
School of Mathematics and Statistics

# Analysis of Kinematic Features of Handwritten Spirals for Parkinson's Disease Classification

by

**Viyanka Moodley (300565283)**

*Supervisor:* Professor Alejandro Frery

*Contributors:* Amelia Bentley and Isabelle Southon

22 October 2024

# Contents

## Executive Summary

Parkinson's disease is a progressive neurological disorder characterised by motor dysfunction, including tremors, stiffness, and slowed movement. Early diagnosis is critical for managing the disease, but current diagnostic methods are costly, time-consuming, and rely on access to specialised neurologists. With a global shortage of neurologists and increasing demand for neurological care, particularly in regions like Aotearoa New Zealand, there is a pressing need for alternative, cost-effective diagnostic tools. Handwriting analysis, particularly the evaluation of spiral drawings, offers a promising non-invasive approach to identifying motor dysfunction in Parkinson's disease.

This study investigates how features derived from handwritten spiral drawings can best distinguish between healthy individuals and Parkinson's patients using machine learning models. The dataset includes spiral drawings from 32 participants, with 2 spirals from each healthy control subject (10 subjects) and 1 spiral from each Parkinson's patient (22 patients). Key features such as X and Y coordinates, pressure, and pen orientation were captured, and additional metrics like mean pressure, pen-down time, and total distance drawn were engineered for further analysis.

Three machine learning models, *Random Forest, Gradient Boosting, and Support Vector Machine (SVM)* were employed for classification. The Random Forest model performed best, achieving an accuracy of 84.6% and an AUC-ROC score of 95.2%, demonstrating its strong capability to differentiate between the two groups. Important predictive features across the models included total distance, mean altitude, pen down time, mean acceleration, and standard deviation of pressure, which reflect key aspects of motor control and movement consistency.

These findings suggest that analysing motor function through spiral drawings, particularly focusing on features such as movement distance and pressure variability, can offer a reliable, non-invasive diagnostic tool for Parkinson's disease. This approach could be especially valuable in areas with limited access to neurologists, providing a scalable and affordable method for early diagnosis. Future applications could integrate this tool into primary care or telemedicine platforms, improving access to Parkinson's screening and potentially advancing patient outcomes.

## 1 Background

Parkinson's disease is a neurodegenerative disorder that impairs motor function, causing tremors, stiffness, and slowed movement. Early diagnosis is critical for managing the disease, yet there are no definitive lab tests or imaging methods to detect Parkinson's, with diagnosis relying heavily on clinical observation (Parkinsons UK, 2022)[11]. This process is further complicated by a shortage of neurologists, particularly in Aotearoa New Zealand, where only 47.8 full-time equivalent (FTE) neurologists are available, well below the necessary 74. This shortage leads to long wait times and delayed diagnoses for neurological conditions like Parkinson's (Neurological Alliance, 2024)[8].

Spiral drawing tasks have shown promise as a non-invasive diagnostic tool for Parkin-

son's. Studies indicate that features derived from spirals—such as pressure consistency and movement patterns—can distinguish between Parkinson's patients and healthy individuals [12]. These insights have paved the way for using machine learning to automate and enhance diagnostic accuracy.

Initially, this project aimed to use regression models to identify effective spiral features, but after exploratory analysis, it became clear that classification models were more appropriate due to the binary nature of the target variable (healthy vs. Parkinson's). The revised research question now focuses on: *What types of features derived from handwritten spiral drawings can best distinguish between healthy individuals and patients with Parkinson's disease in a classification model?*

This analysis derives new features from time-series spiral data, including *mean pressure*, *pen-up time*, *pen-down time*, *total distance drawn*, and motion metrics like *velocity* and *acceleration*. These features are investigated for their ability to accurately classify healthy individuals and those with Parkinson's disease.

Identifying the most predictive features could lead to the development of affordable, non-invasive diagnostic tools, helping to reduce reliance on neurologists and improving early diagnosis and treatment, especially in regions with healthcare shortages. This research contributes to ongoing efforts to enhance diagnostic methods for Parkinson's disease through spiral analysis.

## 2 Data Description

The original dataset consisted of time-series data collected from spiral drawings made by 33 participants, including 10 healthy control subjects and 23 individuals with Parkinson's disease. Each participant's drawing was recorded across eight variables, including X and Y coordinates, pressure, and orientation of the pen. In total, the original dataset contained 121,299 observations and 8 columns.

