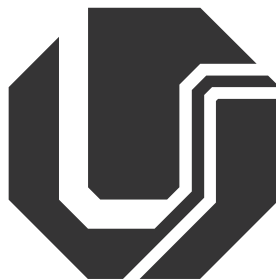


FEDERAL UNIVERSITY OF UBERLÂNDIA

Faculty of Electrical Engineering

POSTGRADUATE PROGRAM IN BIOMEDICAL ENGINEERING



Pedro Henrique Bernardes Caetano Milken

Machine learning applied for the motor pattern classification of
pen-and-paper spiral drawings in Parkinson's disease

UBERLÂNDIA, MG

2025

Pedro Henrique Bernardes Caetano Milken

Machine learning applied for the motor pattern classification of
pen-and-paper spiral drawings in Parkinson's disease

Master's Dissertation presented to the
Postgraduate Program in Biomedical
Engineering at the Federal University
of Uberlândia, as part of the require-
ments for obtaining the title of Master
of Science.

Line of research: Rehabilitation En-
gineering and Biomechanics

Supervisor: Prof. Adriano de
Oliveira Andrade, PhD

Co-supervisor: Lígia Reis Nóbrega,
DSc

UBERLÂNDIA, MG

2025

Ficha Catalográfica Online do Sistema de Bibliotecas da UFU
com dados informados pelo(a) próprio(a) autor(a).

M644
2025

Milken, Pedro Henrique Bernardes Caetano, 1999-
Aprendizado de máquina aplicado à classificação de padrões
motores de desenhos em espiral feitos à mão em pacientes com
doença de Parkinson [recurso eletrônico] / Pedro Henrique
Bernardes Caetano Milken. - 2025.

Orientador: Adriano de Oliveira Andrade.
Coorientadora: Lígia Reis Nóbrega.
Dissertação (Mestrado) - Universidade Federal de Uberlândia,
Pós-graduação em Engenharia Biomédica.
Modo de acesso: Internet.
DOI <http://doi.org/10.14393/ufu.di.2025.488>
Inclui bibliografia.

1. Engenharia biomédica. I. Andrade, Adriano de Oliveira, 1975-,
(Orient.). II. Nóbrega, Lígia Reis, 1993-, (Coorient.). III.
Universidade Federal de Uberlândia. Pós-graduação em
Engenharia Biomédica. IV. Título.

CDU: 62:61



UNIVERSIDADE FEDERAL DE UBERLÂNDIA
Coordenação do Programa de Pós-Graduação em Engenharia
Biomédica

Av. João Naves de Ávila, 2121, Bloco 3N, Sala 115 - Bairro Santa Mônica, Uberlândia-MG,
CEP 38400-902

Telefone: (34) 3239-4761 - www.ppgeb.feelt.ufu.br - ppegb@feelt.ufu.br



ATA DE DEFESA - PÓS-GRADUAÇÃO

Programa de Pós-Graduação em:	Engenharia Biomédica				
Defesa de:	Dissertação de Mestrado Acadêmico, 122, PPGEB				
Data:	oito de agosto de dois mil e vinte e cinco.	Hora de início:	08:00	Hora de encerramento:	10:25
Matrícula do Discente:	12322EBI007				
Nome do Discente:	Pedro Henrique Bernardes Caetano				
Título do Trabalho:	Machine Learning Applied for the Motor Pattern Classification of Pen-and-Paper Spiral Drawings in Parkinson's Disease				
Área de concentração:	Engenharia Biomédica				
Linha de pesquisa:	Engenharia de Reabilitação e Biomecânica				
Projeto de Pesquisa de vinculação:	PARKINSON BRASIL Base de dados de aspectos motores e não motores para o avanço da avaliação automatizada da doença de Parkinson no Brasil				

Reuniu-se via plataforma Teams, Campus Santa Mônica, da Universidade Federal de Uberlândia, a Banca Examinadora, designada pelo Colegiado do Programa de Pós-graduação em Engenharia Biomédica, assim composta: Professores Doutores: Alberto López Delis - Universidade de Oriente; Luanne Cardoso Mendes - UFU; Adriano de Oliveira Andrade - PPGEB/UFU orientador do candidato.

Iniciando os trabalhos, o presidente da mesa, Dr. Adriano de Oliveira Andrade apresentou a Comissão Examinadora e o candidato, agradeceu a presença do público e concedeu ao discente a palavra para a exposição do seu trabalho. A duração da apresentação do discente e o tempo de arguição e resposta foram conforme as normas do Programa.

A seguir, o senhor presidente concedeu a palavra, pela ordem sucessivamente, aos examinadores, que passaram a arguir o candidato. Ultimada a arguição, que se desenvolveu dentro dos termos regimentais, a Banca, em sessão secreta, atribuiu o resultado final, considerando o candidato:

Aprovado.

Esta defesa faz parte dos requisitos necessários à obtenção do título de Mestre.

O competente diploma será expedido após cumprimento dos demais requisitos, conforme as normas do Programa, a legislação pertinente e a regulamentação interna da UFU.

Nada mais havendo a tratar, foram encerrados os trabalhos. Foi lavrada a presente ata que, após lida e achada conforme, foi assinada pela Banca Examinadora.



Documento assinado eletronicamente por **Adriano de Oliveira Andrade, Professor(a) do Magistério Superior**, em 08/08/2025, às 10:27, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Luanne Cardoso Mendes, Professor(a) do Magistério Superior**, em 08/08/2025, às 10:30, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Alberto López Delis, Usuário Externo**, em 13/08/2025, às 08:42, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site https://www.sei.ufu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **6576992** e o código CRC **E5EF29CD**.

ACKNOWLEDGEMENTS

I extend my sincere gratitude to all the volunteers who kindly agreed to participate in this research, as well as to the “Mãos que vibram” project in Uberlândia, Minas Gerais, Brazil, for their valuable collaboration.

I am deeply grateful to my advisor, Prof. Adriano de Oliveira Andrade, for his invaluable guidance, continuous support, and encouragement throughout this journey.

I also express my heartfelt thanks to my co-advisor, Dr. Lígia dos Reis Nóbrega, for her trust, time, and dedication, as well as for the insightful meetings and discussions that helped keep me focused and motivated.

I would like to thank the examination committee for accepting the invitation to evaluate this work and for their constructive feedback and important contributions to its improvement.

I also thank my colleagues in the laboratory for the friendships developed over the years, the knowledge exchanged, and their support at various stages of this research.

A special note of gratitude goes to my grandmother, whose constant presence, wisdom, and emotional support were invaluable throughout both the challenges and triumphs of this journey.

I dedicate heartfelt thanks to my “three little angels” — Jack Brian Juno, Melanie, and Paçoca Bernardes — whose love and companionship brought me strength and inspiration during this period.

I am profoundly grateful to my parents and my brother for their unconditional love, support, and belief in my potential.

To my friends, thank you for walking this path with me, sharing its difficulties, achievements, and moments of joy.

I hope that this work will serve as inspiration to the younger generation in my family—especially my cousin Davi Bernardes Barbosa—encouraging them to pursue knowledge and contribute to scientific progress.

This research was supported by the National Council for Scientific and Technological Development (CNPq - 442150/2023-7, 131534/2024-6, 405365/2023-3, 444437-2024-0 and 302942/2022-0), the Coordination for the Improvement of Higher Education Personnel (CAPES - CAPES Program), and the Research Support Foundation of the State of Minas Gerais (FAPEMIG). Their support is gratefully acknowledged.

ABSTRACT

Parkinson's disease (PD) is a progressive neurodegenerative disorder characterized by impaired voluntary motor control, most commonly manifesting as tremors, bradykinesia, and muscle rigidity. Despite significant advancements in the understanding of PD, clinical evaluation remains the primary method for diagnosis and disease monitoring, often relying on subjective assessments. Spiral analysis has emerged as a promising, low-cost, and non-invasive technique for the objective quantification of motor symptoms through the study of hand-drawn spirals. Unlike previous studies that utilized digitizing tablets and digital pens, the present research investigates the feasibility of using traditional pen-and-paper spirals in a clinical setting. This study aims to develop objective tools to assist healthcare professionals in clinical decision-making. It is part of a broader project called PARKINSON BRASIL, which seeks to develop a comprehensive national database containing multidimensional aspects of individuals with Parkinson's disease. Four deep learning architectures—VGG16, ResNet50, MobileNetV2, and Xception—were employed for feature extraction, followed by classification using error-correcting output codes (ECOC). Participants were classified into either PD or control groups, and PD participants were further categorized into disease stages according to the Hoehn and Yahr scale. Model performance was evaluated using standard metrics, including accuracy and F1-score. Xception achieved the best overall performance. ResNet50 and VGG16 also demonstrated high precision in their performance. The findings demonstrate that pen-and-paper spiral analysis, when integrated with machine learning techniques, holds significant potential as a complementary tool for the diagnosis and monitoring of Parkinson's disease in clinical environments.

Keywords: Parkinson's disease, Deep Learning, Spiral Analysis, Convolutional Neural Network, Machine learning

Graphical Abstract

Machine Learning Application in the Analysis of Spiral Drawings for Pattern Identification in Parkinson's Disease



Data Collection & Processing

- Participants: 25 Parkinson's disease and 15 age-matched healthy individuals
- The 25 participants with Parkinson's disease were classified into Hoehn & Yahr stages (H&Y 1, 2, 3)
- 2,811 spirals were drawn on paper with pen, scanned (Scanner Epson Perfection V370), and preprocessed (CorelDRAW and MATLAB)
- The following models were tested: VGG16, ResNet50, MobileNetV2, Xception



Classification by Severity (H&Y 1, 2, 3)

- Performance decreases as the disease severity increases
- Greater difficulty in differentiating advanced disease stages



Conclusion 1

Handwritten spiral evaluation can be a complementary and accessible method for monitoring and characterizing PD



Conclusion 3

The study suggests that the choice of deep neural network model depends on the application context and clinical scenario. Models like VGG16 and ResNet50 are suitable for simple classification tasks, while Xception and MobileNetV2 are versatile for PD progression monitoring. Techniques like data augmentation and domain adaptation enhance model practical applicability in clinical settings.



Objectives

- Assess whether different machine learning algorithms or tools show significant performance differences in classifying motor assessments based on spiral drawings
- Investigate the correlation between MDS UPDRS scores (especially H&Y's rating) and motor evaluation results derived from spiral drawings
- Explore the potential of using spiral drawings to monitor motor symptoms in Parkinson's disease



Classification Results

- Xception achieved the best overall performance
- ResNet50 and VGG16 also demonstrated high precision



Conclusion 2

The study compared the models architectures across 18 tasks related to tremor, bradykinesia and rigidity. Results showed that neural network performance varies depending on the task, with ResNet50 being the most robust model. MobileNetV2 and Xception achieved competitive results, while VGG16 showed isolated strengths.



List of Figures

4.1	Research methodology.	19
4.2	Overview of the longitudinal data collection, illustrated at both the group and individual levels. (a) Monthly number of collected spiral drawings, with separate bars indicating the respective groups. (b) Timeline of data collection for each participant, showing the month and year of participation. Note that although some participants were assessed multiple times in a single month, only one entry per month is shown in the graph.	20
4.3	Example of an image corrected and individualized during editing without altering the original characteristics of the original drawings.	22
4.4	Samples of the collected dataset. Spirals from individuals with PD (V09, V05, V18, and V03) and a subject of the control group (V35—highlighted) are shown. The spirals were digitised from the actual instrument used for data collection. The arrow in the ideal spiral template indicates the intended starting point of the drawing. Note that some subjects had difficulty following the ideal spiral template strictly.	23
4.5	Data acquired for CG volunteers.	24
4.6	Data acquired for PD volunteers.	24
4.7	Performance statistics of different CNN models for classification between Parkinson's Disease (PD) and Healthy Controls (HC).	32
4.8	Performance statistics of different CNN models for classification between Parkinson's Disease patients in Hoehn & Yahr stages 1–3 and Healthy Controls (HC).	33
4.9	ResNet50 radar plot.	37
4.10	VGG16 radar plot.	37
4.11	MobileNetV2 radar plot.	38
4.12	XCEPTION radar plot.	38
4.13	Statistics for different CNN models comparing different features for classification.	39
4.14	F1-scores by model and group.	42
4.15	Training vs. testing accuracy by model.	42
4.16	F1-score and accuracy in different data collection sessions.	43
5.1	Handwritten spiral protocol.	55
5.2	Spiral drawings using laser protocol.	56

A.1 Example of the instrument used for data collection and a typical data sample
 (on the right) obtained from a person with Parkinson’s disease. 63

List of Tables

2.1	<i>Stages of the Scale</i>	10
3.1	Main characteristics of previous studies on Parkinson's disease detection based on drawing tasks.	15
4.1	Distribution of spiral drawing samples and clinical characterisation by group. .	21
4.2	Overall Accuracy, Cohen's Kappa, Precision, Recall, and F1-score by class (CG vs PD).	31
4.3	Overall Accuracy, Cohen's Kappa, Precision, Recall, and F1-score by class (CG vs H&Y1-3).	32
4.4	Average mean F1-score by classification using VGG16, Xception, MobileNetV2 and ResNet50 comparing different features.	33
4.5	Average mean precision by classification using VGG16, Xception, MobileNetV2, and ResNet50 comparing different features.	34
4.6	Average mean recall by classification using VGG16, Xception, MobileNetV2, and ResNet50 comparing different features.	35
4.7	Average Cohen's Kappa by classification using MobileNetV2, Xception, ResNet50, and VGG16 comparing different features.	36
4.8	Average Metric Values for the Training Set	39
4.9	Average Metric Values for the Test Set	40
4.10	Accuracy, F1-Score and Cohen's Kappa per Collection for MobileNetV2	40
4.11	Accuracy, F1-Score and Cohen's Kappa per Collection for Xception	40
4.12	Accuracy, F1-Score and Cohen's Kappa per Collection for VGG16	41
4.13	Accuracy, F1-Score and Cohen's Kappa per Collection for ResNet50	41

List of Abbreviations

CNN Convolutional neural networks

DNN Deep neural networks

ECOC Error-correcting output codes

GC Control group

H&Y Hoehn & Yahr scale

HOG Histograms of oriented gradients

MDS-UPDRS Movement Disorders Society - Unified Parkinson's Disease Rating Scale

ML Machine learning

PD Parkinson's disease

SA Spiral analysis

Contents

List of Abbreviations	v
1 Introduction	1
1.1 Problem formulation	1
1.2 Relevance of the study	2
1.3 The object of the research	2
1.4 The aim of the thesis	3
1.5 The objective of the thesis	3
1.6 Practical value of the research findings	3
1.7 Approval of the research findings	3
1.8 Note on Terminology	4
1.9 Structure of the dissertation	4
1.10 Academic Qualifications Acquired During the Master's Program	5
2 Parkinson's disease	6
2.1 Physiology and epidemiology	6
2.2 Clinical and motor changes in PD	6
2.3 Clinical evaluation of the upper limbs in Parkinson's disease	7
2.4 MDS-UPDRS and Hoehn & Yahr scale	8
2.5 Hoehn & Yahr stages	10
3 Spiral analysis in Parkinson's disease	11
3.1 Spiral analysis	11
3.1.1 Review of Related Studies	13
3.1.2 Machine learning for spiral analysis	14
4 Spiral drawing protocol: classification of handwritten motor patterns in Parkinson's disease	17
4.1 Approach	17
4.2 Research methodology	18
4.3 Data collection protocol	18
4.4 Scanning procedure	20

4.5	Data storage and dataset	21
4.6	Data processing	24
4.6.1	Feature extraction	24
4.7	Analysis methods	27
4.7.1	Classification models	27
4.8	Statistical analysis	28
4.9	Results	30
4.9.1	Binary and multiclass classification analysis	30
4.9.2	Classification analysis using MDS-UPDRS	31
4.9.3	Classification analysis for monitoring handwritten spirals	39
4.10	Discussion	41
4.10.1	Binary and multiclass classification analysis	41
4.10.2	Classification analysis using MDS-UPDRS	45
4.10.3	Classification analysis for monitoring handwritten spirals	46
4.11	Conclusion	49
4.11.1	Binary and multiclass classification analysis	49
4.11.2	Classification analysis using MDS-UPDRS	49
4.11.3	Classification analysis for monitoring handwritten spirals	50
5	Future studies	52
5.1	General conclusions of the study	52
5.2	Image patterns	53
5.2.1	Handwritten spiral images	53
5.2.2	Digital tool using mouse for spiral drawings images	54
5.2.3	Spiral drawings using laser	54
	References	57
	Glossary	61
A	Appendix 1	62
A.1	Environment for data collection	62
B	Appendix 2	64
B.1	Codes for the analysis	64
B.2	Open access database	64

Introduction

1.1 Problem formulation

Parkinson’s disease (PD) is a progressive neurodegenerative disorder primarily characterized by motor impairment, including resting tremor, bradykinesia, muscle rigidity, and alterations in gait and posture. Despite growing scientific understanding of the disease, clinically validated methods for quantitatively assessing motor function in individuals with PD remain limited.

