early PD detection, potentially improving diagnostic accuracy and patient outcomes.

II. Related Work

Recent advances in computer vision and deep learning have significantly revolutionized medical imaging, particularly enhancing the accuracy and efficiency of diagnostics [11]. These advancements are especially notable in the multifaceted approaches to diagnosing Parkinson's disease (PD), which leverage symptomatic data across various modalities including speech, handwriting, gait, EEG, and other physiological markers for early detection and precise diagnostics [11]–[13].

One notable approach employs multimodal assessments to diagnose Parkinson's disease using deep learning techniques that analyze movement and speech patterns. Similar to the approach in this study, these works utilize CNNs akin to ResNet50 and MobileNetV2, reinforcing the potential of advanced computational techniques in PD diagnosis [11]. This aligns with the broader goals of the present research, which explores automated methods for PD detection through artistic outputs.

Further research has leveraged deep learning to analyze the dynamics of handwriting in early Parkinson's detection, capturing subtle motor impairments indicative of the disease. This methodology parallels the objectives of this study, which focuses on extracting diagnostic insights from artistic expressions. The application of deep learning in these studies underscores its efficacy in deriving accurate diagnostic conclusions through detailed computational analysis [12].

Additionally, efforts have been directed towards quantifying the severity of Parkinson's disease through deep learning models. These projects demonstrate the adaptability of deep learning in evaluating PD progression and severity, contributing significantly to ongoing research in disease management [14].

Other initiatives have explored the use of deep learning to interpret EEG signals for PD diagnosis, showcasing the broad applicability of this technology across various data modalities. Although focused on different types of data, these studies highlight the potential of high-level computational techniques to significantly enhance PD diagnosis, complementing the objectives of this study, which applies deep learning to analyze artistic outputs for similar purposes [13].

Historically, PD diagnosis has relied on expensive and inconvenient methods, posing signifi-

cant challenges, particularly in developing regions. However, recent advancements in machine learning algorithms, such as Logistic Regression, Support Vector Classification, K-Nearest Neighbor, and Random Forest Classifier, have achieved impressive classification accuracy rates of nearly 91 percent, offering promising alternatives to traditional diagnostic approaches. Spiral drawing—a commonly used diagnostic tool that often deteriorates early in PD—holds significant potential as a marker for disease progression. Quantitative analysis of these drawings can highlight specific motor impairments characteristic of PD, providing valuable insights into disease severity and progression [15].

Further studies have shown that spiral drawings can serve as effective pre-clinical markers for PD. Quantitative analysis of these drawings highlights specific motor abnormalities, supporting their use as viable biomarkers for early PD detection [16].

A notable study using an automated deep learning approach to detect Parkinson's disease utilized the hand-drawn spirals and waves dataset from 2004, which consisted of sketches from 55 participants [17], [18]. This dataset comprised pencil sketches on white paper backgrounds. Transfer learning with ResNet50 models was used to distinguish between healthy and Parkinson's cohorts, resulting in a performance of 96.67

Additionally, Muhammed Erdem Isenkul and colleagues advanced PD monitoring through the use of a digitized graphics tablet that collects comprehensive motor function data via handwriting tests. Their study introduced the Dynamic Spiral Test (DST), enhancing the conventional Static Spiral Test (SST), to measure the impact of PD on handwriting capabilities. Conducted among PD patients and healthy controls, their findings illustrated that computerized tests can offer accurate, non-invasive diagnostic support and simplify the monitoring process. Such innovations could potentially replace traditional, more laborious diagnostic methods, offering significant benefits in environments where regular clinical visits pose logistical challenges [19]. This study leverages that innovative dataset to further validate the effectiveness of automated PD detection systems.

Specifically, the dataset [19] used in this research contains the coordinates of drawings created by patients, which were subsequently rendered into 2D plots for training deep learning models. This methodology aligns with the approach described by [18], but introduces distinct variations in both the dataset used and the data modality. Unlike previous work that utilized scanned images of pencil drawings, this process involved converting coordinate data directly into 2D images. Furthermore, this study explores the efficacy of using the energy-efficient MobileNetV2 architecture to assess the performance of lighter models in accurately predicting Parkinson's disease.