### Data Adjustments and Processing

During the exploratory data analysis (Phase 1), it was observed that each control subject had two spiral drawings, whereas most Parkinson's patients had only one. However, one Parkinson's patient, T005, also had two spiral drawings. Upon closer inspection, the coordinates for T005 showed abnormally large values compared to the rest of the dataset, leading to the exclusion of T005's data from the analysis. This exclusion was justified by the fact that T005 was the only Parkinson's patient with two spirals, creating a structural imbalance in the dataset.

Additionally, to address the class imbalance, the two spirals for each control subject were split into separate observations using a threshold of 5000 on the Y-coordinate, which provided a natural point of division between the two spirals. This process ensured each spiral was treated as an independent observation. Originally, the dataset had 10 spiral drawings for the control group and 23 for the Parkinson's group, but after splitting, the control group had 20 spirals. With the removal of T005, the final dataset contained 20

spirals from control subjects and 22 from Parkinson's patients.

The final dataset used for analysis consisted of 42 spirals (20 from control subjects and 22 from Parkinson's patients) and 16 columns, including both the original variables and newly engineered features derived from the raw data. These features captured important aspects of movement, pressure, and pen orientation, which are relevant for distinguishing between Parkinson's and healthy subjects.

### Engineered Features

The analysis focused on newly derived features that were created based on the original time-series data. These engineered features were designed to capture important movement dynamics and pen behaviour, which could be predictive of whether a participant had Parkinson's disease. The key engineered features included mean and variability of pressure, pen-up time, pen-down time, total distance drawn, and motion metrics like velocity, acceleration and jerk. The final dataset used for analysis consisted of 42 entries and 15 columns. To see a full list of the new features, visit Appendix A.

By splitting the control spirals, excluding outliers, and engineering new features, the dataset became more balanced and meaningful for classification analysis. This allowed for a more robust investigation and better interpretability of the differences in motor function between Parkinson's patients and healthy individuals.

## 3 Ethics, Privacy, and Security

### Ethical Considerations

This project involves the use of a dataset containing spiral drawings from individuals with Parkinson's disease and healthy control subjects. The ethical considerations for such data usage are significant due to the medical nature of the information. The data must be handled in a way that ensures integrity, accuracy, and fairness in the analysis, especially given that the findings could impact how Parkinson's disease is diagnosed and understood.

First, it is essential that the data used was collected with informed consent, where participants were fully aware of the study's aims and their rights, including the right to withdraw from the study at any time without consequences. In accordance with ethical research practices, the data has been used for academic and research purposes, as authorised by the study's supervisor, ensuring it remains within the scope of the original consent.

A core ethical concern is to avoid introducing bias into the analysis, which could lead to unfair or inaccurate conclusions. For example, we must ensure that the methods used to distinguish between Parkinson's patients and healthy individuals are based on sound statistical reasoning and do not disproportionately misclassify certain groups. Additionally, it is crucial to avoid perpetuating stereotypes or stigmatising individuals with Parkinson's

disease in any way. Researchers must be transparent about the limitations of the dataset and cautious in how findings are presented to avoid potential misinterpretation or harm (Beauchamp & Childress, 2001) [1].

Finally, the public dissemination of results should be subject to ethical review, and any errors found during the research must be corrected promptly to prevent misinformation. Prior to releasing any findings, ethical oversight will ensure compliance with standards in medical research.

## Privacy Concerns

Privacy concerns are central to this project because the dataset involves sensitive medical information, specifically whether individuals have Parkinson's disease. Even though the dataset is anonymised, without direct identifiers such as names or addresses, there remains a risk of re-identification through indirect identifiers. Time series data like timestamps, pressure patterns, and unique drawing characteristics could, in theory, be cross referenced with other datasets to identify participants, compromising their privacy.

In accordance with New Zealand's Privacy Act 2020, strict measures have been implemented to safeguard the privacy of participants. The Privacy Act emphasises that personal data should only be used for the purposes for which it was collected and should be protected from unauthorised access (New Zealand Government, 2020)[10]. To mitigate the risk of re-identification, individual-level data is not shared, and results are presented in aggregated form. This ensures that the analysis focuses on overall trends rather than exposing any participant's personal health information. Additionally, features like patient IDs were replaced with anonymised identifiers, and findings are shared in generalised formats.

Given the sensitive nature of the medical data, unauthorised disclosure of a participant's disease status could have severe consequences, including stigma and discrimination. As such, care was taken to ensure that patient privacy is fully protected throughout the research process.

## Security Steps

The analysis for this project was conducted on personal devices and university library computers, utilising Python and R. While basic security protocols were followed, including password protection and biometric authentication on personal devices, as well as logging out of shared university computers after use, several additional security measures could have further enhanced data protection.