Spiral analysis (SA) has emerged as a non-invasive, objective technique to evaluate motor performance through the analysis of hand-drawn spirals. This method captures essential motor characteristics by analyzing spatial and temporal features such as tightness, regularity, speed, pressure, and drawing duration. These parameters are extracted from conventional clinical spiral drawing tasks and converted into computer-based assessments, allowing for quantifiable evaluation of motor function (24).

The core motor symptoms of PD—bradykinesia, resting tremor, rigidity, and postural instability—are typically assessed using the Unified Parkinson’s Disease Rating Scale (UPDRS). When administered by a trained movement disorder specialist, the UPDRS serves as a reliable indicator of disease severity. Functional neuroimaging studies have also demonstrated correlations between UPDRS scores and reduced fluorodopa uptake. However, the accuracy of UPDRS assessments decreases when conducted by general neurologists, due to the scale’s subjective nature and dependence on examiner expertise (29).

Our group initiated research on spiral analysis in 2019 with the study by Folador et al. (11), which introduced a method for recognizing spiral and sine wave patterns using Histograms of Oriented Gradients (HOG). Data were collected from 15 PD patients and 12 healthy controls, each of whom produced three to four spiral and sine wave drawings. A total of 102 images were digitized and classified based on gradient and edge orientation patterns using HOG features.

Subsequent studies expanded upon this work. Rosebrock (28) extended classification from sine waves to spirals using H-based descriptors. DeSipio et al. (9) applied machine

learning and deep learning techniques to analyze tremor patterns via the Archimedean Spiral Test. Their proposed Fourier Domain analysis method converted spiral drawings into frequency-based features, achieving an accuracy of 81.5%—a 6% improvement over previous methods.

These advancements contribute to the characterization of motor dysfunction in PD and support its monitoring and progression tracking. The present study, as part of the PARKINSON BRASIL project, aims to evaluate the feasibility and effectiveness of using pen-and-paper spiral drawings as a clinical tool for PD assessment. Given that clinical observation remains the primary diagnostic and monitoring approach in current practice, this research seeks to validate whether low-cost, non-digital spiral analysis can offer reliable, objective data. Ultimately, the findings are intended to inform clinical decision-making and support the integration of new technologies into routine care for individuals with Parkinson’s disease.

1.2 Relevance of the study

Despite numerous advances in the understanding of Parkinson’s disease (PD), the processes of monitoring, measuring, treating, and diagnosing the condition remain complex and challenging in clinical practice. Diagnosis often depends on clinical examinations and subjective assessments, which can introduce variability and potential bias. In this context, the development of standardized assessment tools capable of generating consistent, objective, and reproducible data would be highly beneficial. Such tools would support healthcare professionals in making more accurate and evidence-based decisions when evaluating individuals with PD. Therefore, the creation of reliable instruments designed for use by trained practitioners represents a valuable step toward enhancing diagnostic precision and treatment monitoring.

1.3 The object of the research

The spiral analysis tools employed in this study were developed as part of the PARKINSON BRASIL project (ClinicalTrials.gov Identifier: NCT06536348), a nationwide initiative led by the Federal University of Uberlândia. The project seeks to build a comprehensive database encompassing both motor and non-motor symptoms of Parkinson’s disease (PD) in Brazil. Its overarching goal is to support clinical, computational, and neuroscientific research, thereby contributing to a deeper understanding of PD within the national context.

As of March 14, 2025, the dataset was last updated. Further information about the project is available at ClinicalTrials.gov/NCT06536348. The present master’s research is a derivative of this broader initiative.

In this study, pen-and-paper spiral drawings were used. Each participant was asked to draw five spirals on an A4 sheet. The sheets were then digitally scanned, and individual spirals were segmented and labeled for subsequent analysis.

1.4 The aim of the thesis

The general objective of this study is to investigate the feasibility of using spiral-based motor assessment techniques for evaluating motor symptoms in Parkinson's disease through the application of machine learning algorithms.

1.5 The objective of the thesis

The specific objectives of this study are:

- To assess whether different machine learning (ML) algorithms exhibit significant differences in performance when classifying motor symptoms using spiral-based assessment methods.
- To analyze the correlation between the MDS-UPDRS scores and the results obtained from spiral-based motor assessments.
- To investigate the potential of spiral drawings as a tool for monitoring motor symptoms in individuals with Parkinson's disease.

1.6 Practical value of the research findings

This study demonstrated significant practical value by enabling the application of the proposed tools in clinical settings without incurring high costs or requiring advanced technological infrastructure. These tools are designed to enhance patient care and streamline clinical workflows, ensuring accessibility for healthcare professionals across diverse contexts. By emphasizing simplicity and effectiveness, the proposed methods are well-suited for real-world implementation in routine clinical practice. Furthermore, these tools offer objective and quantitative insights into motor symptoms associated with Parkinson's disease, serving as valuable adjuncts to support clinical decision-making.

1.7 Approval of the research findings

Milken2024.

1.8 Note on Terminology

To facilitate the understanding of technical terminology and acronyms used throughout this dissertation, a glossary is provided at the end of the document. Readers are encouraged to consult it as needed for clarification of key concepts.

1.9 Structure of the dissertation

This thesis is organized to guide the reader through a coherent sequence of topics, beginning with the contextualization of the research problem and culminating in the discussion of findings and recommendations for future investigations. The document opens with a list of abbreviations, consolidating the principal acronyms used throughout the text to enhance clarity and readability.

Chapter 1 outlines the foundational elements of the research. It presents the problem statement, research significance, subject definition, general and specific objectives, and the methodology employed. Additionally, this chapter addresses the practical relevance of the findings, discusses the validation of results, and offers commentary on the terminology adopted. It concludes with an overview of the thesis structure.

Chapter 2 provides a theoretical framework on Parkinson's disease, including an analysis of its physiological and epidemiological aspects, key clinical and motor manifestations, and upper limb assessment methods. Emphasis is placed on widely adopted scales such as the MDS-UPDRS and the Hoehn and Yahr scale.

Chapter 3 explores spiral analysis as a tool for motor function assessment in Parkinson's disease. It introduces the theoretical foundations of spiral analysis and investigates the application of machine learning techniques for interpreting these drawings.

Chapter 4 describes the experimental protocol developed in this study. It details the overall methodology, data acquisition procedures, scanning methods, data organization, and dataset construction. The chapter then addresses data processing stages, including feature extraction, analysis techniques, classification models, statistical approaches, and the results obtained. A comprehensive analysis of the three classification scenarios is provided, along with corresponding conclusions.

Chapter 5 presents recommendations for future research, with emphasis on exploring innovative approaches in visual pattern analysis. It proposes investigating alternative data collection methods for spiral drawings, including digital tools such as a mouse and laser tracking technologies.

The thesis concludes with a references section listing all bibliographic sources cited, followed by a glossary that defines key technical terms used in the study.

Two appendices supplement the document: Appendix 1 describes the environment used for data collection, while Appendix 2 contains the analysis code and details about

the open-access database created during the research.

This structure is designed to ensure clarity, logical flow, and ease of navigation, thereby facilitating a comprehensive understanding of the research undertaken.

1.10 Academic Qualifications Acquired During the Master's Program

During the course of the master's program, clinical and research competencies were enhanced through the acquisition of several internationally recognized certifications related to the assessment of Parkinson's disease and cognitive screening. Three certification programs were successfully completed through the Movement Disorder Society (MDS): the MDS Non-Motor Rating Scale (MDS-NMS), the Unified Dyskinesia Rating Scale (UDysRS), and the MDS-Unified Parkinson's Disease Rating Scale (MDS-UPDRS). These credentials ensure a standardized and globally recognized level of expertise in the clinical evaluation of both motor and non-motor symptoms in individuals with Parkinson's disease.

In addition, official MoCA Certification was obtained from MoCA Cognition, an organization dedicated to improving cognitive screening tools through research and innovation. The MoCA Training & Certification Program is a required qualification for professionals administering the paper-based version of the Montreal Cognitive Assessment (MoCA), as it ensures standardized administration and scoring procedures.

Furthermore, active participation in the Mãos que Vibram Extension Project contributed to the development of interdisciplinary practices targeting individuals with Parkinson's disease. This initiative promotes health, leisure, and cultural engagement through various physical and cognitive activities aimed at supporting longevity and improving quality of life. The project provided opportunities for direct interaction with individuals affected by Parkinson's disease, enabling integrated therapeutic approaches that address both motor and non-motor aspects of patient care.

Parkinson's disease

2.1 Physiology and epidemiology

Parkinson's disease (PD) is a progressive neurodegenerative disorder characterized by rigidity, tremor, and bradykinesia. It is associated with the gradual degeneration of dopaminergic neurons in the substantia nigra, as well as other affected brain regions. In addition to motor manifestations, non-motor symptoms such as dysautonomia and dementia are commonly observed, particularly in the later stages of the disease. Notably, the term "Parkinson's disease" is not universally defined among clinicians and researchers, nor is it necessarily regarded as a single disease entity. Multiple pathological mechanisms may underlie the condition, and the term is often applied based on clinical presentation rather than a uniform etiological framework (33).

Estimates of PD incidence range from 5 to over 35 cases per 100,000 individuals annually. The incidence increases five- to ten-fold between the sixth and ninth decades of life. Similarly, prevalence rises with age. In a meta-analysis involving four North American cohorts, prevalence increased from under 1% among individuals aged 45–54 to approximately 2% of males and 4% of females aged 85 years or older. As global populations age, the prevalence of PD is projected to more than double over the next two decades. Consequently, the associated social and economic burdens are also expected to increase, underscoring the urgent need for more effective preventive, therapeutic, and curative interventions (30).

2.2 Clinical and motor changes in PD

Additional risk factors for Parkinson's disease (PD) include a family history of PD or essential tremor, chronic constipation, and absence of tobacco use—all of which are associated with at least a twofold increase in the risk of developing the disease. Diagnosis is primarily clinical, with supplementary diagnostic testing reserved for cases exhibiting atypical features.

Historically, the clinical diagnosis of PD has centered on the presence of motor symptoms. According to the Movement Disorder Society Parkinson's Disease (MDS-PD) criteria, motor manifestations remain the cornerstone of diagnosis. Specifically, the identification of parkinsonism—defined as bradykinesia in combination with either resting tremor, rigidity, or both—is required. These motor signs must be clearly demonstrable and not attributable to confounding factors (27).

Nonetheless, non-motor symptoms are prevalent in the majority of patients and can often overshadow motor features in the clinical presentation. A wide range of non-motor manifestations has now been incorporated into the diagnostic criteria. Furthermore, the neurodegenerative process in PD often begins in nondopaminergic regions of the brain or the peripheral nervous system, where non-motor symptoms typically predominate. This understanding has led to the recognition of prodromal PD as a distinct and clinically meaningful stage of the disease. Importantly, prodromal PD is not merely a risk state, but rather a phase within the spectrum of PD itself.

Prognostic considerations are often guided by PD subtypes and the duration of the prodromal phase, which can precede motor symptom onset by several years. At present, the recognition of early or prodromal features has limited therapeutic implications beyond symptomatic management. However, the ability to diagnose prodromal PD is expected to have substantial clinical relevance once disease-modifying therapies become available (2).

Treatment objectives in PD vary across individuals, highlighting the importance of personalized care strategies. There is no clinical justification for delaying symptomatic treatment. Levodopa remains the most commonly prescribed first-line pharmacological therapy. Optimal management should begin at the time of diagnosis and requires a multidisciplinary approach, incorporating both pharmacological and non-pharmacological interventions. Although no current therapy halts disease progression, several promising approaches are under investigation, particularly in light of recent advances in the understanding of genetic risk factors and the molecular mechanisms underlying neuronal degeneration (2).

2.3 Clinical evaluation of the upper limbs in Parkinson's disease

Several clinical features observed in patients with Parkinson's disease (PD) are consistent with parkinsonian tremor. This tremor is typically asymmetric and most prominent at rest, although it may persist—albeit with reduced amplitude—when the affected limb is held against gravity. Parkinsonian tremor often diminishes or disappears during voluntary movement, such as during a finger-to-nose test or with postural adjustments, and reappears after a brief delay, a phenomenon referred to as re-emergent tremor. Depend-

ing on the context of its appearance, this symptom may be classified as resting, postural, or re-emergent tremor. It is frequently observable in the affected upper limb while the individual is walking. The most specific pathological criteria for the diagnosis of PD include asymmetric resting tremor and a positive clinical response to levodopa; however, approximately 30% of patients do not exhibit tremor at the time of initial presentation (22).

Patients with the tremor-dominant subtype of PD, which is characterized by a coarse resting tremor, are often misdiagnosed with essential tremor. Tremor-dominant PD generally presents a more benign disease course and is initially unaccompanied by significant motor deficits, rigidity, or bradykinesia. Furthermore, the tremor in this subtype is relatively unresponsive to levodopa therapy. Additional clinical signs—such as facial hypomimia or asymmetric arm swing during gait—may assist in differentiating PD from other tremor disorders. Adult-onset dystonic tremor is another differential diagnosis for upper limb tremor, though it occurs less frequently in PD. Like parkinsonian tremor, dystonic tremor can be unilateral or asymmetric and may occur at rest, during specific postures, or with movement. However, distinguishing features of dystonic tremor include position- or task-specificity and the presence of additional dystonic signs, such as abnormal limb posturing. Importantly, patients with dystonic tremor typically do not present with other cardinal PD features, such as decremental bradykinesia, hyposmia, or rapid eye movement sleep behavior disorder (RBD) (22).

Clinicians should also evaluate for bradykinesia-related limb symptoms, such as difficulty performing repetitive or rotational tasks (e.g., brushing teeth, washing hair) and challenges with fine motor coordination (e.g., handwriting, buttoning clothes). Inquiry into axial motor difficulties—including trouble rising from a chair, exiting a car, or turning in bed—may further support a diagnosis of parkinsonism. On physical examination, in addition to tremor, key clinical signs include reduced spontaneous facial expression and blinking, resulting in a fixed or masked facial appearance. Increased muscle tone (rigidity) is typically noted in the limbs and trunk. Bradykinesia may be demonstrated by reduced amplitude in repeated finger-tapping tasks. A hypokinetic gait, characterized by short steps and diminished arm swing—often more evident on the affected side—is also common. Resting tremor may re-emerge in the affected upper limb while walking (22).

2.4 MDS-UPDRS and Hoehn & Yahr scale

The assessment of Parkinson’s disease (PD) is critical for evaluating both motor and non-motor symptoms, which vary widely in severity and impact among patients. Among the instruments developed for this purpose, the Movement Disorder Society–Unified Parkinson’s Disease Rating Scale (MDS-UPDRS) has become the most widely used and accepted

tool in both clinical and research settings. The MDS-UPDRS was designed to provide a comprehensive, flexible, and standardized method for quantifying the impairments and disabilities associated with Parkinson’s disease, integrating elements from earlier assessment scales to enhance clinical utility. The Unified Parkinson’s Disease Rating Scale (UPDRS) was originally developed in 1984 by a dedicated committee responding to the need for consistent clinical evaluation of PD. In 1987, Fahn and Elton formally introduced the UPDRS, which encompassed both clinician-administered motor assessments and patient-reported evaluations of daily functional abilities. In 2008, the International Parkinson and Movement Disorder Society revised the scale extensively, resulting in the MDS-UPDRS. This updated instrument is now regarded as the gold standard for diagnosing and monitoring the progression of Parkinson’s disease (26).