III. Methods and Materials

In this section, the research methodology implemented to conduct the experiments is described. This section begins with an overview of the dataset used, followed by a description of the experimental setup.

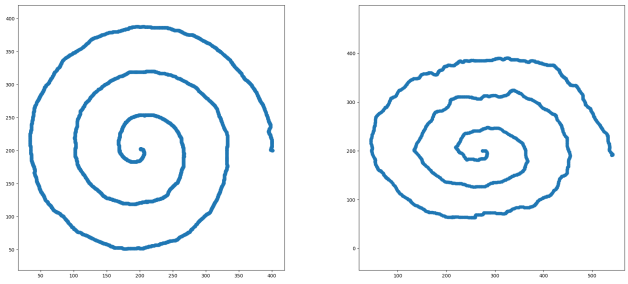
A. Dataset

The dataset used in this study was collected by Muhammed Erdem Isenkula and colleagues [19] in 2009. It comprises handwriting data recorded using the Wacom Cintiq 12WX, a device that combines a graphics tablet with an LCD monitor. This technology enables the display of a computer's screen while simultaneously capturing interactions with a digitized pen. The primary objective of this setup was to assess motor coordination in individuals, particularly Parkinson's disease (PD) patients, through handwriting tasks.

The recorded data consists of several parameters: X, Y, and Z coordinates, pressure levels, grip angle, timestamp, and test ID, all stored in comma-separated values (CSV) format. The X and Y coordinates were used to generate spiral drawings, as shown in Figure 1, which served as the primary input for the convolutional neural network (CNN) models. The dataset was partitioned into three subsets: 70% for training, 14% for testing, and 16% for validation.

The dataset used in this study, titled *"Improved Spiral Test Using Digitized Graphics Tablet for*

Monitoring Parkinson's Disease", is publicly available and can be accessed at: UCI Machine Learning Repository.



(a) Healthy (b) Parkinson's

Fig. 1: Spiral drawings samples from healthy and Parkinson's diagnosed patients

B. Methods

This study employs two pre-trained convolutional neural network architectures—MobileNetV2 and ResNet50—to evaluate their effectiveness in detecting Parkinson's disease from spiral drawing images. MobileNetV2 was selected for its computational efficiency and lightweight structure, making it suitable for deployment on devices with limited hardware. ResNet50 was chosen for its established high performance across various image classification tasks.

Both architectures were adapted by appending a dense layer of 100 neurons with a ReLU activation function, followed by a softmax output layer to classify between healthy individuals and Parkinson's patients. Training involved experimenting with epochs ranging from 10 to 50 and learning rates between 0.00001 and 0.005. Hyperparameters were tuned based on validation accuracy, and the best-performing model for each architecture was selected.

Finalized models were evaluated on a separate test subset using performance metrics such as accuracy, precision, recall, and F1-score. These metrics offer a comprehensive evaluation of the models' diagnostic capabilities.

This methodological approach ensures a fair evaluation of MobileNetV2 and ResNet50 as potential diagnostic tools in clinical settings, specifically for detecting Parkinson's disease through spiral drawing analysis.

IV. Results

In this section, we detail the results of our empirical evaluation, where we compare the performance of two distinct algorithms, MobileNetV2 and ResNet50, in the context of predicting Parkinson's disease from spiral drawings. The primary objective of this study is to determine whether MobileNetV2, a computationally lighter model, can match or exceed the performance of the more complex ResNet50 model in this specific application.

The validation results of the MobileNetV2 and ResNet50 hyperparameter tuning during the training phase are shown in in Figure 2 and Figure 3, respectively.