Conducting analysis on public or shared Wi-Fi networks, such as at home, the university, or public libraries, poses potential risks for data security. To mitigate these risks, data encryption could have been implemented to safeguard the dataset during storage and transmission. Encryption is crucial in ensuring that data, even if intercepted, remains unreadable without the appropriate decryption keys (New Zealand Government, 2020)[10]. Additionally, the use of a virtual private network (VPN) could have provided

an additional layer of security by encrypting all internet traffic, further reducing the likelihood of unauthorised access to data when using public or shared networks (Cisco, n.d.) [3].

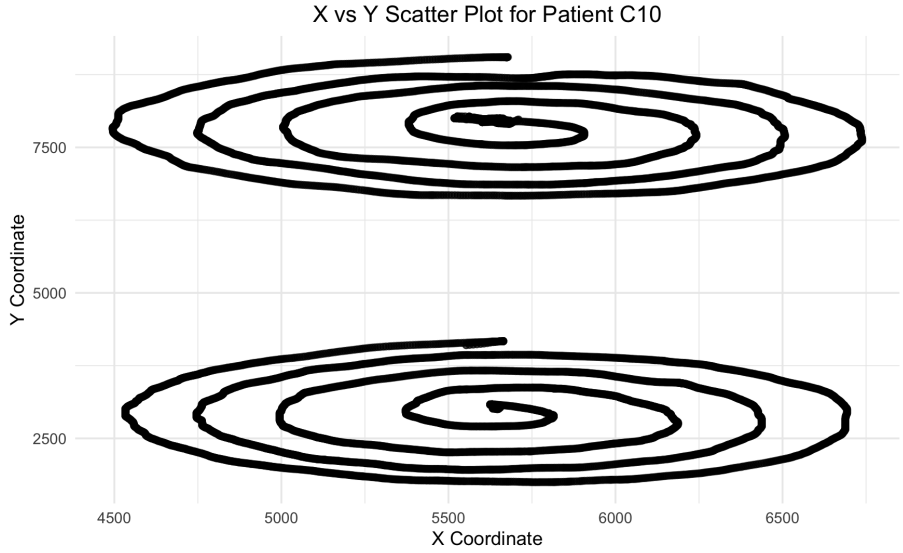
Furthermore, regular backups of the dataset should have been stored in secure, encrypted cloud environments, such as Google Cloud or AWS, to ensure data recovery in the event of device failure or accidental deletion. Implementing intrusion detection systems and real-time monitoring tools would have allowed for the detection and prevention of any unauthorised access attempts, helping to ensure the integrity of the data throughout the analysis process (Office of the Privacy Commissioner, 2020)[13]. At the completion of the project, securely deleting the dataset using approved data disposal methods would have guaranteed that no sensitive information could be retrieved or misused, protecting participant privacy even after the analysis was concluded (Office of the Privacy Commissioner, 2020)[13].

In future projects, adopting more advanced security protocols, such as encrypting all data, using secure private networks, and employing real-time monitoring systems, would offer greater protection for sensitive medical information throughout the research process.

# 4 Exploratory Data Analysis

The two plots of the X vs. Y coordinates for Patients C10 (control) and T008 (Parkinson's) reveal stark differences in motor function that are both visually evident and analytically significant. In the case of Patient C10, the spiral appears smooth and consistent,

Figure 1: Spiral Drawing done by a Healthy Participant




with no visible interruptions or irregularities. The loops of the spiral are well-defined and evenly spaced, indicative of steady and controlled hand movements. This consistency suggests that Patient C10, who is part of the control group, exhibits normal motor function without any noticeable signs of tremors or instability. The spiral maintains a near-perfect symmetry, which reflects the patient's ability to maintain a stable pen grip, direction,

and pressure throughout the drawing task.

On the other hand, Patient T008, who is diagnosed with Parkinson's disease, produces a much more erratic and inconsistent spiral. The plot reveals frequent deviations from the intended path, with jagged lines and irregular loop patterns. These irregularities are typical of the motor impairments caused by Parkinson's disease, particularly the tremors and involuntary movements that can significantly disrupt fine motor tasks. The loops in Patient T008's spiral are unevenly spaced, with areas of higher density where the pen appears to jitter uncontrollably, alongside gaps that suggest interruptions in the pen's contact with the surface. Moreover, the presence of outlier points on the far ends of the graph further suggests that the patient experienced moments of difficulty maintaining control, possibly lifting the pen unintentionally or deviating from the path due to tremors.