The MDS-UPDRS comprises four parts, each addressing a distinct aspect of the disease. Part I evaluates non-motor experiences, including fatigue, anxiety, and apathy, with input from both the patient and clinician. Part II assesses motor experiences of daily living, such as walking, eating, and dressing, and is completed solely by the patient. Part III consists of a motor examination performed by the clinician, utilizing standardized instructions and demonstrations to evaluate motor function. Part IV focuses on motor complications—including dyskinesia, dystonia, and motor fluctuations—and is completed by the clinician. Each item is rated on a 5-point scale from 0 (normal) to 4 (severe), allowing a graded evaluation of symptom severity and functional impact (26).

In parallel, the Hoehn and Yahr (H&Y) scale remains a widely utilized instrument for staging Parkinson’s disease, particularly valued for its simplicity and clinical applicability. Originally developed in the 1960s, the H&Y scale was subsequently reviewed by the Movement Disorder Society Task Force on PD Rating Scales. Its advantages include a strong correlation with conventional clinical measures of motor impairment, disability, quality of life, and neuroimaging markers of dopaminergic neuron loss. However, despite its widespread acceptance, the H&Y scale has notable limitations. These include an uneven stage structure that conflates impairment and disability, and an overemphasis on postural instability as the primary indicator of disease severity. Consequently, the H&Y scale does not fully capture the spectrum of motor dysfunctions and provides no assessment of non-motor symptoms (14).

Although direct clinimetric evaluation of the H&Y scale has been limited, it demonstrates acceptable reliability and validity, especially within its intermediate stages (Stages 2–4). A modified version employing 0.5-point increments has gained popularity in clinical practice, but lacks formal clinimetric validation. In light of these factors, the Movement Disorder Society Task Force recommends the continued use of the original H&Y scale, particularly for demographic characterization of patient cohorts. It further advises reporting medians and ranges for group comparisons and utilizing nonparametric statistical methods for longitudinal analyses. In research contexts, the scale remains valuable for defining

inclusion and exclusion criteria. In clinical practice, its application should emphasize simplicity and direct observation, summarized as “rate what you see.” Until further validation is conducted, the original five-stage H&Y scale should remain the standard (14).

2.5 Hoehn & Yahr stages

In order to clarify the different stages proposed by the Hoehn & Yahr scale, the following table 2.1 differentiates the characteristics of the 5 different stages proposed by the staging (14).

Table 2.1 – *Stages of the Scale*

Hoehn and Yahr scale	
1	Unilateral involvement is typically associated with minor or no functional disability
2	Bilateral or midline involvement without loss of balance
3	Bilateral disease: mild to moderate impairment, altered postural reflexes; physically independent
4	Severely impairing sickness; still able to walk or stand without assistance
5	Confinement to bed or wheelchair unless assisted

Spiral analysis in Parkinson's disease

3.1 Spiral analysis

Non-invasive, objective assessments of motor function can facilitate early detection of Parkinson's disease (PD), quantify disease severity, and monitor progression in both clinical practice and research trials. Validated kinematic tests that measure motor performance provide objective and reproducible metrics amenable to statistical analysis. However, few methods are currently available for clinical evaluation of motor function in PD patients (29).

Spiral analysis is a non-invasive technique that quantifies motor performance by analyzing kinematic and physiological characteristics of handwritten spiral drawings. This method employs computational algorithms to measure parameters such as force, speed, time, tightness, and homogeneity of the spiral, capturing spatial and temporal data. Spiral data are collected along the x, y, and pressure axes, resulting in virtual tri-axial recordings when using a digital pen and tablet, or as two-dimensional coordinate data when employing pen-and-paper drawings (29).

The growing number of published studies highlights spiral analysis (SA) as a promising tool in PD diagnostic research. Its practicality and cost-effectiveness have led to SA-based PD detection outperforming traditional clinical, laboratory, and imaging methods (24).

SA provides an objective assessment of PD severity by extracting detailed motor features from the standard clinical spiral drawing test. Ideally, objective motor performance measures should correlate with disease severity as assessed by the Unified Parkinson's Disease Rating Scale (MDS-UPDRS). However, variations between clinical rating scales and computerized motor data often result in modest correlations. Spiral analysis captures motor performance components relevant to clinical status but not evaluated by MDS-UPDRS, which may contribute to weaker associations (29).

Despite the widespread use of spiral drawing in neurological examinations, quantitative analysis remains underutilized. Computerized spiral analysis (CSA) objectively evaluates upper limb kinematics through spiral drawings captured on digitizing tablets, providing

reliable and rapid assessments of motor function. Certain SA metrics, including second-order smoothness, severity degree, and mean speed, correlate with multiple motor function subdomains, with strongest associations in the arm, followed by side and whole-body measures—consistent with the hand-based nature of the task. The strong correlation with total motor MDS-UPDRS scores in early-stage PD patients indicates SA’s utility for assessing both arm-specific and global motor function. Moreover, as a neuropsychological tool, spiral drawing may probe brain circuits implicated in parkinsonian symptoms beyond limb movement. However, the strong link with total MDS-UPDRS scores may partially reflect the influence of arm scores on the overall scale (29).

Specific spiral metrics align with distinct MDS-UPDRS subdomain scores, underscoring SA’s potential for isolated domain evaluation. For example, second-order smoothness, which quantifies deviations from an ideal spiral, correlates with bradykinesia, rigidity, and action tremor but not rest tremor. Mean speed associates with rigidity, bradykinesia, and action tremor, whereas first-order crossing correlates with bradykinesia and action tremor but not rigidity or rest tremor. Rest tremor shows minimal correlation with most spiral metrics, except severity degree, which incorporates tremor quantified by computer-based scales derived from expert ratings (29).

Computerized spiral analysis has been employed for years in the clinical evaluation of movement disorders, including upper limb tremors in PD, essential tremor, dystonia, and cerebellar diseases. Traditionally, clinicians visually rate spirals using subjective severity scales (e.g., 0–4). While simple, such rating scales are examiner-dependent, prone to inter-rater variability (e.g., MDS-UPDRS), and insensitive to subtle changes. In contrast, CSA provides precise, objective quantifications of drawing control, spiral tightness, drawing speed, tremor frequency and amplitude, movement direction, and loop width variation. These parameters facilitate detailed characterization of motor deficits, disease progression, initial severity, treatment effects, and early PD detection (19).

Although the MDS-UPDRS demonstrates high reliability among trained movement disorder specialists, its reliability diminishes among general neurologists or trainees. Video-based assessments cannot capture complex motor control dynamics or clinical features such as rigidity, which are reflected in some spiral indices. Objective measures therefore support general neurologists and provide accurate data on challenging clinical features like rigidity. While CSA cannot replace clinical scales, it complements them by correlating well with MDS-UPDRS and offering quantitative metrics to enhance clinical evaluations. The MDS-UPDRS remains a widely used and validated measure of motor function in PD, serving as a reliable proxy for disease severity. Nevertheless, the MDS-UPDRS has limitations: its ordinal 5-point scale limits sensitivity to changes smaller than one point, and its non-ratio nature implies that score differences do not uniformly reflect clinical change magnitude. Although MDS-UPDRS reliability is well established in early untreated and advanced treated PD, it remains untested in early treated cases (29).

Several factors may explain the modest correlations between spiral analysis and MDS-UPDRS. Measurement errors and inconsistencies in spiral analysis could weaken associations. Additionally, MDS-UPDRS and SA may capture distinct but complementary aspects of PD pathology. Comparative studies with external benchmarks such as neuroimaging, clinical progression, treatment outcomes, and functional status are warranted. While MDS-UPDRS has undergone extensive validation, spiral analysis lacks similar external validation. Nonetheless, SA’s continuous, simple measurements could complement MDS-UPDRS in detecting subtle changes, especially in early PD stages when MDS-UPDRS scores remain low. Precise measurements aligned with MDS-UPDRS motor components could quantify mild disabilities and intermediate disease stages. Discrepancies between SA and MDS-UPDRS may reveal clinically important facets of PD severity not captured by clinical scales (29).

Therefore, combining spiral analysis with MDS-UPDRS application can enhance characterization and monitoring of Parkinson’s disease patients.

3.1.1 Review of Related Studies

Handwritten spiral drawings have become a widely investigated method for assessing motor symptoms in individuals with Parkinson’s disease (PD), offering a low-cost, accessible, and reproducible alternative to traditional clinical scales. These drawing-based tasks are particularly attractive in the context of inclusive healthcare strategies, where frequent in-person assessments by specialists may not be feasible. Recent advances in computational analysis and pattern recognition have further expanded the potential of spiral drawings to support diagnosis and monitoring across diverse clinical settings.

In line with the objectives of the present study—namely, to explore the viability of spiral-based assessments as a tool for broader PD follow-up—we conducted a review of related works. Table 3.1 presents the main characteristics of prior studies focused on detecting motor signs of PD using spiral or sketching tasks. The studies were organized according to several key dimensions: the number and type of participants (PDG and control groups), the number of spiral samples collected, the device or medium used (e.g., digital tablets or pen-and-paper), the specific motor tasks performed (such as the Drawing Spiral Task—DST, or the Spiral Sketching Task—SST), the machine learning methodologies applied, the comparison groups considered, and the performance metrics reported.

Across the literature, there is considerable variability in experimental protocols and technologies employed. While many studies use digital acquisition tools, such as graphics tablets and smart pens, others rely on traditional pen-and-paper drawings that are subsequently digitized for analysis. The classification methods range from classical machine learning algorithms (e.g., k-Nearest Neighbors, Support Vector Machines, and Random Forests) to deep learning models (e.g., CNNs, AlexNet, VGG, and ViT). Reported per-

formance metrics frequently include accuracy (ACC), area under the ROC curve (AUC), F1-score, and, in some cases, Matthews Correlation Coefficient (MCC), with several studies reporting accuracies above 90%.

However, most of these studies have been developed and validated under tightly controlled conditions or with relatively small and homogeneous datasets, thereby limiting their generalizability to broader clinical populations. Moreover, many works overlook the practicality of implementing such tools in real-world scenarios, particularly those constrained by limited resources or varying levels of technological infrastructure.

3.1.2 Machine learning for spiral analysis

In several healthcare applications, such as the identification of Parkinson’s disease (PD) and associated memory impairments, machine learning (ML) has emerged as a useful technique with great promise for the study of neurodegenerative diseases. With distinct capabilities in pattern recognition and prediction, ML comprises various methodologies, including supervised, unsupervised, deep, and ensemble learning. Supervised learning stands out because it uses labeled data to teach algorithms by reducing the difference between what is expected and what actually happens, allowing them to work well with new data they haven’t seen before. This approach includes both classification and regression tasks, where classification models are important because they give specific labels to input data, helping to sort samples into separate groups and aiding in decision-making for both yes/no and multiple category situations. At the same time, regression techniques are often used to understand how different factors relate to each other and to predict ongoing results, which helps in analyzing trends and making forecasts based on data (1).

Some of the most commonly used algorithms are tree-based methods like Decision Trees (DT), Random Forests (RF), and Gradient Boosting Trees (GBT), which are appreciated for being easy to understand, simple to use, and good at working with organized data by breaking it down into a series of decision steps. Another robust supervised approach is the Support Vector Machine (SVM), which excels in binary classification and can also be extended to regression and multi-class problems. By creating the best possible dividing line that keeps the data classes as far apart as possible using support vectors, which are the key data points near the decision boundary, SVMs enhance the accuracy of classification. Together, these diverse machine learning strategies establish a comprehensive framework for addressing complex healthcare challenges, especially those involving neurodegenerative diseases such as Parkinson’s disease (1).

One specific task that requires a high degree of coordination and motor control is spiral drawing, which has been recognized as a reliable measure of motor function. Parkinson’s disease, which affects various motor abilities such as speech, writing, gait, and coordination, is often evaluated using established rating tools like the Motion Rating Scale and its widely used subscale, the Unified Parkinson’s Disease Rating Scale (MDS-UPDRS-III).

Table 3.1 – Main characteristics of previous studies on Parkinson’s disease detection based on drawing tasks.

Reference	Year	Participants	Number of Spirals	Device / Material	Tasks	Methodology	Groups Compared	Performance Metrics
(35)	2017	27 PDG, 28 CG	55	Digital pen and tablet	DST	Multiclass ROC	PDG vs. CG; H&Y vs. CG	PDG vs. CG: ACC = 79.1%, AUC = 86.2% H&Y vs. CG: ACC = 68.2%, AUC = 83.2%
(18)	2019	37 PDG, 38 CG	75	Digital pen and tablet	DST	KNN, SVM (RBF/Linear), NB, LDA, RF, ADA	PDG vs. CG	ACC = 73.38%, AUC = 75.0%
(13)	2019	62 PDG, 15 CG	77	Public database (digital pen and tablet)	DST	CNN (spectrum)	PDG vs. CG	ACC = 96.5%, AUC = 99.2%
(4)	2020	27 PDG, 28 CG	55	Digital pen and tablet	SST	CNN	PDG vs. CG	ACC = 93.3%, F1 = 93.94%
(21)	2022	193 PDG, 142 CG	26,078	Public database (digital pen and tablet)	DST	AlexNet (fine-tuned)	PDG vs. CG	ACC = 99.22%
(20)	2020	25 PDG, 15 CG	40	Digital pen and tablet	SST	Logistic Regression	PDG vs. CG	ACC = 91.6%, AUC = 98.1%
(3)	2021	25 PDG, 15 CG	40	Digital pen and tablet	SST	DST, GoogLeNet, AlexNet	PDG vs. CG	ACC = 86%, F1 = 88%
(6)	2021	62 PDG, 15 CG	DST: 72, SST: 66	Digital pen and tablet	CMT, SST, DST	Random Forest	PDG vs. CG	SST: ACC = 97.6%, F1 = 98.5%, AUC = 99.9% DST: ACC = 96.4%, F1 = 97.7%, AUC = 99.6%
(12)	2021	56 PDG, 15 CG	71	Digital pen and tablet	SST, DST	KNN, DT, GA-KNN, GA-DT	PDG vs. CG	ACC = 100%
(17)	2023	15 PDG, 12 CG	102	Public database (pen and paper)	SST	VGG16, VGG19, ResNet18, ResNet50, ResNet101, ViT	PDG vs. CG	ACC = 87%, MCC = 74%
(25)	2024	15 PDG, 12 CG	102	Public database (pen and paper)	SST	Random Forest, KNN	PDG vs. CG	ACC = 86.67%, F1 = 87.5%

Acronyms: PDG = Parkinson’s Disease; CG = Control Group; DST = Drawing Spiral Task; SST = Spiral Sketching Task; CMT = Clock Drawing Task; ACC = Accuracy; AUC = Area Under the ROC Curve; F1 = F1-score; MCC = Matthews Correlation Coefficient; KNN = k-Nearest Neighbors; SVM = Support Vector Machine; NB = Naive Bayes; LDA = Linear Discriminant Analysis; RF = Random Forest; ADA = AdaBoost; DT = Decision Tree; GA = Genetic Algorithm; ViT = Vision Transformer.

Note: When two spiral counts are shown, the first refers to DST and the second to SST. Performance metrics are grouped by task and comparison group.

To assess the severity of PD, researchers have proposed numerous techniques aimed at identifying both motor deterioration and non-motor biomarkers. However, several challenges persist in the classification and care of PD, including the difficulty many caregivers face in bringing patients to clinical facilities and the dependence on trained professionals to perform physical examinations and provide diagnoses based on observational evaluations. These challenges are even greater when trying to detect the disease in its early stages, as invasive clinical procedures can be effective but are costly and carry significant risks, especially in developing countries (10).