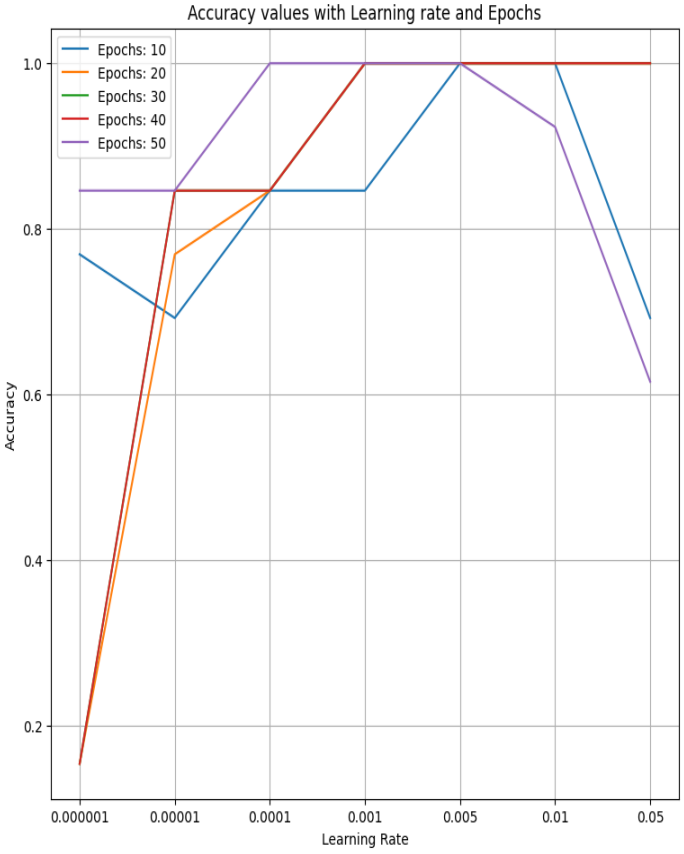


Fig. 2: MobileNetV2 Validation Results

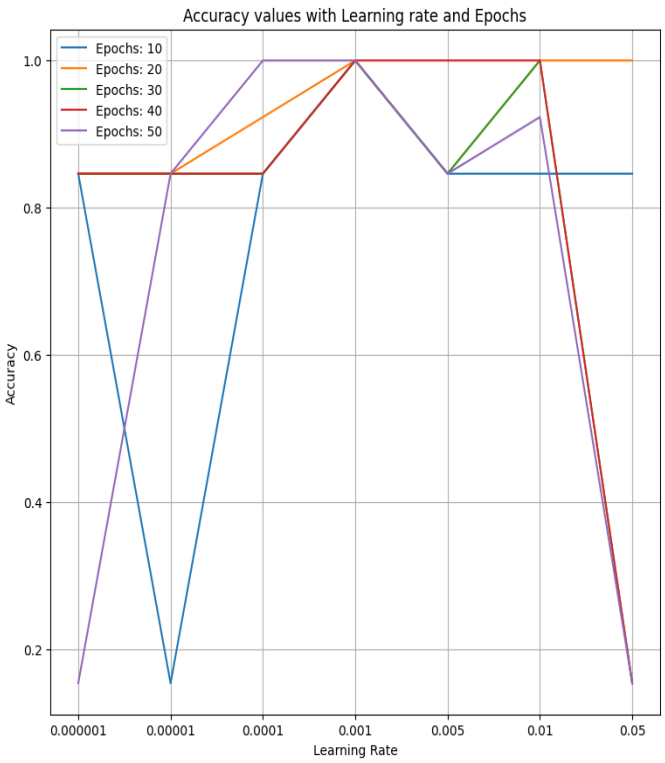


Fig. 3: ResNet50 Validation Results

Both models demonstrate high efficiency on the task, with MobileNetV2 showing slightly more sensitivity to initial learning rates, but quickly achieving high performance as training progresses. In contrast, ResNet50 displays more stability in different learning rates, though it is not immune to performance drops at higher rates and longer training times.

From these results, it is evident that MobileNetV2, despite being a lighter model, can match and occasionally exceed the performance of the more complex ResNet50 in predicting Parkinson's disease from spiral drawings. This outcome suggests that MobileNetV2 is not only computationally efficient, but also highly effective for tasks requiring high-precision pattern recognition, making it an attractive option for real-world applications where computational resources are limited.