The control patient's smooth and regular spiral reflects a level of motor control that is disrupted in the Parkinson's patient, whose spiral shows erratic and involuntary hand movements. These visual patterns not only differentiate between healthy and affected individuals but also provide valuable insights into the severity of motor dysfunction. As such, spiral drawing tasks could serve as effective diagnostic tools, where the smoothness, regularity, and presence of interruptions in the X and Y coordinates could be quantified and analysed to distinguish between patients with Parkinson's disease and healthy individuals. The contrast between the two plots highlights the potential of such assessments in identifying early signs of motor impairment in clinical settings.

Figure 2: Spiral Drawing done by a Parkinson's patient

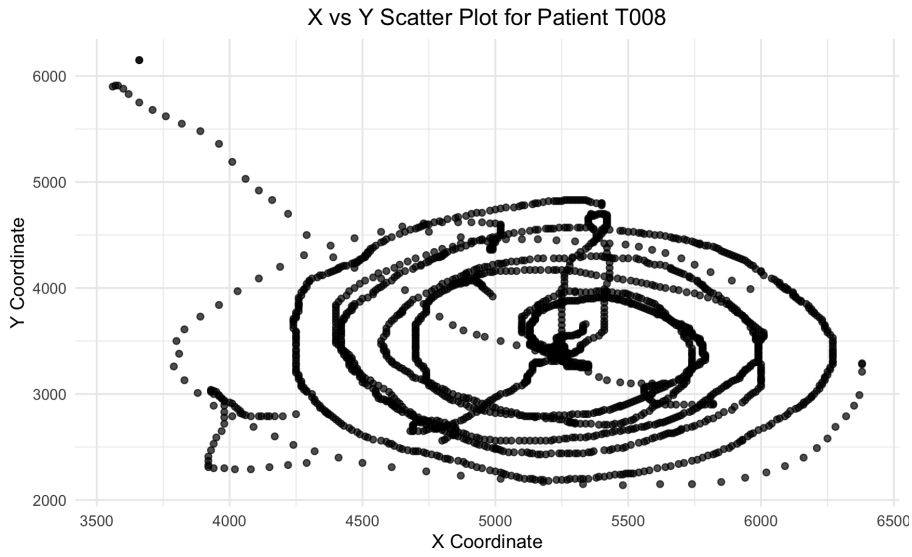



Figure 3 shows the proportions of pen states ("On" and "Off") for both Parkinson's patients and healthy individuals. The pen state indicates whether the pen was in contact with the drawing surface ("On") or lifted off the surface ("Off") during the drawing of spirals.

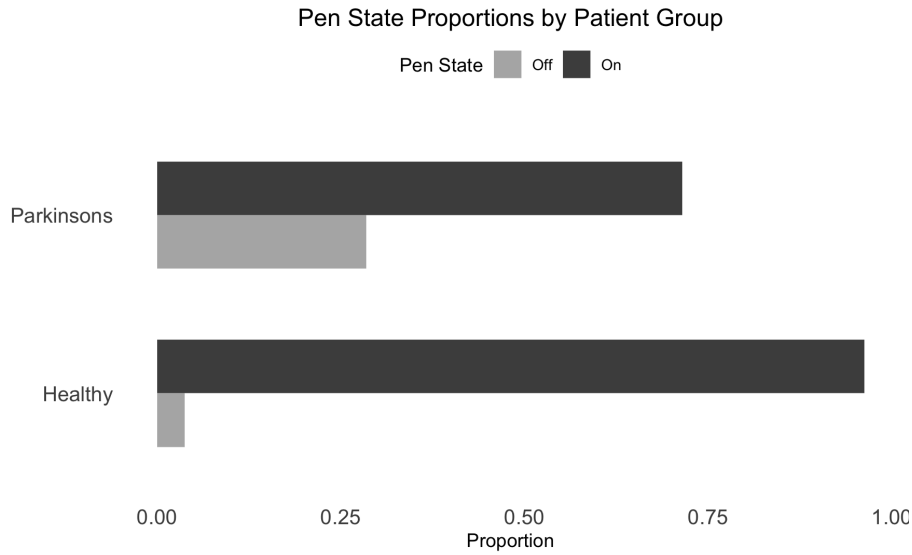
For Parkinson's patients, the plot reveals a notable amount of time when the pen is not in contact with the surface (pen "Off"), represented by the lighter gray bar. This is likely due to tremors, instability, or involuntary hand movements associated with Parkinson's

disease, which may cause the pen to be lifted unintentionally. However, a majority of the time is still spent with the pen in contact with the surface ("On"), as indicated by the darker bar.

In contrast, healthy individuals exhibit a much smaller proportion of time with the pen lifted off the surface ("Off"). The healthy group spends the vast majority of the time with the pen in contact with the drawing surface, as shown by the significantly longer dark gray bar. This suggests more stable and controlled hand movements during the spiral drawing task compared to those with Parkinson's.