Furthermore, there is no universally accepted objective standard for diagnosing PD, leading to a notable rate of misclassification, especially when evaluations are conducted by nonspecialists, with error rates reaching as high as 20%. While a detailed analysis of key symptoms—such as stiffness, bradykinesia, and tremors—can improve diagnostic accuracy, subjective clinical interpretation may introduce bias. Consequently, the development of decision support systems in medicine is gaining attention for their potential to enhance objectivity and accelerate early diagnosis. Early detection is crucial not only for initiating timely interventions but also for enabling the development of individualized treatment strategies. In this context, the search for reliable biomarkers continues to be a central aim in neurodegenerative disease research. Many studies have looked at speech, especially everyday talking and long vowel sounds, as safe ways to help diagnose PD, and analyzing movement and walking has also been helpful in spotting and tracking motor symptoms (10).

Beyond speech and movement analysis, researchers have developed several techniques to examine handwriting in PD patients, investigating both static and dynamic features such as writing speed and the progressive decline in pen pressure. However, factors like vision, language skills, and writing style influence individual variability, complicating handwriting analysis. As a result, alternative tasks such as drawing—especially spiral illustrations—have gained prominence for being less affected by these personal differences. At the same time, deep learning (DL) has changed the way we analyze medical and biomedical images, showing great success in identifying complex details and improving how accurately we can classify images for tasks like segmentation, detection, and diagnosis. In particular, convolutional neural networks (CNNs) have been crucial in this change because they are very good at generalizing and performing well in many medical image classification tasks, which has significantly improved the accuracy and usefulness of automated PD assessment (10).

Spiral drawing protocol: classification of handwritten motor patterns in Parkinson's disease

4.1 Approach

This section details the development of a database comprising spiral drawings executed with pen and paper, aimed at investigating their utility in detecting Parkinson's disease (PD). The spirals were digitized, segmented, standardized, and individually labeled to construct the dataset. Participants were initially categorized into two groups: the control group (CG), consisting of individuals without a PD diagnosis, and the experimental group (PD), comprising diagnosed patients.

According to the Hoehn & Yahr (H&Y) classification employed in Part III of the MDS-UPDRS scale, the PD group was further subdivided into stages 1, 2, and 3. Classification techniques were applied to identify patterns within the drawings and categorize them according to these predefined groups.

Initially, a binary classification (CG vs. PD) was performed. Subsequently, a multiclass classification (CG vs. H&Y stages 1, 2, and 3) was conducted to evaluate the potential of neural networks in detecting disease progression patterns. Additional analyses were performed based on specific motor variables from Part III of the MDS-UPDRS, exploring the feasibility of spiral drawings as a tool for longitudinal monitoring. It is noteworthy that data collection spanned 16 sessions.

Furthermore, analyses were conducted based on the specific motor variables of the MDS-UPDRS Part III scale obtained during data collection. The analyses aimed to identify which attributes of the spiral drawings each neural network evaluated, correlating these evaluations with the clinical parameters of the aforementioned scale. This approach made it possible to verify the relationship between the characteristics extracted by the neural networks and the motor signs evaluated in the MDS-UPDRS, enriching the interpretation of the results.

For image analysis, feature extraction was performed using pre-trained convolutional

neural networks (CNNs) including ResNet-50, VGG16, Xception, and MobileNetV2. Extracted features underwent comprehensive analysis, and model performance was assessed using metrics such as accuracy, recall, and F1-score. Cross-validation was employed to ensure result robustness.

Model implementations were primarily developed in Python (version 3.9.12) within the Visual Studio Code (VS Code) environment, utilizing the SciPy library (version 1.7.3) for statistical analyses. Additionally, MATLAB was used with the *Deep Learning Toolbox* to train and evaluate the selected deep neural network models.

4.2 Research methodology

This observational, cross-sectional study with a control group consists of three main components: data collection, data preprocessing, and data processing using machine learning techniques. Figure 4.1 illustrates the methodological pipeline employed to automatically classify the stages of Parkinson’s disease based on spiral drawings.

Data collection involved participants producing spiral drawings using pen and paper. These drawings were subsequently scanned, segmented, and labeled. The resulting images were organized in the HEMATO database and subjected to a preprocessing stage, which included resizing to $256 \times 256 \times 3$ pixels and normalization of pixel intensity values to a range between 0 and 1.

The preprocessed images were input into various convolutional neural network (CNN) architectures—MobileNetV2, VGG16, Xception, and ResNet50—for feature extraction. The resulting feature vectors were passed through a GlobalMaxPooling2D layer, followed by two fully connected (dense) layers comprising 256 and 128 neurons, respectively, both utilizing the ReLU activation function.

Finally, classification was performed using a Softmax output layer to assign participants to one of four categories: a control group or stages 1, 2, or 3 of the Hoehn & Yahr scale, depending on whether the individual had been diagnosed with Parkinson’s disease.

4.3 Data collection protocol

This **longitudinal, observational** study was approved by the **Research Ethics Committee (CEP)** of the Federal University of Uberlândia (Brazil) under process number **69064623.0.0000.5152**. Data collection was conducted between October 2023 and August 2024. A total of 40 individuals, aged between 49 and 84 years, participated in the study. The sample consisted of 15 healthy individuals in the control group (CG: 7 males and 8 females) and 25 individuals diagnosed with Parkinson’s disease in the PD group (PD: 12 males and 13 females), as detailed in Table 4.1.

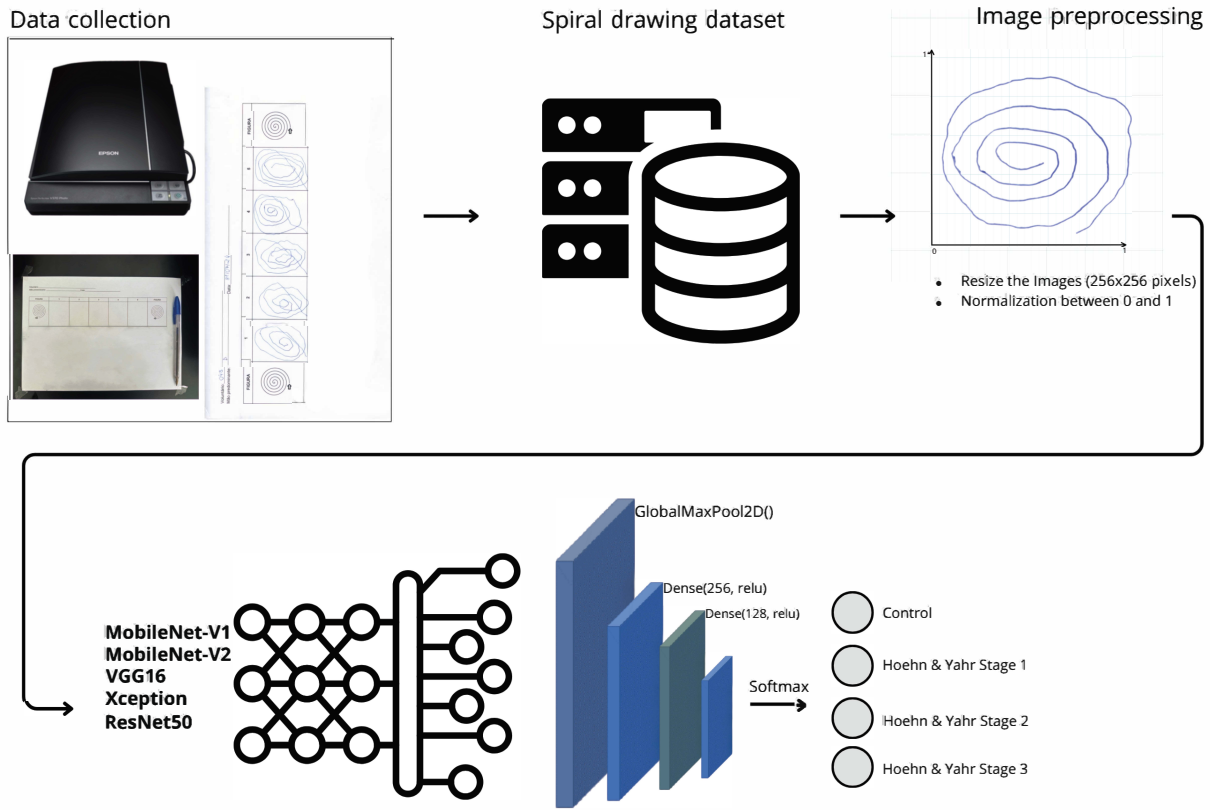


Figure 4.1 – Research methodology.

To ensure demographic comparability between the groups, appropriate statistical tests were applied. The normality of the age variable was assessed using the Shapiro–Wilk test. Since the data did not follow a normal distribution ($p < 0.05$), the **nonparametric** Mann–Whitney U test was used to compare age distributions, confirming no significant difference between the groups ($U = 183.50, p = 0.92$). For the categorical variable **sex**, Fisher’s exact test was employed, as it is particularly suitable for analyzing contingency tables with small sample sizes. This test also indicated no significant difference in sex distribution between the groups ($p = 1.00$), suggesting that the samples were well matched.

The control group (CG) had a mean age of **62.7 (SD 8.6)** years, while the Parkinson’s disease group (PD) had a mean age of **63.7 (SD 6.9)** years. Participants in the PD group were clinically staged according to the Hoehn and Yahr (H&Y) scale, an ordinal scale ranging from 0 to 5, with a mean disease stage of **1.72 ± 0.74**. All control group participants were classified as stage 0. Clinical motor assessments were conducted using the Movement Disorder Society–Unified Parkinson’s Disease Rating Scale (MDS-UPDRS) Part III, where the maximum possible score is 64 points. The mean scores were 0 for the control group and **15.50 ± 10.47** for the PD group.

Each participant completed multiple test sessions, during which they were instructed to perform five sequential spiral drawings from left to right on a sheet of paper without interruption. A total of **1,412** spiral images were collected from the control group and **1,834** from the PD group. The average number of spiral drawings per participant was

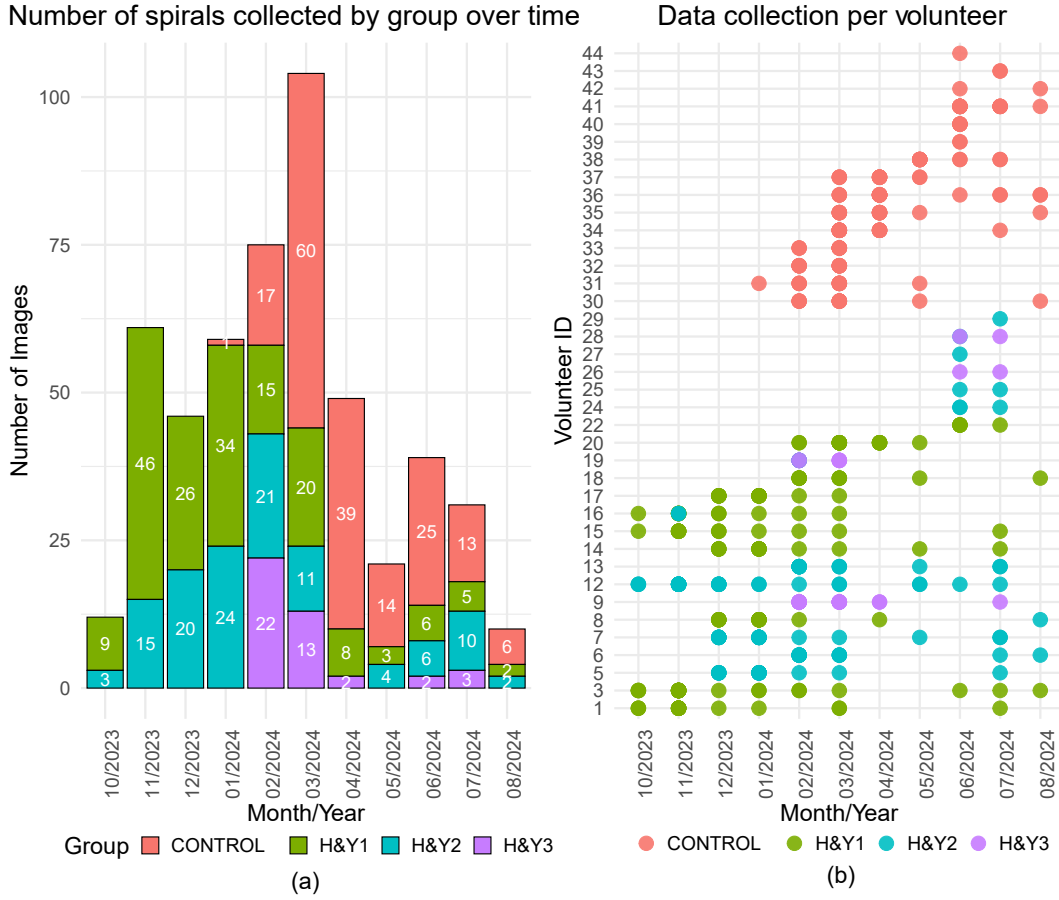


Figure 4.2 – Overview of the longitudinal data collection, illustrated at both the group and individual levels. **(a)** Monthly number of collected spiral drawings, with separate bars indicating the respective groups. **(b)** Timeline of data collection for each participant, showing the month and year of participation. Note that although some participants were assessed multiple times in a single month, only one entry per month is shown in the graph.

60.1 ± 39.0 in the control group and 75.4 ± 53.4 in the PD group. A statistical comparison of the number of samples between groups, using the Mann–Whitney U test, revealed no significant difference ($p = 0.43$).

In comparison to existing datasets in the literature, this study offers a distinctive longitudinal dataset with repeated recordings collected over nearly twelve months, thereby enabling the analysis of intra-subject variability over time.

4.4 Scanning procedure

After data acquisition, all collected drawings were digitized using an Epson Perfection V370 scanner to ensure image standardization. CorelDRAW software was employed to adjust and isolate each spiral, producing standardized individual images. Minor distortions were corrected during editing without altering the original characteristics of the drawings as in the Figure 4.3. These processed images were subsequently used for analysis via machine learning techniques.

Table 4.1 – Distribution of spiral drawing samples and clinical characterisation by group.

Control Group (CG)							Parkinson’s Disease Group (PD)						
ID	<i>n</i>	Sex	H&Y	Hand	UPDRS-III	Age	ID	<i>n</i>	Sex	H&Y	Hand	UPDRS-III	Age
V30	75	F	0	Right	0	62	V01	100	F	1	Right	10	62
V31	86	M	0	Right	0	56	V03	180	M	1	Right	12	56
V32	75	F	0	Right	0	55	V04	110	F	1	Right	7	58
V33	65	F	0	Right	0	67	V05	105	M	2	Right	26	68
V34	110	M	0	Right	0	59	V06	85	F	2	Right	13	62
V35	90	M	0	Right	0	66	V07	115	M	2	Right	9	58
V36	125	F	0	Right	0	55	V08	100	M	1	Right	15	64
V37	65	F	0	Right	0	67	V09	84	F	3	Right	14	67
V38	70	M	0	Right	0	72	V12	175	F	2	Right	18	65
V39	10	F	0	Right	0	65	V13	75	F	2	Right	28	84
V40	30	F	0	Right	0	53	V14	85	F	1	Right	10	63
V41	76	M	0	Right	0	62	V15	85	F	1	Right	5	63
V42	5	M	0	Right	0	61	V16	95	F	1	Right	1	64
V43	15	M	0	Right	0	84	V17	110	M	1	Right	13	68
V44	5	M	0	Right	0	84	V18	55	M	1	Right	6	52
— no more participants —							V19	80	M	3	Right	47	59
							V20	95	F	1	Right	2	49
							V22	30	M	1	Right	12	68
							V23	10	F	2	Right	2	68
							V24	15	M	2	Right	18	59
							V25	10	F	2	Right	26	68
							V26	10	M	3	Right	21	67
							V27	5	M	2	Right	27	62
							V28	15	M	3	Right	14	77
							V29	5	F	2	Right	23	53
Mean \pm SD: 62.7 \pm 8.6 years (Total <i>n</i> = 1412)							Mean \pm SD: 63.7 \pm 6.9 years (Total <i>n</i> = 1834)						
							H&Y: 1.72 \pm 0.74 UPDRS-III: 15.50 \pm 10.47						

F: Female; M: Male; SD: Standard deviation; H&Y: Hoehn & Yahr scale; *n*: Number of spiral drawings

PD: Parkinson’s Disease group; CG: Control group

UPDRS-III: Unified Parkinson’s Disease Rating Scale Part III (Motor Examination)

4.5 Data storage and dataset

The spiral images were stored in folders labeled according to the groups to which participants were originally assigned. The MDS-UPDRS assessment provided Hoehn & Yahr (H&Y) classifications, restricting the Parkinson’s disease (PD) group to stages 1, 2, and 3. All images were standardized in .jpeg format to ensure compatibility with the analysis software and subsequently organized.