The MobileNetV2 model that was trained with a learning rate of 0.001 and 20 epochs was chosen as the best model. In the case of ResNet50, the model with a learning rate of 0.001 and 10 epochs was chosen as the best model. In the test phase, MobileNetV2 achieved an accuracy of 91%, with a weighted average precision of 0.92 and a recall of 0.91. These results underscore the model's robustness and its ability to generalize well from training to unseen real-world data, affirming the validation findings. Conversely, ResNet50 demonstrated an accuracy of 82% in the test phase, with a lower average weighted precision of 0.67 and a recall of 0.82. The disparity between the validation performance and test results highlights potential overfitting issues or a lack of generalizability to new data, which was not as evident in the controlled validation environment.

The test results, when compared with the validation accuracy, suggest that MobileNetV2 not only outperforms ResNet50 in terms of computational efficiency but also in overall predictive accuracy and reliability when applied to new, unseen data, as shown in Table I. The superior performance of MobileNetV2 in both precision and recall metrics indicates a more balanced model that is better suited for practical applications in the medical diagnostic field.

The experimental findings support our hypothesis that a lighter model like MobileNetV2 can perform at par with, and sometimes better than, a more computationally intensive model such as ResNet50 in specific AI applications. This study underscores the potential of employing lightweight models in medical diagnostic applications, particularly in environments where computational efficiency is important. The discrepancies observed between validation and test performances for ResNet50 also prompt a deeper investigation into model training techniques to enhance generalizability and prevent overfitting.

TABLE I: Weighted Average Precision, Recall and F1 Score

	Precision	**Recall**	**F-1 Score**
MobileNetV2	0.92	0.91	0.90
ResNet50	0.67	0.82	0.74

V. Discussion

This section delves into the comparative analysis and implications of the experimental results obtained from testing MobileNetV2 and ResNet50

in the detection of Parkinson's disease using spiral drawings. The core aim was to evaluate whether the computationally lighter MobileNetV2 could achieve comparable or superior performance to the more complex ResNet50. The validation performance of the two model architectures provides insights into the model behaviour for this dataset.

The performance of MobileNetV2 varied significantly with changes in the learning rate. At very low learning rates (0.000001 and 0.00001), the model showed poor performance during the initial epochs but improved substantially as the number of epochs increased. Specifically, the accuracy increased from 0.7692 to 0.8462 at a learning rate of 0.00001 over 50 epochs, illustrating the model's capacity to adapt and learn progressively over time, albeit slowly. The model achieved optimal performance quite rapidly at a learning rate of 0.001. By the 20th epoch, the accuracy had reached 1.0, maintaining this level consistently through the 50th epoch. This suggests that a learning rate of 0.001 is effective for quick convergence in tasks involving the detection of Parkinson's disease from spiral drawings. At higher learning rates (0.01 and 0.05), although the model quickly reached high accuracy, it showed signs of instability in later epochs. For instance, while the learning rate of 0.01 yielded perfect accuracy from the 10th to the 30th epoch, there was a slight decline to 0.9231 by the 50th epoch. At the highest tested learning rate (0.05), the model's performance was erratic, peaking at 1.0 at the 20th and 30th epochs but then dropping to 0.6154 by the 50th epoch. This fluctuation suggests that while high learning rates can accelerate the initial learning process, they may adversely affect the model's stability over longer training periods. Across all learning rates, extending the number of training epochs generally resulted in improved performance or stabilization of accuracy.

The performance of ResNet50 at the lowest learning rates (0.000001 and 0.00001) showed generally stable performance across the epochs, maintaining an accuracy around 0.8462, with a notable exception at 50 epochs where accuracy significantly dropped to 0.1538. This indicates that while low learning rates ensure model stability up to a point, they might contribute to underfitting, particularly over extended training periods. At moderate to high learning rates (0.0001 to 0.01), the model generally performed better, reaching optimal accuracy (1.0) at a learning rate of 0.001 from the 10^{th} epoch onward. Notably, at a learning rate of 0.01, the model achieved perfect accuracy by the 20^{th} epoch and maintained this performance through the 40th epoch. The highest learning rate tested (0.05) showed a decline in model performance over time, with accuracy dropping significantly after the 20th epoch, suggesting that while higher learning rates can initially accelerate learning, they may lead to instability or overfitting as training progresses. The fluctuating accuracies, particularly noted at extreme learning rates, indicate potential issues with model reliability and generalizability.