This analysis reflects a clear distinction between the two groups in terms of pen contact consistency. The greater proportion of pen lifts among Parkinson's patients provides insight into their motor challenges, likely caused by involuntary movements or difficulty maintaining steady hand motion, which is a characteristic of the disease. This variable, representing the pen "On" and "Off" states, could be an important feature when analysing and distinguishing between Parkinson's patients and healthy controls. The plot suggests that healthy participants exhibit greater control and consistency during the task, whereas Parkinson's patients experience more frequent interruptions in their movements.

Figure 3: Pen State Proportions




## 5 Detailed Analysis Results

The objective of this analysis was to classify patients as either healthy or having Parkinson's disease based on features derived from their spiral drawings. For this purpose, I employed three machine learning classifiers: Random Forest, Gradient Boosting, and Support Vector Machine (SVM). These models were chosen for their ability to handle small datasets and capture non-linear relationships between variables, which are common in medical data [5]. Furthermore, all three models are robust against multicollinearity, which is a known concern when working with derived or engineered features.

### 5.1 Data Manipulation and Preprocessing

To make the data suitable for classification, I engineered several new features from the original time-series dataset, which included spatial (X, Y) coordinates, timestamps, pressure, azimuth, and altitude values recorded during the spiral drawing. These raw variables were transformed into more meaningful statistical and motion-related features, such as mean pressure, standard deviation of pressure, total distance, mean velocity, acceleration, and jerk, which have been found useful in similar handwriting analysis studies [12].

Given the differing scales of the features, standardisation was applied to the continuous variables. Standardising these features ensures that all variables contribute equally to the model, preventing those with larger numeric ranges (e.g., total distance) from dominating the learning process [6]. Furthermore, the target variable (*patient_group*) was encoded as a binary variable, with 0 representing healthy controls and 1 representing Parkinson's patients. This binary classification setup is required for the models used.

During feature engineering, missing and infinite values emerged, particularly in the calculation of velocity, acceleration, and jerk, due to boundary issues (e.g., undefined values for the first and last data points). To ensure the models could be trained without errors, I replaced missing and infinite values with zeros. This approach is standard in scenarios where edge effects cause missing data, particularly for features that involve rate of change [9]. While this method can minimise disruption to model training, it is important to recognise that imputing values in this manner might slightly bias the results, particularly if the zero-filled values are not representative of actual patient behaviour.

The data then dropped the "spiral_id" variable and was separated into X, which was the input variables and Y for the target feature "patient_group". The data was then split using a 70 to 30 train test split before carrying out any modelling techniques, using "random_state=301" for reproducibility.

### 5.2 Feature Selection and Model Fitting

Once the dataset was cleaned and prepared, I initially trained the models using all available features. For the tree-based models (Random Forest and Gradient Boosting), feature importance was calculated based on each feature's contribution to reducing impurity in decision splits. These models inherently rank features, allowing us to understand which variables are most important in distinguishing between the two patient groups. For SVM, I employed Recursive Feature Elimination (RFE), a wrapper method that iteratively removes the least important features based on model performance until only the most significant ones remain [4].

The top features across models consistently included *total distance* and *standard deviation of pressure*, which align with previous studies showing that motor control metrics such as movement distance and pressure variability are effective in differentiating patients with motor impairments [7]. The feature importances from Random Forest and Gradient Boosting are visualised in the figures 4 & 5 below. The results from SVM's RFE are displayed in Appendix A.

After identifying the most important features, I retrained each model using only the