The complete dataset was stored in a digital repository and made freely available, as detailed in Appendix 1 and exemplified, as in Image 4.4. Representative images are presented in Figure 4.6 for the PD group and Figure 4.5 for the control group (CG).

CoreIDRAW software was utilized to isolate each spiral individually, enabling separation of spirals into distinct images in a standardized manner. Each spiral was resized to a width of 1040 pixels and a height of 850 pixels and saved as an individual Joint Photographic Experts Group (JPG) file for subsequent analysis using machine learning techniques. The files were named according to a 24-character string following this con-

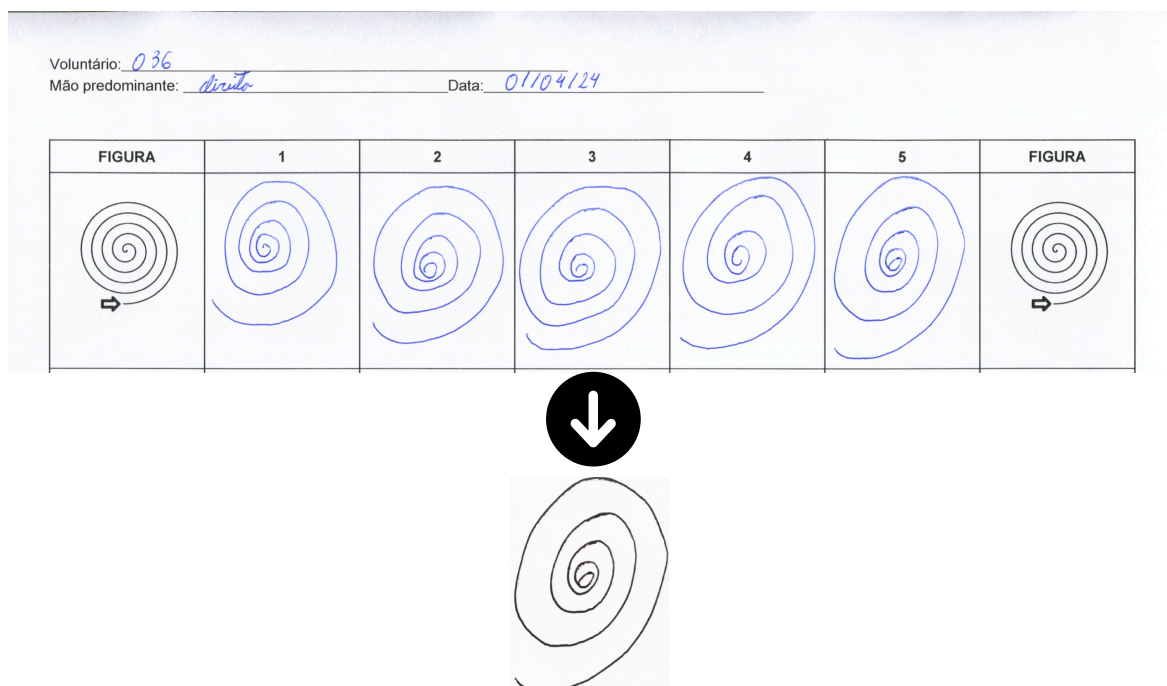


Figure 4.3 – Example of an image corrected and individualized during editing without altering the original characteristics of the original drawings.

vention:

- **V** — Initial marker for volunteer identification.
- **XX** — Two-digit volunteer identifier.
- **YY** — Two-character code for group identification (**GC** for Control, **PD** for Parkinson's disease).
- **F** — Initial marker for data collection instrument (sheet) identification.
- **ZZZZZZZ** — Seven-digit identifier of the data collection instrument.
- **TYP** — Three-character code for drawing type (**ESP** denotes spiral).
- **Q** — Single-digit spiral identifier (1 to 5).
- **COL** — Initial marker for session identification.
- **MM** — Two-digit session number.
- **LL** — Two-character code for the hand used in drawing (**MD**: right hand; **ME**: left hand).

For instance, the file name **V30GCF0000132ESP1COL12MD.jpg** decodes as: **V XX YY F ZZZZZZZ TYP Q COL MM LL** or **V 30 GC F 0000132 ESP 1 COL 12 MD**.

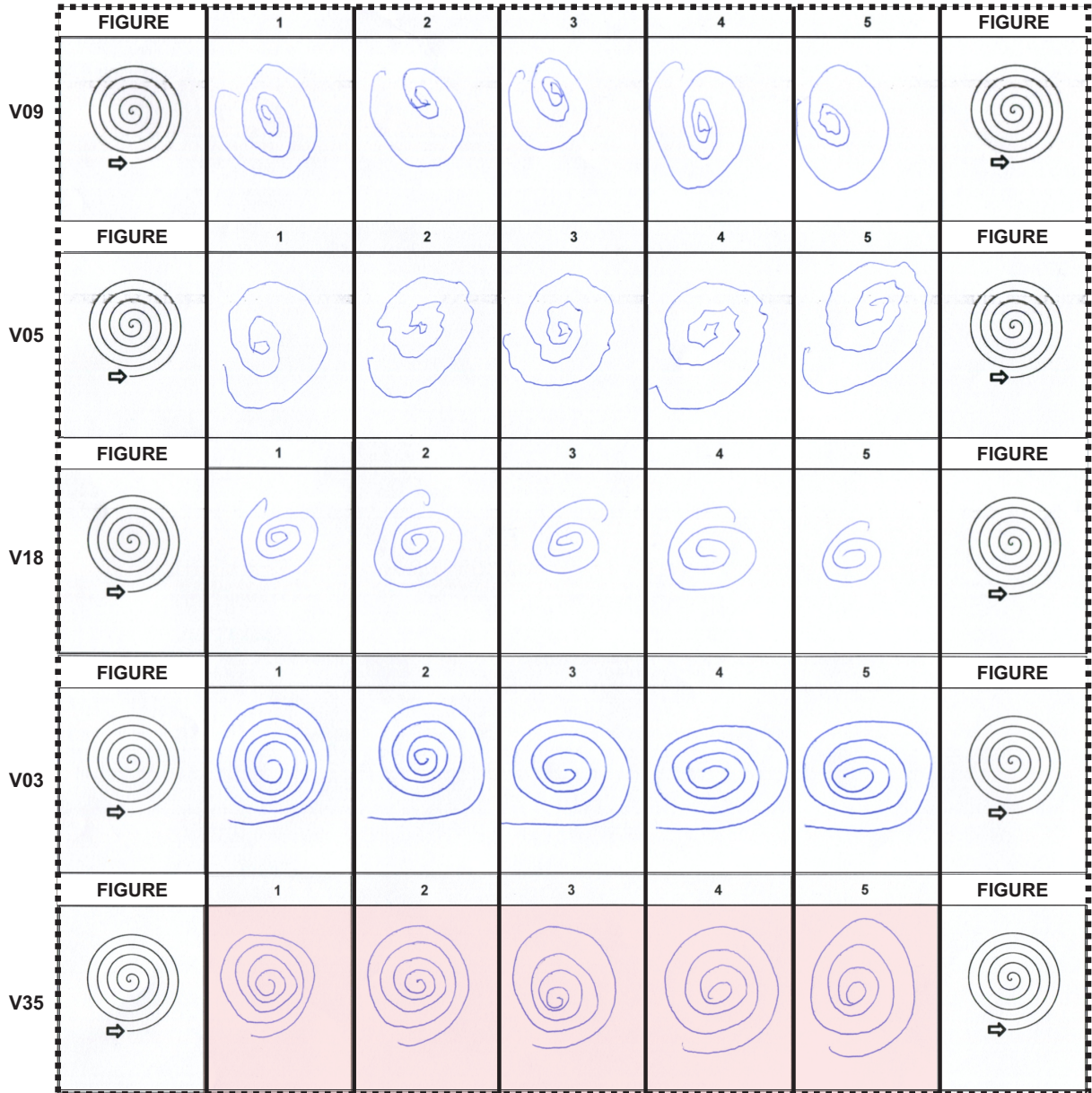


Figure 4.4 – Samples of the collected dataset. Spirals from individuals with PD (V09, V05, V18, and V03) and a subject of the control group (V35—highlighted) are shown. The spirals were digitised from the actual instrument used for data collection. The arrow in the ideal spiral template indicates the intended starting point of the drawing. Note that some subjects had difficulty following the ideal spiral template strictly.

This filename convention facilitates data organization and automates the data analysis workflow. Comparison of spiral drawings produced by the CG and PD groups reveals significant differences in patterns and graphical characteristics. Spirals from the CG typically exhibit smooth, continuous, and well-centred strokes with regular spacing between turns and clear motor control, as illustrated in Figure 4.5. These features reflect stability, coordination, and precision in fine motor movements.

In contrast, spirals produced by the PD group display irregular features, including noticeable variations in line thickness, deviations in the drawing path, and evident tremors,

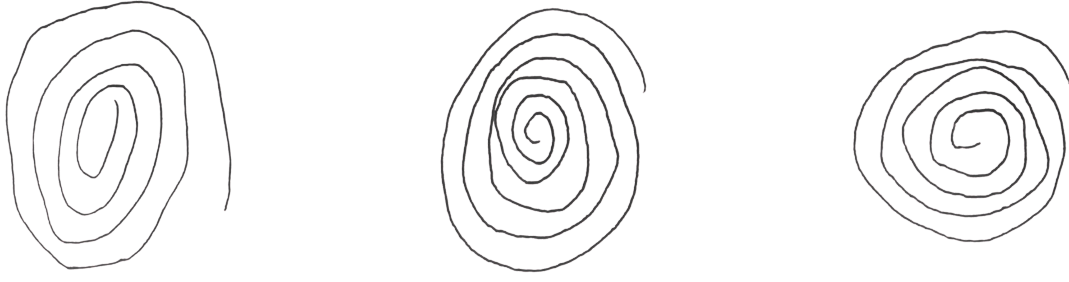


Figure 4.5 – Data acquired for CG volunteers.



Figure 4.6 – Data acquired for PD volunteers.

especially in curves and turns, as observed in Figure 4.5. Many drawings exhibit overlapping lines and misaligned or hesitant strokes, indicating difficulty executing movements fluidly and with control. Additionally, micro-interruptions or rhythmic oscillations consistent with resting tremors typical of Parkinson's disease are often observed.

These differences suggest impaired motor function in the PD group, characterized by deficits in visuomotor coordination, accuracy, and stroke control, unlike the CG group, which demonstrates smooth and coordinated movements. Thus, the graphical analysis of the spirals visually and qualitatively highlights the impact of Parkinson's disease on fine motor skills.

4.6 Data processing

4.6.1 Feature extraction

Feature extraction refers to the process of deriving meaningful information from raw data. In this study, we exclusively employed deep feature extraction using pre-trained models.

Feature extraction using pre-trained models

In addition to our proposed method, which utilizes autonomous feature extraction, we investigated the application of pre-trained deep neural network models for feature extraction, inspired by the approach of DeSipio et al. (9).

Although spiral images are not included in the ImageNet database, we adopted the methodology suggested by DeSipio et al. (9) in the conclusions of their work, which suggested the potential use of neural networks for the autonomous extraction of attributes from spiral images, without defining which types of data should be extracted. In the case of the work conducted by DeSipio et al. (9), VGG16 and ResNet50 were used, but in preliminary analyses, we opted for the additional analysis of MobileNetV2 and Xception, given the potential results reported in the literature by Chollet et al. and Howard et al., respectively (16), (7). Future studies could explore fine-tuning these pre-trained models specifically for spiral drawings.

In this study, feature extraction was performed using MobileNetV2, VGG16, ResNet-50, and Xception. The raw images were resized to $224 \times 224 \times 3$ pixels for MobileNetV2, VGG16, and ResNet-50, whereas images for the Xception model were resized to $299 \times 299 \times 3$ pixels to conform to the model's input requirements.

Following the dataset preparation procedures, the neural networks autonomously performed feature extraction in the layers listed below, utilizing the labels assigned according to the Hoehn and Yahr scale collected during the MDS-UPDRS assessment, and classified individuals as either affected or unaffected by Parkinson's disease, thereby simulating a supervised machine learning process.

- **VGG16:** fc7
- **ResNet50:** avg_pool
- **Xception:** block14_sepconv2_act
- **MobileNetV2:** global_average_pooling2d_1
- **MobileNetV1:** global_average_pooling2d

The models were loaded with pre-trained weights, and their fully connected layers were removed. Each image was then passed through the modified models to obtain output feature maps from the final convolutional layer. VGG16 generated 512 feature maps of size 14×14 per image, while ResNet-50 produced 2048 feature maps of size 7×7 . These feature maps were subsequently used as inputs for classification tasks.

Xception

Xception is a deep convolutional neural network introduced by Google in 2017, known for its novel approach to feature learning. It uses a different kind of convolution called

depthwise separable convolutions, which first applies a depthwise spatial convolution and then a pointwise (1×1) convolution to create a new set of features. This design significantly reduces the number of parameters while maintaining high performance. The architecture consists of 14 modules with 36 convolutional layers, most of which include linear residual connections. Xception achieved a top-5 classification accuracy of 95.5% on the ImageNet dataset (7).

MobileNetV2

MobileNet is a family of efficient convolutional neural networks specifically designed for mobile and edge devices. Originally introduced by Google in 2017, MobileNets emphasize computational efficiency while maintaining competitive performance. These networks utilize depthwise separable convolutions to significantly reduce the number of parameters and computational cost. Multiple iterations of the architecture have been developed, including MobileNetV1, MobileNetV2, and MobileNetV3, each incorporating architectural enhancements and optimization techniques. MobileNetV1 achieves a top-5 classification accuracy of 89.9% on the ImageNet dataset (16).

VGG16

VGG16 is a deep convolutional neural network developed by the Visual Geometry Group at the University of Oxford. Renowned for its simplicity and consistent architectural design, VGG16 achieved a top-5 accuracy of 92.7% in the 2014 ImageNet Large Scale Visual Recognition Challenge (ILSVRC). The network comprises 13 convolutional layers with 3×3 filters, followed by 3 fully connected layers. Although computationally demanding, VGG16 has had a significant impact on the development of deeper neural networks due to its ability to effectively capture fine-grained visual features(31).

ResNet50

ResNet-50 is a deep convolutional neural network architecture introduced by Microsoft Research in 2015. It utilizes residual blocks with skip connections, which facilitate gradient flow during training and help mitigate the vanishing gradient problem. Comprising 50 layers, ResNet-50 enables the effective training of very deep networks, resulting in improved performance. The model achieved a top-5 classification accuracy of 92.2% on the ImageNet dataset (15).

4.7 Analysis methods

4.7.1 Classification models

Classification, a machine learning and statistical technique, assigns data to predefined classes or categories. A classifier is a mapping function that learns from input data and predicts the category to which new data belongs. Classification can be broadly divided into three types: binary, multi-class, and multi-label. In binary classification, each input is assigned one of two possible outcomes (e.g., true or false). In multi-class classification, more than two classes exist, but each input is associated with only one class. In multi-label classification, each input may simultaneously belong to multiple classes (32).

We initially employed the same deep neural networks (DNNs) used for feature extraction to classify the data, allowing for methodological consistency and enabling direct performance comparison with traditional methods (23). For multi-class problems, we utilized the Error-Correcting Output Codes (ECOC) method, a common technique that combines multiple binary classifiers to address classification tasks involving more than two classes. Two distinct classification tasks were performed using spiral images.