Both MobileNetV2 and ResNet50 models displayed sensitivities to varying learning rates during hyper-parameter tuning, showing a pattern in which they performed optimally at certain rates but exhibited signs of instability or underfitting at others. MobileNetV2 achieved high accuracy swiftly at a moderate learning rate of 0.001, maintaining perfect accuracy from the 20th epoch onward. It also demonstrated the ability to progressively learn at very low learning rates, albeit at a slower pace. In contrast, ResNet50 reached optimal performance earlier, starting from the 10th epoch at the same learning rate of 0.001, which underscores that both models are capable of rapid convergence when tuned correctly.

However, at the extremities—both the lowest and highest learning rates—the models encountered challenges. MobileNetV2 showed instability in the later epochs at high rates, while ResNet50 experienced a significant performance drop at the 50th epoch at the lowest rates, suggesting underfitting. Furthermore, while both models achieved high accuracy, MobileNetV2 exhibited a decline in later epochs at elevated learning rates, indicating a nuanced difference in handling extreme training conditions.

These observations highlight that although both architectures are capable of achieving high performance, their responses to hyper-parameter adjustments vary, impacting their stability and reliability over extended training periods. This differential behavior suggests that careful calibration of learning

rates is crucial for optimizing each model's performance.

In testing, MobileNetV2 recorded an accuracy of 91%, with a weighted average precision of 0.92 and a recall of 0.91, emphasizing the model's robustness and its capacity to generalize effectively from training to unseen real-world data. In contrast, ResNet50 demonstrated an accuracy of 82% during the test phase, with a lower average weighted precision of 0.67 and a recall of 0.82. The notable disparity between validation performance and test results for ResNet50 suggests issues with generalizability to new data, which were less apparent during the controlled validation phase. This divergence in test performance further illustrates the critical importance of rigorous testing under varied conditions.

Ultimately, the results from both models contribute to a broader understanding of how different deep learning architectures can behave differently for the same hyper-parameter values. The results also emphasize the need for ongoing research to refine these models, enhance their generalizability, and ensure they can be deployed reliably in real-world diagnostic settings.

VI. Conclusion

In conclusion, this study has demonstrated that convolutional neural networks, particularly the MobileNetV2 model, can provide a viable, non-invasive alternative for the early detection of Parkinson's disease using spiral drawing analysis. The experiments conducted reveal that MobileNetV2, with its lighter architecture, not only achieves high accuracy but also outperforms the more complex ResNet50 model in specific test scenarios. These findings are significant as they suggest that the efficiency and accuracy of MobileNetV2 can potentially reduce the economic and physical burdens associated with traditional, more invasive diagnostic methods.

Furthermore, the sensitivity of both models to different learning rates highlights the importance of careful hyper-parameter tuning in achieving optimal performance. While MobileNetV2 generally showed greater stability and higher performance at moderate learning rates, ResNet50 required precise adjustments to avoid underfitting or overfitting, underscoring the different approaches needed for different architectures.

By integrating such AI-driven tools, healthcare providers can achieve earlier and more accurate diagnoses, which is crucial for the effective management and treatment of Parkinson's disease. Moving forward, further studies are recommended to explore the integration of these models into real-world clinical settings, assessing their performance across broader datasets and diverse patient demographics to validate their practical utility and robustness.

Code Availability

The code used in this study will be made available upon request.

References

[1] J. W. Han, Y. D. Ahn, W.-S. Kim, C. M. Shin, S. J. Jeong, Y. S. Song, Y. J. Bae, and J.-M. Kim, "Psychiatric manifestation in patients with parkinson's disease," *Journal of Korean medical science*, vol. 33, no. 47, 2018.

[2] L. V. Kalia and A. E. Lang, "Parkinson's disease," *The Lancet*, vol. 386, no. 9996, pp. 896–912, 2015.