Figure 4: Best Features According to Random Forest

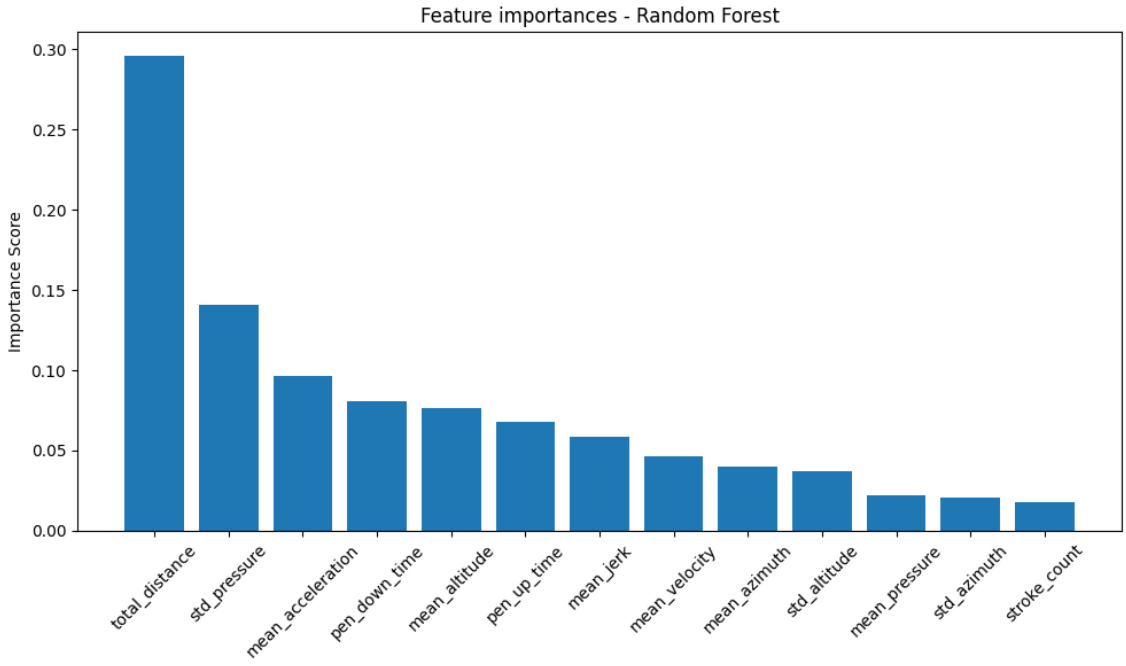
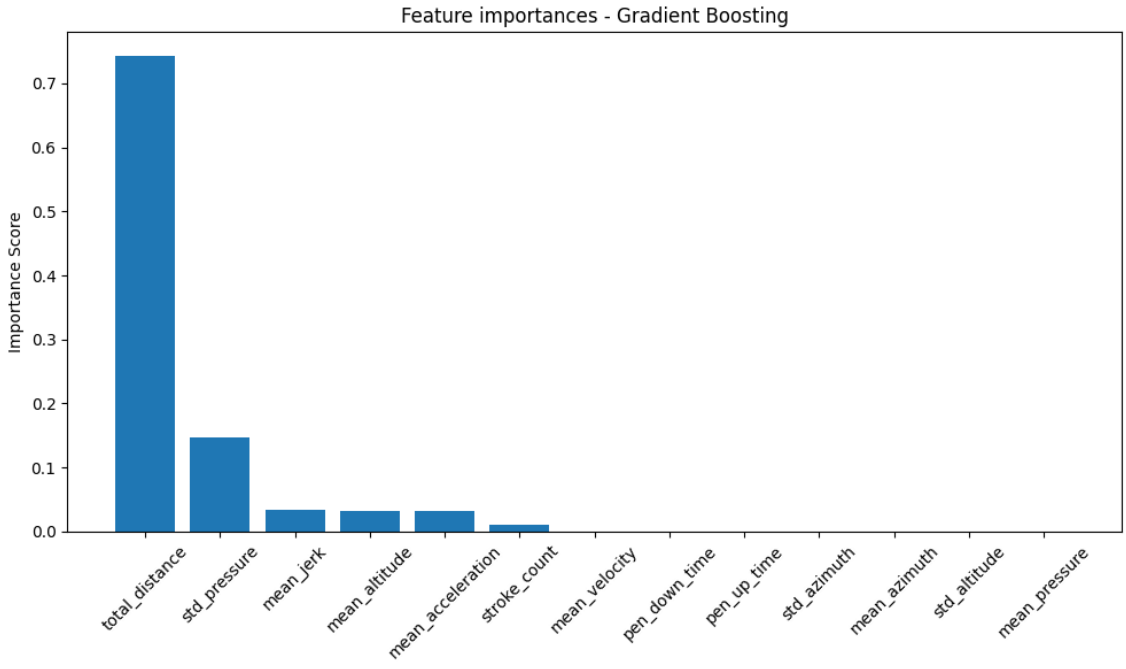



Figure 5: Best Features According to Gradient Boosting




top five features from each algorithm, improving model simplicity and reducing the risk of overfitting.

### 5.3 Baseline Model Performance

The initial models, trained with all features, provided baseline performance metrics, which are summarised below in Table 1.

Table 1: Baseline Model Performance

| Model | Accuracy | AUC-ROC | Precision | Recall | F1-Score |
|---|---|---|---|---|---|
| Random Forest | 0.77 | 0.88 | 0.75 | 0.86 | 0.80 |
| Gradient Boosting | 0.69 | 0.69 | 0.71 | 0.71 | 0.71 |
| SVM | 0.77 | 0.83 | 1.00 | 0.57 | 0.73 |

Random Forest and SVM showed comparable results, with Random Forest achieving a higher AUC-ROC, indicating a better ability to distinguish between the classes. Gradient Boosting, while showing lower overall performance, still demonstrated reasonable accuracy and recall, making it a viable alternative model.