The first task aimed to discriminate individuals with Parkinson’s disease (PD) from healthy controls. The second task classified participants according to the Hoehn & Yahr (H&Y) scale, which categorizes PD participants into stages 1, 2, or 3, alongside the control group. Model performance was evaluated using standard classification metrics—accuracy, precision, recall, and F1-score—providing a comprehensive assessment of classification quality.

Further classification strategies, leveraging the same feature extraction framework, were explored using spiral images. Initially, features related to tremor from the Unified Parkinson’s Disease Rating Scale (MDS-UPDRS) were used. Subsequent classifications incorporated hand movement-related features, specifically the mean and median scores, followed by other motor-related indicators reported in Part III of the MDS-UPDRS. These labels were assigned by a trained specialist during test sessions, aiming to enhance discrimination between PD patients and healthy controls.

Additionally, the H&Y-based classification was reinforced through data collected across 16 assessment sessions, providing a clinically grounded framework for categorizing disease progression. This approach improved the model’s ability to classify spiral drawings according to PD severity (stages 1, 2, or 3), thereby capturing varying levels of motor impairment.

All analyses focus exclusively on the test phase outcomes, with training performance included only for comparative purposes against test results in the context of evaluating the potential application of spiral drawings for monitoring Parkinson’s disease. The data split was standardised across all experiments, allocating 70% of the samples for training and 30% for testing. This approach ensured methodological consistency and optimised

performance, akin to the strategy utilised by Zham et al (35).

4.8 Statistical analysis

Classic evaluation metrics, such as accuracy and precision, are widely used to assess the predictive performance of machine learning algorithms, particularly in classification tasks. Accuracy is defined as the proportion of correctly classified samples, representing an overall measure of performance that considers all classes and their respective sample counts. Mathematically, for binary or multiclass problems, accuracy is calculated as the ratio of correctly classified examples to the total number of examples available for classification. However, this metric has limitations in scenarios involving class imbalance—that is, when one or more classes have substantially more samples than others. In such cases, accuracy may obscure the true performance of the model on underrepresented classes (8).

To address this limitation, the macro F1-score is employed, providing a more balanced evaluation of model performance in multiclass and imbalanced classification scenarios. Unlike accuracy, the macro F1-score is computed as the harmonic mean of precision and recall for each class, followed by averaging across all classes, thus assigning equal weight to each. This approach prevents dominant classes from disproportionately influencing the overall score, offering a fairer assessment of model performance (8).

It is important to note that beyond these quantitative metrics, other factors such as model interpretability, robustness, and usability should also be considered—especially for real-world applications. In this study, the workflow began with data preprocessing, followed by transformation of data into numerical feature vectors and subsequent application of classification algorithms. Model effectiveness was ultimately evaluated using accuracy and macro F1-score (8).

For clarity, the following terms are defined:

- **TP (True Positives):** Positive cases correctly identified by the model.
- **TN (True Negatives):** Negative cases correctly identified.
- **FP (False Positives):** Negative cases incorrectly classified as positive.
- **FN (False Negatives):** Positive cases incorrectly classified as negative.
- **class (Classes):** categories into which each instance (or sample) in the dataset can be assigned.

Cohen's Kappa coefficient

The kappa statistic is commonly employed to assess inter-rater reliability. Evaluator reliability is crucial as it indicates the degree to which the data gathered in the study

accurately reflect the measured variables. Inter-rater reliability refers to the degree to which data collectors (raters) provide consistent scores for the same variable. Various methods have been employed to assess inter-rater reliability, with the traditional approach being the calculation of percentage agreement, determined by the ratio of agreement scores to the total number of scores. This index assesses the accuracy of a machine learning model in multiclass classification, specifically addressing the issue of class imbalance in multi-category predictions. This is particularly beneficial in cases of class imbalance (34).

$$\kappa = \frac{p_o - p_e}{1 - p_e} = \frac{\frac{TP + TN}{T} - \left(\frac{TP + FP}{T} \cdot \frac{TP + FN}{T} + \frac{FN + TN}{T} \cdot \frac{FP + TN}{T} \right)}{1 - \left(\frac{TP + FP}{T} \cdot \frac{TP + FN}{T} + \frac{FN + TN}{T} \cdot \frac{FP + TN}{T} \right)}. \quad (4.1)$$

- p_o is the observed proportion of agreement between raters,
- p_e is the expected proportion of agreement by chance.
- T = total of observations ($T = TP + TN + FP + FN$).

Accuracy

Accuracy measures the proportion of correct predictions relative to the total number of examples:

$$\text{accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

F1-score

The F1-score is the harmonic mean of Precision and Recall, seeking a balance between both metrics.

Precision

Precision answers the question: “Of all instances predicted as positive, how many were actually positive?”

$$\text{Precision} = \frac{TP}{TP + FP}$$

Recall (Sensitivity)

Recall answers the question: “Of all actual positive cases, how many did the model correctly identify?”

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

F1-score (Binary Classification)

The general formula of the F1-score is:

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

An equivalent expression in terms of TP, FP, and FN is:

$$\text{F1-score} = \frac{\text{TP}}{\text{TP} + 0.5 \times (\text{FP} + \text{FN})}$$

F1-score (Multiclass Classification)

For multiclass classification problems, the F1-score is computed individually for each class using a One-vs-Rest approach, and then aggregated using various averaging strategies:

Micro F1-score Calculated by aggregating True Positives, False Positives, and False Negatives across all classes before computing the F1-score. In single-label multiclass classification, the micro F1-score is equivalent to accuracy.

$$\text{micro F1-score} = \frac{\sum \text{TP}_{\text{classes}}}{\sum \text{TP}_{\text{classes}} + 0.5 \times (\sum \text{FP}_{\text{classes}} + \sum \text{FN}_{\text{classes}})}$$

Macro F1-score This is the unweighted arithmetic mean of the F1-scores computed for each class individually. It treats all classes equally, regardless of their frequency.

$$\text{macro F1-score} = \frac{\text{F1}_{\text{Class A}} + \text{F1}_{\text{Class B}} + \text{F1}_{\text{Class C}} + \dots}{\text{Number of Classes}}$$

Weighted F1-score This is the average of the class-wise F1-scores weighted by the number of true instances (support) for each class.

$$\text{Weighted F1-score} = \frac{(\text{F1}_{\text{Class A}} \times \text{Support}_{\text{Class A}}) + (\text{F1}_{\text{Class B}} \times \text{Support}_{\text{Class B}}) + \dots}{\text{Total Support}}$$

4.9 Results

4.9.1 Binary and multiclass classification analysis

Figures 4.14 and 4.15, as well as Tables 4.2 and 4.3, present the accuracy and F1-score results for both the training and test datasets. It is important to emphasize that this

study utilized an original dataset. The primary objective was to distinguish between the control group (individuals without Parkinson’s disease) and the PD group. Subsequently, differentiation among the Hoehn & Yahr scale levels was performed exclusively within the PD group.

Table 4.2 – Overall Accuracy, Cohen’s Kappa, Precision, Recall, and F1-score by class (CG vs PD).

Metrics	Mobile NetV2	VGG16	Xception	Res Net50
Accuracy	91.45%	92.64%	94.06%	94.06%
Cohen’s Kappa	0.82	0.84	0.87	0.86
Class: Control				
Precision	90.32%	91.02%	91.42%	90.51%
Recall	82.35%	85.66%	90.07%	91.18%
F1-score	86.15%	88.26%	90.74%	90.84%
Class: PDG				
Precision	91.92%	93.34%	95.30%	95.77%
Recall	95.79%	95.96%	95.96%	95.44%
F1-score	93.81%	94.64%	95.63%	95.61%

4.9.2 Classification analysis using MDS-UPDRS

The F1-score results for the training and test datasets are presented in Table ?? and figure 4.14. It is important to note that a unique dataset was specifically created for this study. Initially, individual F1-score results were reported, followed by the calculation of the mean F1-score to facilitate comparisons.

Table 4.3 – Overall Accuracy, Cohen’s Kappa, Precision, Recall, and F1-score by class (CG vs H&Y1-3).

Metric	MobilenetV2	VGG16	ResNet50	Xception
Overall Accuracy	80.29	78.03	81.95	81.00
Cohen’s Kappa	0.72	0.69	0.74	0.73
CONTROL				
Precision	88.69	88.26	90.22	86.29
Recall	89.34	85.66	91.54	94.85
F1-score	89.01	86.94	90.88	90.37
Stage H&Y1				
Precision	78.41	77.02	79.38	79.56
Recall	77.43	74.61	79.62	79.31
F1-score	77.92	75.80	79.50	79.43
Stage H&Y2				
Precision	78.31	78.89	82.16	83.73
Recall	82.22	78.89	84.44	77.22
F1-score	80.22	78.89	83.29	80.35
Stage H&Y3				
Precision	59.38	49.44	57.38	54.24
Recall	53.52	61.97	49.30	45.07
F1-score	56.30	55.00	53.03	49.23

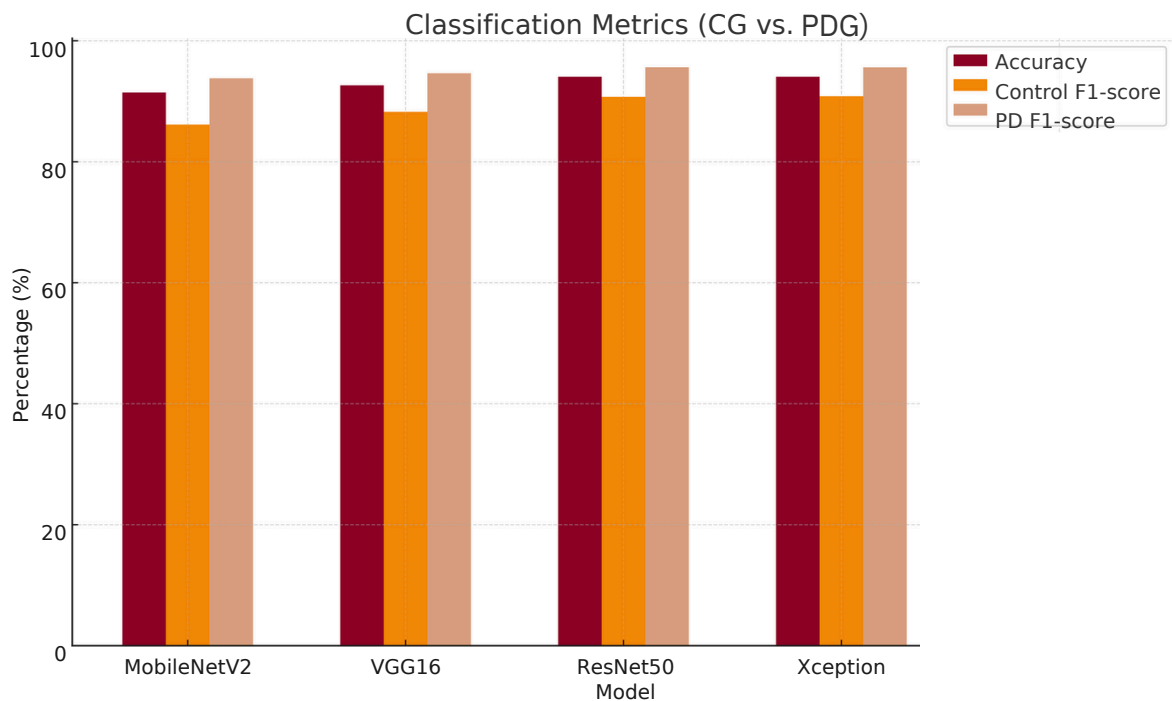


Figure 4.7 – Performance statistics of different CNN models for classification between Parkinson’s Disease (PD) and Healthy Controls (HC).

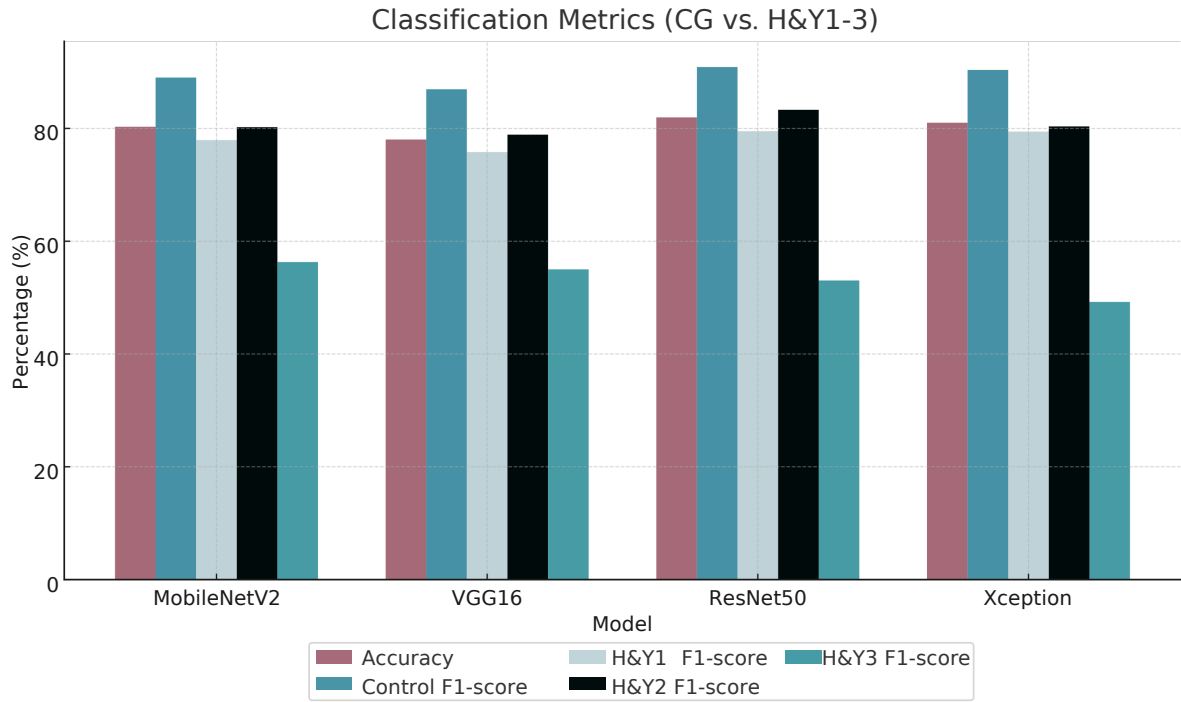


Figure 4.8 – Performance statistics of different CNN models for classification between Parkinson’s Disease patients in Hoehn & Yahr stages 1–3 and Healthy Controls (HC).

Table 4.4 – Average mean F1-score by classification using VGG16, Xception, MobileNetV2 and ResNet50 comparing different features.

Task	MobileNetV2	Xception	ResNet50	VGG16
Speech (3.1)	0.48182	0.62050	0.49824	0.41864
Facial Expression (3.2)	0.60660	0.60343	0.60601	0.56097
Rigidity (3.3)	0.64400	0.62772	0.69136	0.63352
Pinch (3.4)	0.42180	0.44052	0.44573	0.50096
Hand Movements (3.5)	0.59610	0.63020	0.59910	0.48680
Pronation-Supination (3.6)	0.55158	0.59133	0.55710	0.48984
Toe Tapping (3.7)	0.52711	0.52711	0.44171	0.44615
Leg Agility (3.8)	0.39921	0.56913	0.43473	0.44899
Chair Rising (3.9)	0.50490	0.53064	0.51935	0.50669
Gait (3.10)	0.58217	0.58218	0.58857	0.56899
Freezing (3.11)	0.45106	0.80056	0.79977	0.76701
Postural Stability (3.12)	0.53043	0.56284	0.58682	0.51287
Posture (3.13)	0.54912	0.67483	0.71121	0.62105
Body Bradykinesia (3.14)	0.54102	0.47689	0.64792	0.64792
Rest Tremor in Hands (3.15)	0.62023	0.65826	0.71952	0.60034
Action Tremor in Hands (3.16)	0.56271	0.69148	0.74859	0.62446
Tremor Amplitude (3.17)	0.60273	0.68358	0.67691	0.58286
Tremor Persistence (3.18)	0.54045	0.60593	0.73009	0.55610

Note: The numbers in parentheses indicate the corresponding item codes of Part III of the MDS-UPDRS (Motor Examination).