[3] "Parkinson's disease: Causes, symptoms, and treatments." https://www.nia.nih.gov/health/parkinsons-disease/parkinsons-disease-causes-symptoms-and-treatments. Accessed: 2024-5-1.

[4] "Statistics." https://www.parkinson.org/understanding-parkinsons/statistics. Accessed: 2024-5-1.

[5] G. DeMaagd and A. Philip, "Parkinson's disease and its management: Part 1: Disease entity, risk factors, pathophysiology, clinical presentation, and diagnosis," *P T*, vol. 40, pp. 504–532, Aug. 2015.

[6] A. Melão, "Parkinson's study predicts more than 1 million cases in US by 2030." https://parkinsonsnewstoday.com/news/1-2-million-cases-parkinsons-predicted-2030-united-states-study/, July 2018. Accessed: 2024-5-1.

[7] 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, no. 6, pp. 566–576, 2016.

[8] "Statistics — parkinson.org." www.parkinson.org/understanding-parkinsons/statistics#:~:text=Incidencepercent20ofpercent20Parkinson. [Accessed 08-05-2024].

[9] "Positron emission tomography scan - Mayo Clinic — mayoclinic.org." https://www.mayoclinic.org/tests-procedures/pet-scan/about/pac-20385078#:~:text=The%20PET%20scan%20uses%20a,magnetic%20resonance%20imaging%20(MRI). [Accessed 08-05-2024].

[10] M. Isenkul and B. Sakar, "Parkinson Disease Spiral Drawings Using Digitized Graphics Tablet." UCI Machine Learning Repository, 2017. DOI: https://doi.org/10.24432/C5Q01S.

[11] J. C. Vásquez-Correa, T. Arias-Vergara, J. R. Orozco-Arroyave, B. Eskofier, J. Klucken, and E. Nöth, "Multimodal assessment of parkinson's disease: a deep learning approach," *IEEE journal of biomedical and health informatics*, vol. 23, no. 4, pp. 1618–1630, 2018.

[12] C. R. Pereira, S. A. Weber, C. Hook, G. H. Rosa, and J. P. Papa, "Deep learning-aided parkinson's disease diagnosis from handwritten dynamics," in *2016 29th SIBGRAPI conference on graphics, patterns and images (SIBGRAPI)*, pp. 340–346, Ieee, 2016.

[13] S. L. Oh, Y. Hagiwara, U. Raghavendra, R. Yuvaraj, N. Arunkumar, M. Murugappan, and U. R. Acharya, "A deep learning approach for parkinson's disease diagnosis from eeg signals," *Neural Computing and Applications*, vol. 32, pp. 10927–10933, 2020.

[14] S. Grover, S. Bhartia, A. Yadav, K. Seeja, *et al.*, "Predicting severity of parkinson's disease using deep learning," *Procedia computer science*, vol. 132, pp. 1788–1794, 2018.

[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] M. San Luciano, C. Wang, R. A. Ortega, Q. Yu, S. Boschung, J. Soto-Valencia, S. B. Bressman, R. B. Lipton, S. Pullman, and R. Saunders-Pullman, "Digitized spiral drawing: A possible biomarker for early parkinson's disease," *PloS one*, vol. 11, no. 10, p. e0162799, 2016.

[17] Y. El Ghzizal, N. Aharrane, G. Khaissidi, and M. Mrabti, "Transfer learning and pressure effect for handwriting to early detection of parkinson's disease," in *International Conference on Digital Technologies and Applications*, pp. 460–469, Springer, 2022.

[18] N. Jahan, A. Nesa, and M. A. Layek, "Parkinson's disease detection using resnet50 with transfer learning," *International Journal of Computer Vision and Signal Processing*, vol. 11, no. 1, pp. 17–23, 2021.

[19] M. Isenkul, B. Sakar, O. Kursun, *et al.*, "Improved spiral test using digitized graphics tablet for monitoring parkinson's disease," in *The 2nd international conference on e-health and telemedicine (ICEHTM-2014)*, vol. 5, pp. 171–175, 2014.