### 5.4 Improved Model Performance with Feature Selection

Once feature selection was performed, the models were retrained using only the top-ranked features, leading to more efficient and interpretable models. The refined models provided the following performance metrics, which are summarised in Table 2. The feature-selected Random Forest model showed the most significant improvement, with an AUC-ROC of 0.95, suggesting a highly effective model for distinguishing between healthy and Parkinson's patients. This model effectively utilised the most important features while maintaining a balance between precision and recall. Gradient Boosting maintained its baseline performance, while SVM's results showed a trade-off between high precision and lower recall, indicating that it may be less effective at identifying all cases of Parkinson's disease but more reliable when it does.

Table 2: Improved Model Performance after Feature Selection

| Model | Accuracy | AUC-ROC | Precision | Recall | F1-Score |
|---|---|---|---|---|---|
| Random Forest | 0.85 | 0.95 | 0.86 | 0.86 | 0.86 |
| Gradient Boosting | 0.69 | 0.71 | 0.71 | 0.71 | 0.71 |
| SVM | 0.77 | 0.83 | 1.00 | 0.57 | 0.73 |

### 5.5 Uncertainty and Bias

There are several areas of uncertainty and potential bias in this analysis. First, the dataset is relatively small, with only 42 spiral drawings after augmentation. While Random Forest is robust to small datasets, the model's complexity may still lead to overfitting, especially when the number of samples per class is imbalanced [6]. The splitting of control subjects' spirals into two observations somewhat mitigated the class imbalance, but the dataset remains small overall, which could affect the generalisability of the models.

Moreover, the engineered features, such as mean jerk and acceleration, depend on the quality of the raw time-series data. Any noise or inaccuracies in the original measurements could propagate through to the derived features, potentially introducing bias into the models. Future analyses could benefit from more data and robust cross-validation techniques to further reduce bias and ensure the models' reliability.

# 6 Conclusions and Recommendations

The analysis demonstrated that several engineered features derived from handwritten spiral drawings, such as total distance, pen down time, mean altitude, mean acceleration, and standard deviation of pressure, were highly predictive in distinguishing between healthy individuals and patients with Parkinson's disease. These features, capturing key aspects of motor dysfunction, were consistently ranked as important across models like Random Forest, Gradient Boosting, and Support Vector Machines (SVM). The Random Forest classifier achieved the highest accuracy of 84.6% and an AUC-ROC of 95.2%, emphasising the utility of these features in effectively classifying the two groups. This suggests that such motion-related variables provide significant insights into motor control difficulties experienced by Parkinson's patients.

However, the study had some limitations that should be acknowledged. Firstly, the relatively small sample size (42 spirals) may limit the generalisability of the findings, and the dataset's class imbalance (initially 10 control spirals and 23 Parkinson's spirals) required splitting the control spirals to partially address the issue. Despite this approach, the small sample size still poses a challenge for building highly generalisable models, and the risk of overfitting remains a concern, particularly for complex models like Random Forest. Secondly, the handling of missing and edge-case data points by filling them with zeros may have introduced noise into the analysis. Although this method enabled the completion of model training, it is not always an accurate reflection of real patient behaviour, especially for critical features like acceleration and jerk.

Future directions for this research should focus on expanding the dataset, both in terms of the number of participants and the diversity of the population included, to improve the robustness and generalisability of the classification models. Additionally, using more sophisticated imputation techniques to address missing data, such as multiple imputation or using time-series-specific methods, could enhance the accuracy of the engineered features. Exploring advanced time-series models, such as Long Short-Term Memory (LSTM) networks, could also capture the temporal dependencies and dynamics in spiral drawing tasks more effectively. Another potential direction is validating these models with a broader and more clinically diverse population to ensure their effectiveness in real-world clinical settings.

Lastly, collaboration with clinical experts in neurology could provide valuable insights for refining feature engineering and identifying additional metrics that may further improve diagnostic accuracy. Integrating such models into telemedicine platforms or primary healthcare settings could make Parkinson's diagnosis more accessible, particularly in regions facing a shortage of neurologists. This would represent a significant step toward developing cost-effective, non-invasive diagnostic tools for early detection of Parkinson's disease.

## References


[1] Beauchamp, T. L., & Childress, J. F. (2001). *Principles of biomedical ethics (5th ed.)*. Oxford University Press.