Table 4.5 – Average mean precision by classification using VGG16, Xception, MobileNetV2, and ResNet50 comparing different features.

Task	MobileNetV2	Xception	ResNet50	VGG16
Speech (3.1)	0.47124	0.62050	0.67467	0.41739
Facial Expression (3.2)	0.60985	0.60343	0.60601	0.56097
Rigidity (3.3)	0.67121	0.62772	0.71561	0.69667
Pinch (3.4)	0.44573	0.44052	0.44573	0.50096
Hand Movements (3.5)	0.59910	0.63020	0.59910	0.48680
Pronation-Supination (3.6)	0.51908	0.59133	0.55710	0.48984
Toe Tapping (3.7)	0.46769	0.46769	0.40728	0.44331
Leg Agility (3.8)	0.56913	0.56913	0.43473	0.44025
Chair Rising (3.9)	0.53064	0.53064	0.51935	0.51191
Gait (3.10)	0.58218	0.58218	0.58857	0.56899
Freezing (3.11)	0.80056	0.80056	0.79977	0.76701
Postural Stability (3.12)	0.56284	0.56284	0.58682	0.51287
Posture (3.13)	0.67483	0.67483	0.71121	0.62105
Body Bradykinesia (3.14)	0.47689	0.47689	0.64792	0.64792
Rest Tremor in Hands (3.15)	0.65826	0.65826	0.71952	0.60034
Action Tremor in Hands (3.16)	0.69148	0.69148	0.74859	0.62446
Tremor Amplitude (3.17)	0.68358	0.68358	0.67691	0.58286
Tremor Persistence (3.18)	0.60593	0.60593	0.73009	0.55610

Note: The numbers in parentheses indicate the corresponding item codes of Part III of the MDS-UPDRS (Motor Examination).

Table 4.6 – Average mean recall by classification using VGG16, Xception, MobileNetV2, and ResNet50 comparing different features.

Task	MobileNetV2	Xception	ResNet50	VGG16
Speech (3.1)	0.50187	0.50187	0.48426	0.42584
Facial Expression (3.2)	0.60343	0.60343	0.60601	0.56097
Rigidity (3.3)	0.62772	0.62772	0.69136	0.63352
Pinch (3.4)	0.44052	0.44052	0.49463	0.50096
Hand Movements (3.5)	0.63020	0.63020	0.57000	0.48680
Pronation-Supination (3.6)	0.59133	0.59133	0.55947	0.48984
Toe Tapping (3.7)	0.52711	0.52711	0.44171	0.44615
Leg Agility (3.8)	0.39921	0.39921	0.46363	0.44899
Chair Rising (3.9)	0.50490	0.50490	0.51192	0.50669
Gait (3.10)	0.58218	0.58218	0.60039	0.55858
Freezing (3.11)	0.80765	0.80765	0.83985	0.73148
Postural Stability (3.12)	0.56284	0.56284	0.56904	0.48557
Posture (3.13)	0.65030	0.65030	0.70935	0.71020
Body Bradykinesia (3.14)	0.41986	0.41986	0.48759	0.48759
Rest Tremor in Hands (3.15)	0.65689	0.65689	0.71443	0.60283
Action Tremor in Hands (3.16)	0.64680	0.64680	0.71461	0.60027
Tremor Amplitude (3.17)	0.62492	0.62492	0.63899	0.60721
Tremor Persistence (3.18)	0.61986	0.61986	0.60966	0.56766

Note: The numbers in parentheses indicate the corresponding item codes of Part III of the MDS-UPDRS (Motor Examination).

Table 4.7 – Average Cohen’s Kappa by classification using MobileNetV2, Xception, ResNet50, and VGG16 comparing different features.

Task	MobileNetV2	Xception	ResNet50	VGG16
Speech (3.1)	0.51	0.48	0.52	0.42
Facial Expression (3.2)	0.52	0.51	0.51	0.55
Rigidity (3.3)	0.69	0.69	0.65	0.62
Pinch (3.4)	0.50	0.42	0.48	0.50
Hand Movements (3.5)	0.55	0.57	0.55	0.47
Pronation-Supination (3.6)	0.58	0.56	0.50	0.44
Toe Tapping (3.7)	0.49	0.46	0.42	0.44
Leg Agility (3.8)	0.34	0.41	0.43	0.43
Chair Rising (3.9)	0.51	0.47	0.50	0.42
Gait (3.10)	0.56	0.58	0.55	0.50
Freezing (3.11)	0.45	0.45	0.45	0.47
Postural Stability (3.12)	0.56	0.53	0.51	0.45
Posture (3.13)	0.57	0.55	0.51	0.53
Body Bradykinesia (3.14)	0.54	0.54	0.59	0.52
Rest Tremor in Hands (3.15)	0.67	0.62	0.71	0.59
Action Tremor in Hands (3.16)	0.62	0.56	0.64	0.58
Tremor Amplitude (3.17)	0.68	0.60	0.61	0.54
Tremor Persistence (3.18)	0.62	0.54	0.64	0.52

Note: The numbers in parentheses indicate the corresponding item codes of Part III of the MDS-UPDRS (Motor Examination).

ResNet50 – Radar Chart

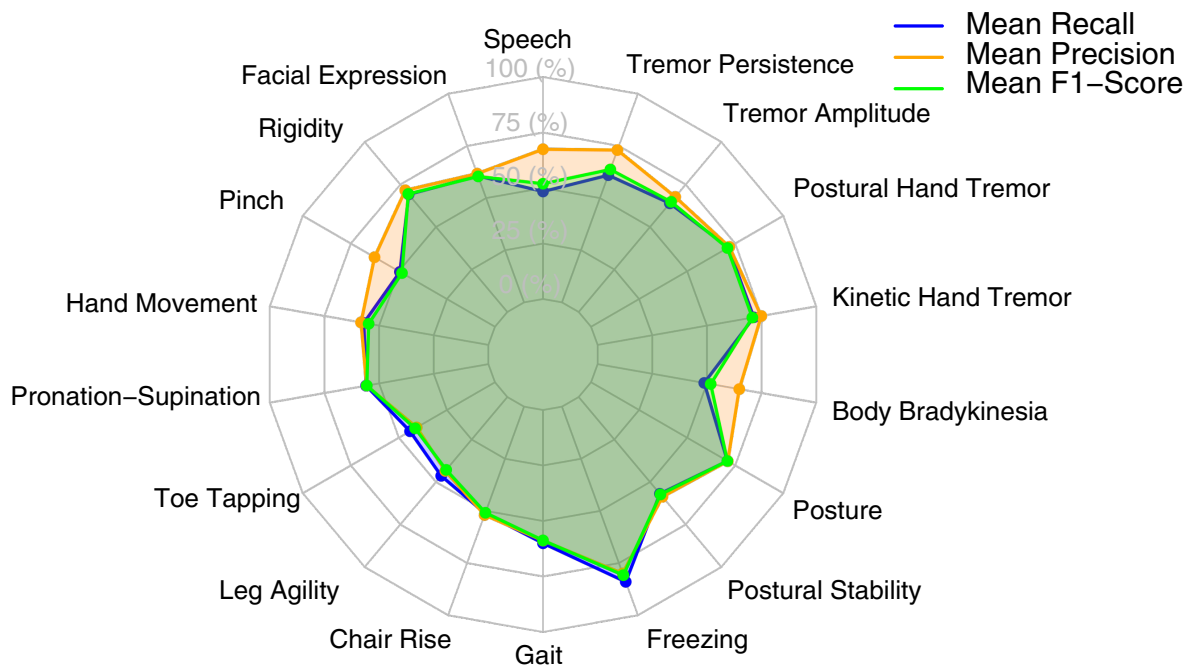


Figure 4.9 – ResNet50 radar plot.

VGG16 – Radar Chart

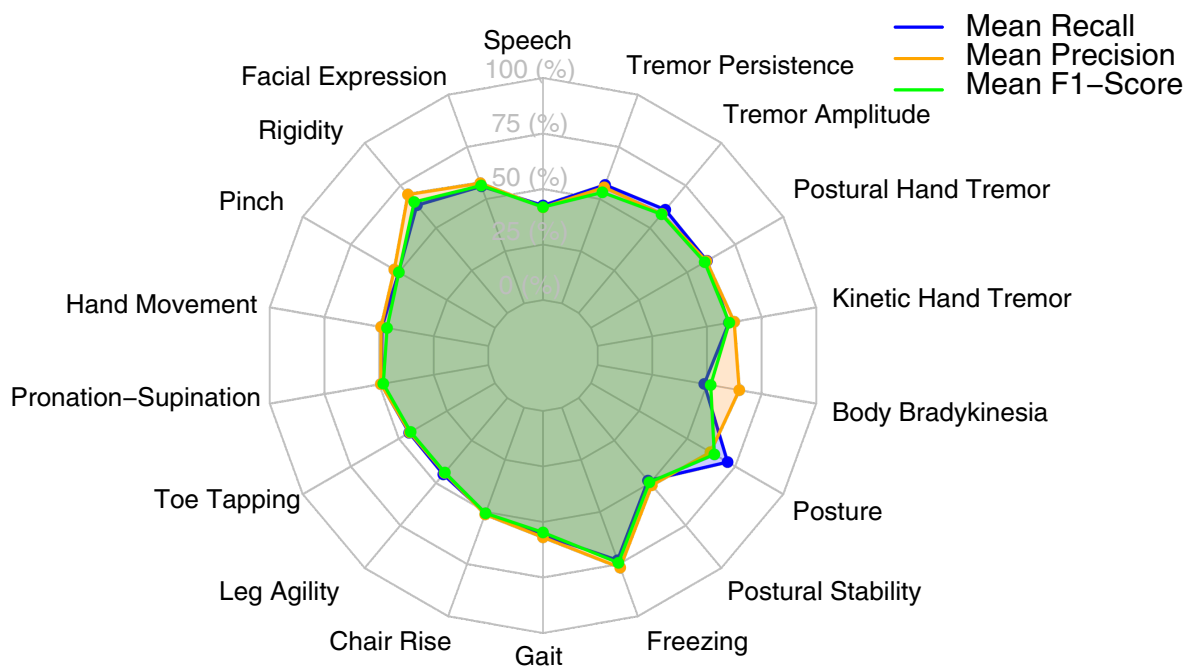


Figure 4.10 – VGG16 radar plot.

MobilenetV2 – Radar Chart

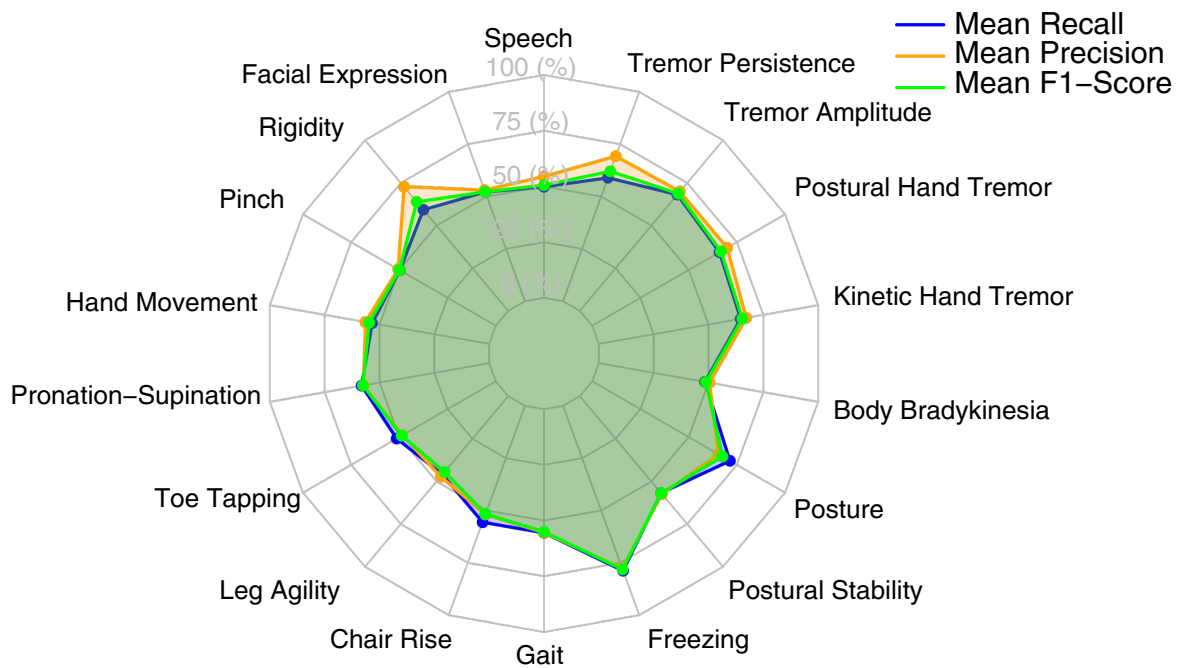


Figure 4.11 – MobileNetV2 radar plot.

Xception – Radar Chart

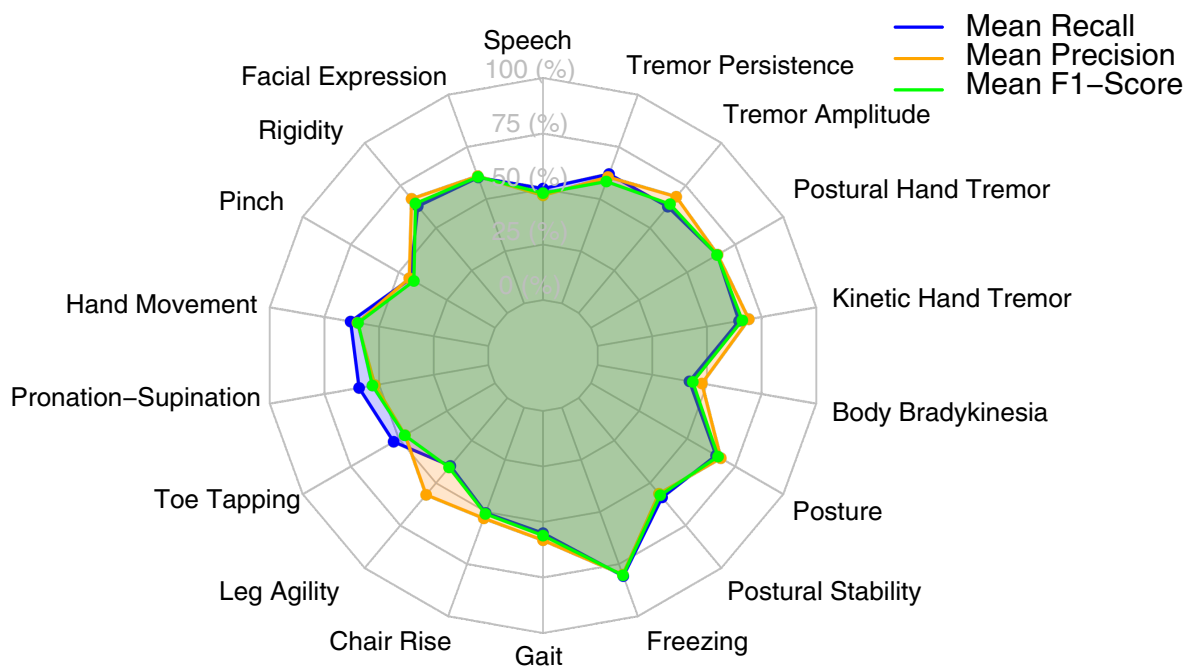


Figure 4.12 – XCEPTION radar plot.