[2] Bender, K. (2022, February 17). The "Grave Threat" Posed by the Shortage of Neurologists. *NeurologyLive*. Retrieved from https://www.neurologylive.com/view/the-grave-threat-posed-by-the-shortage-of-neurologists

[3] Cisco. (n.d.). *What is a firewall?*. Cisco. Retrieved from: https://www.cisco.com/c/en_us/products/security/firewalls/what-is-a-firewall.html

[4] Guyon, I.M., Weston, J., Barnhill, S.D., & Vapnik, V.N. (2002). *Gene Selection for Cancer Classification using Support Vector Machines*. Machine Learning, 46, 389-422. Retrieved from: https://link.springer.com/article/10.1023/A:1012487302797

[5] Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer.

[6] James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning: With Applications in R*. Springer.

[7] Legrand, A. P., Rivals, I., Richard, A., Apartis, E., Roze, E., Vidailhet, M., & Hainque, E. (2017). New insight in spiral drawing analysis methods – Application to action tremor quantification. *Clinical Neurophysiology*, 128(10), 1823-1834.

[8] Neurological Alliance of New Zealand. (2024, February). *Briefing for the Incoming Minister of Health*. Retrieved from: https://www.neurologicalalliance.org.nz/neurologicAlliance/pdf/BIM-Final-v2_3-Feb_2024.pdf

[9] Nielsen, F. (2016). *Introduction to machine learning*. MIT Press.

[10] New Zealand Government. (2020).*Privacy Act 2020*. Retrieved from: https://www.legislation.govt.nz/act/public/2020/0031/latest/whole.html

[11] Parkinson's UK. (2022). How is Parkinson's diagnosed? Retrieved from https://www.parkinsons.org.uk/information-and-support/how-parkinsons-diagnosed

[12] Pullman, S. L. (1998). Spiral analysis: A new technique for measuring tremor with a digitizing tablet. *Movement Disorders*, 13(S3), 85-89. https://doi.org/10.1002/mds.870131315

[13] Office of the Privacy Commissioner. (2020).*Guide to secure data disposal*. Retrieved from: https://www.privacy.org.nz/assets/Uploads/Data-disposal-guidance.pdf

[14] Wang, H., Yu, Q., Kurtis, M. M., Floyd, A. G., Smith, W. A., & Pullman, S. L. (2008). Spiral analysis—Improved clinical utility with center detection. *Journal of Neuroscience Methods*, 171(2), 264–270. https://www.sciencedirect.com/science/article/pii/S016502700800174X?via%3Dihub

# A Appendix

## New Features

Table 3: Final Dataset Structure

| Column | Type | Description |
|---|---|---|
| spiral_id | Integer | A unique identifier for each spiral. |
| mean_pressure | Float | The average pressure applied during the drawing session. |
| std_pressure | Float | The standard deviation of pressure applied during the drawing session. |
| mean_altitude | Float | The average altitude angle (vertical orientation) of the pen. |
| std_altitude | Float | The standard deviation of the altitude angle during the drawing session. |
| mean_azimuth | Float | The average azimuth angle (horizontal orientation) of the pen. |
| std_azimuth | Float | The standard deviation of the azimuth angle during the drawing session. |
| pen_up_time | Float | The total time when the pen was not in contact with the surface (pen lifted). |
| pen_down_time | Float | The total time when the pen was in contact with the surface (pen down). |
| stroke_count | Integer | The number of times the pen was lifted off the surface during the drawing. |
| total_distance | Float | The total Euclidean distance covered during the drawing session. |
| mean_velocity | Float | The average speed at which the pen moved between points. |
| mean_acceleration | Float | The average change in velocity over time during the drawing session. |
| mean_jerk | Float | The average rate of change of acceleration (smoothness of movement). |
| patient_group | Object | The group classification: "Healthy" for controls and "Parkinson's" for patients. |

## SVM Recursive Feature Elimination(RFE) Results

Table 4: SVM RFE Results

| | Feature | Ranking | Selected |
|---|---|---|---|
| 0 | mean_pressure | 5 | False |
| 1 | std_pressure | 1 | True |
| 2 | mean_altitude | 6 | False |
| 3 | std_altitude | 8 | False |
| 4 | mean_azimuth | 1 | True |
| 5 | std_azimuth | 2 | False |
| 6 | pen_up_time | 1 | True |
| 7 | pen_down_time | 1 | True |
| 8 | stroke_count | 3 | False |
| 9 | total_distance | 1 | True |
| 10 | mean_velocity | 4 | False |
| 11 | mean_acceleration | 7 | False |
| 12 | mean_jerk | 9 | False |