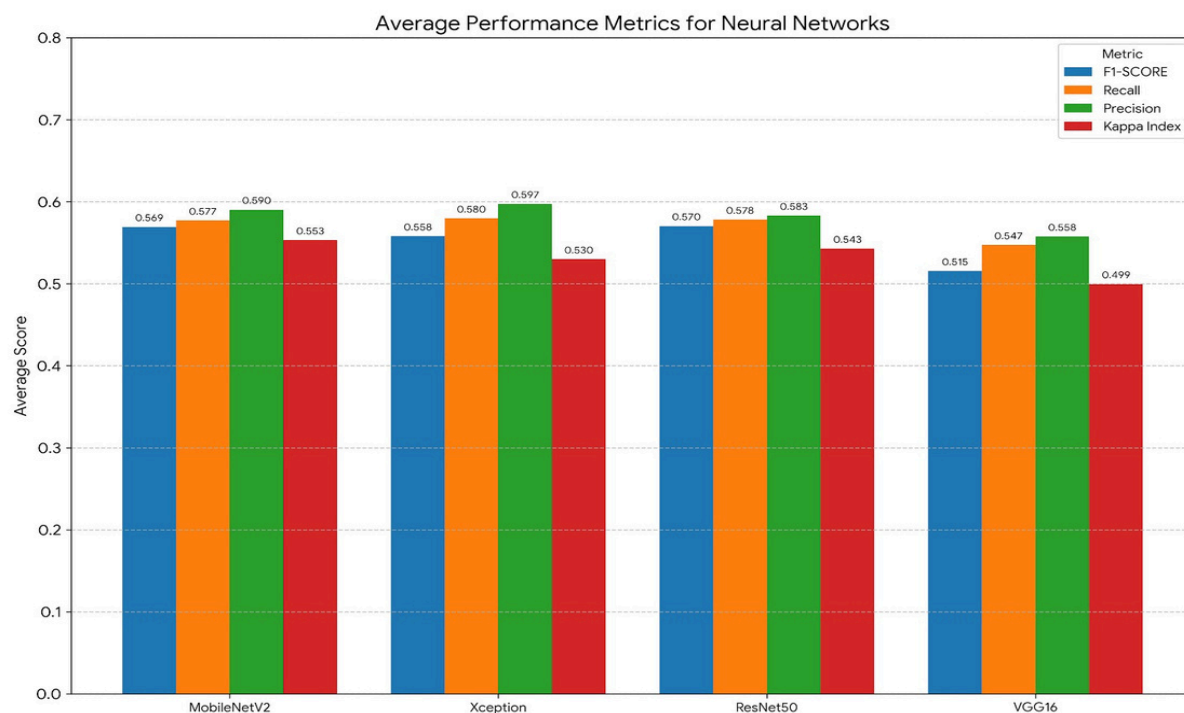


Figure 4.13 – Statistics for different CNN models comparing different features for classification.

4.9.3 Classification analysis for monitoring handwritten spirals

Training accuracy, testing accuracy, and F1-score results for the different data collection sessions are presented in Figures 4.14, 4.15 and 4.16, as well as in Tables 4.8 and 4.9. It is important to note that a unique dataset was specifically created for this study. Initially, individual results were reported, followed by the calculation of the mean training accuracy, testing accuracy, and F1-score to facilitate comparisons.

Additionally, the detailed performance of each neural network across the 16 different data collection sessions is shown in Tables 4.10, 4.11, 4.12, and 4.13, which present accuracy and F1-score values for both the control group (training) and the Parkinson's disease group (testing). These results provide a granular view of model performance, enabling the identification of variations between collections and network architectures.

Table 4.8 – Average Metric Values for the Training Set

Neural Network	F1-score	Precision	Recall	Cohen's Kappa
MobileNetV2	0.9963	0.9965	0.9961	0.9958
Xception	0.9495	0.9471	0.9530	0.9462
VGG16	0.9979	0.9980	0.9977	0.9976
ResNet50	0.9982	0.9985	0.9979	0.9980

Table 4.9 – Average Metric Values for the Test Set

Neural Network	F1-score	Precision	Recall	Cohen’s Kappa
MobileNetV2	0.2438	0.2440	0.2505	0.2400
Xception	0.2484	0.2464	0.2583	0.2471
VGG16	0.2545	0.2525	0.2621	0.2532
ResNet50	0.2891	0.2892	0.2958	0.2875

Table 4.10 – Accuracy, F1-Score and Cohen’s Kappa per Collection for MobileNetV2

Collection	CG Group (Training)			PD Group (Test)		
	Accuracy	F1-Score	Cohen’s Kappa	Accuracy	F1-Score	Cohen’s Kappa
Collection1	0.9947	0.9921	0.9894	0.3580	0.3694	0.2160
Collection2	0.9876	0.9938	0.9752	0.2609	0.2553	0.0218
Collection3	1.0000	0.9811	1.0000	0.2727	0.2655	0.0454
Collection4	0.9762	0.9879	0.9524	0.1852	0.1770	-0.2296
Collection5	1.0000	1.0000	1.0000	0.1778	0.1882	-0.2444
Collection6	1.0000	1.0000	1.0000	0.2600	0.2857	0.0200
Collection7	1.0000	1.0000	1.0000	0.2400	0.2162	-0.0200
Collection8	1.0000	1.0000	1.0000	0.2273	0.2247	-0.0454
Collection9	1.0000	1.0000	1.0000	0.2745	0.2641	0.0490
Collection10	1.0000	1.0000	1.0000	0.2500	0.2202	0.0000
Collection11	1.0000	1.0000	1.0000	0.3000	0.2885	0.1000
Collection12	1.0000	1.0000	1.0000	0.2222	0.2273	-0.0556
Collection13	1.0000	1.0000	1.0000	0.1956	0.2308	-0.1088
Collection14	1.0000	1.0000	1.0000	0.3235	0.2716	0.0470
Collection15	0.9857	0.9857	0.9714	0.1333	0.1311	-0.7334
Collection16	1.0000	1.0000	1.0000	0.2222	0.2857	-0.0556

Table 4.11 – Accuracy, F1-Score and Cohen’s Kappa per Collection for Xception

Collection	CG Group (Training)			PD Group (Test)		
	Accuracy	F1-Score	Cohen’s Kappa	Accuracy	F1-Score	Cohen’s Kappa
Collection1	0.9735	0.9364	0.9470	0.3333	0.2609	0.1666
Collection2	0.9503	0.9503	0.9006	0.3913	0.3553	0.2826
Collection3	0.9385	0.9313	0.8770	0.1091	0.1091	-0.7818
Collection4	0.9048	0.9383	0.8096	0.2778	0.2941	0.0556
Collection5	0.9333	0.9515	0.8666	0.2000	0.2169	-0.2000
Collection6	0.9565	0.9483	0.9130	0.1600	0.1569	-0.2800
Collection7	0.9478	0.9646	0.8956	0.2200	0.2500	-0.1600
Collection8	0.9118	0.9300	0.8236	0.1591	0.1687	-0.3618
Collection9	0.9664	0.9237	0.9328	0.3529	0.3051	0.7058
Collection10	0.8938	0.9395	0.7876	0.2292	0.2340	-0.5416
Collection11	0.9652	0.9487	0.9304	0.2200	0.2444	-0.5600
Collection12	0.9429	0.9474	0.8858	0.2444	0.2444	-0.5112
Collection13	0.9449	0.9406	0.8898	0.1956	0.2069	-0.6088
Collection14	1.0000	0.9939	1.0000	0.2941	0.2817	-0.4118
Collection15	0.9714	0.9714	0.9428	0.1667	0.2083	-0.6666
Collection16	0.9524	0.9756	0.9048	0.3889	0.4375	0.2778

Table 4.12 – Accuracy, F1-Score and Cohen’s Kappa per Collection for VGG16

Collection	CG Group (Training)			PD Group (Test)		
	Accuracy	F1-Score	Cohen’s Kappa	Accuracy	F1-Score	Cohen’s Kappa
Collection1	1.0000	0.9974	1.0000	0.4321	0.3665	0.8642
Collection2	1.0000	1.0000	1.0000	0.3044	0.2763	0.6088
Collection3	0.9846	0.9846	0.9692	0.2727	0.2679	0.0454
Collection4	0.9841	0.9841	0.9682	0.2593	0.2353	0.0186
Collection5	1.0000	1.0000	1.0000	0.2000	0.2046	-0.2000
Collection6	1.0000	1.0000	1.0000	0.1400	0.1628	-0.7200
Collection7	1.0000	1.0000	1.0000	0.3000	0.3030	0.6000
Collection8	1.0000	1.0000	1.0000	0.1818	0.1798	-0.3636
Collection9	1.0000	1.0000	1.0000	0.2745	0.2800	0.0490
Collection10	1.0000	1.0000	1.0000	0.2708	0.2857	0.0416
Collection11	1.0000	1.0000	1.0000	0.1800	0.1956	-0.3600
Collection12	1.0000	1.0000	1.0000	0.1778	0.1633	-0.6444
Collection13	1.0000	1.0000	1.0000	0.1522	0.1772	-0.6956
Collection14	1.0000	1.0000	1.0000	0.2941	0.3226	-0.4118
Collection15	1.0000	1.0000	1.0000	0.2667	0.2759	-0.4666
Collection16	1.0000	1.0000	1.0000	0.3333	0.3750	0.6666

Table 4.13 – Accuracy, F1-Score and Cohen’s Kappa per Collection for ResNet50

Collection	CG Group (Training)			PD Group (Test)		
	Accuracy	F1-Score	Cohen’s Kappa	Accuracy	F1-Score	Cohen’s Kappa
Collection1	1.0000	0.9947	1.0000	0.4198	0.3977	0.8396
Collection2	1.0000	1.0000	1.0000	0.3478	0.3077	0.6956
Collection3	0.9846	0.9884	0.9692	0.2182	0.2202	-0.1636
Collection4	0.9921	0.9881	0.9842	0.4074	0.3284	0.8148
Collection5	1.0000	1.0000	1.0000	0.2222	0.2532	-0.1112
Collection6	1.0000	1.0000	1.0000	0.2600	0.2524	0.0200
Collection7	1.0000	1.0000	1.0000	0.2600	0.3059	0.0200
Collection8	1.0000	1.0000	1.0000	0.1818	0.1882	-0.3636
Collection9	1.0000	1.0000	1.0000	0.3725	0.3140	0.7450
Collection10	1.0000	1.0000	1.0000	0.2500	0.2500	0.0000
Collection11	1.0000	1.0000	1.0000	0.3000	0.3191	0.1000
Collection12	1.0000	1.0000	1.0000	0.2444	0.2716	-0.5112
Collection13	1.0000	1.0000	1.0000	0.3044	0.3294	0.6088
Collection14	1.0000	1.0000	1.0000	0.3824	0.4062	0.7648
Collection15	1.0000	1.0000	1.0000	0.2333	0.2593	-0.5334
Collection16	1.0000	1.0000	1.0000	0.2222	0.2222	-0.5556

4.10 Discussion

4.10.1 Binary and multiclass classification analysis

The use of spiral drawings as a biomarker for Parkinson’s disease (PD) is supported by positive outcomes reported in the literature (35). Spiral drawings can be collected non-invasively, requiring only that the patient draws specific figures on a tablet using a

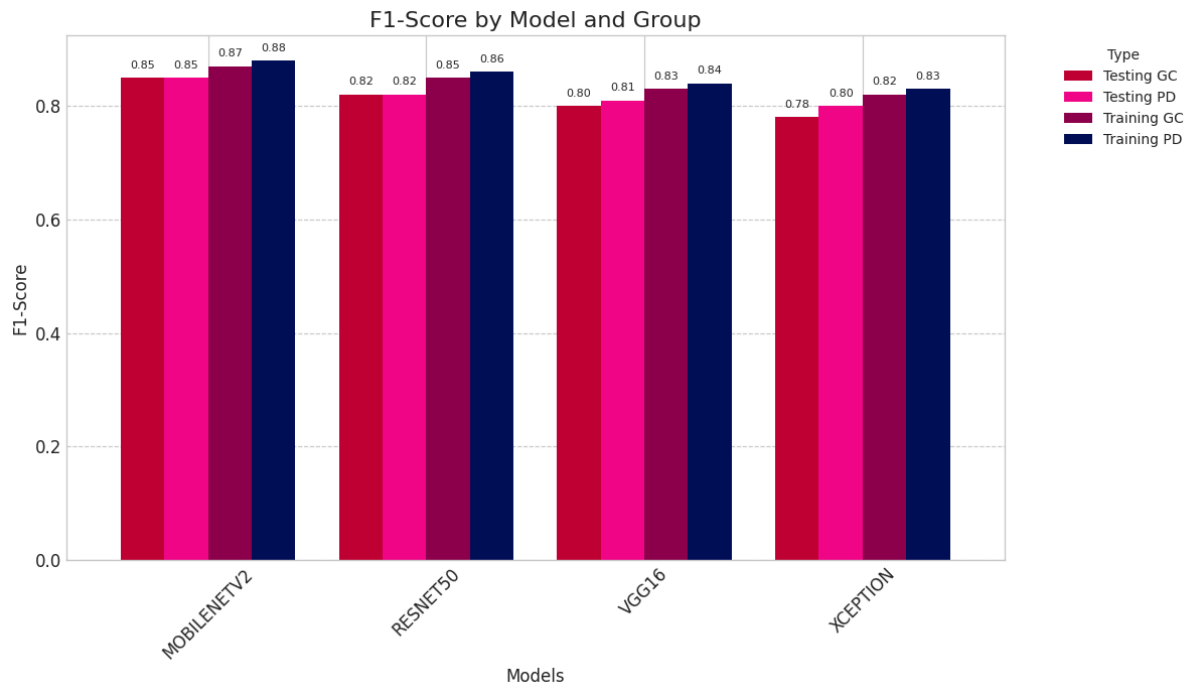


Figure 4.14 – F1-scores by model and group.

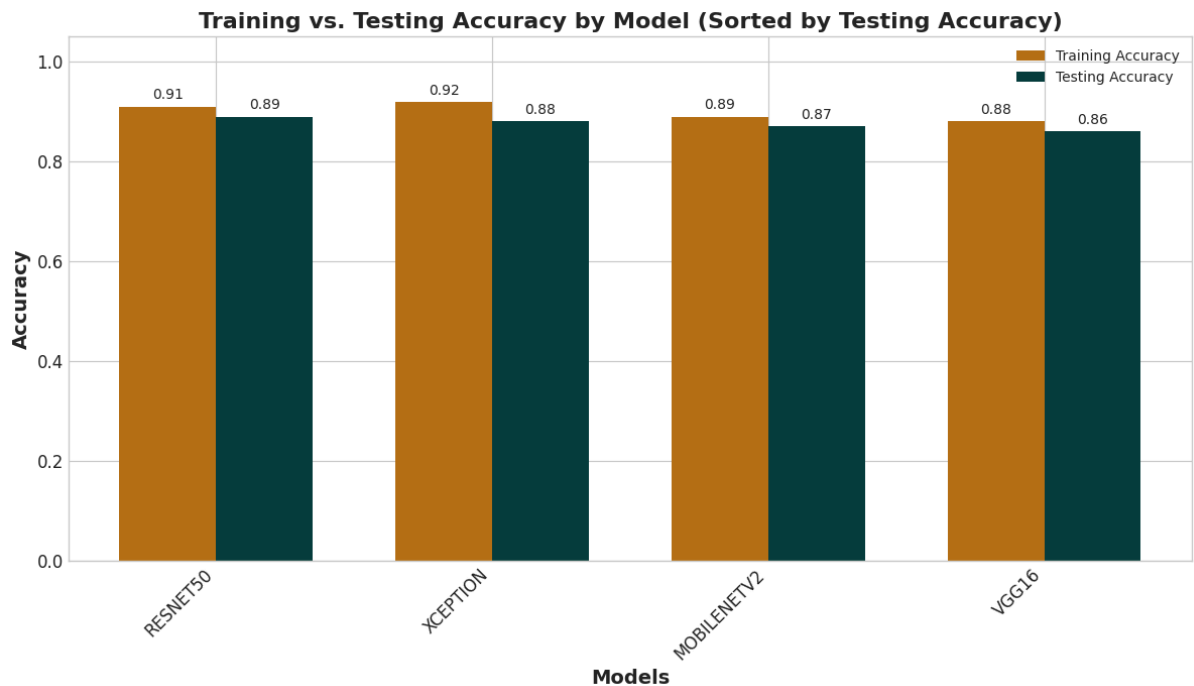


Figure 4.15 – Training vs. testing accuracy by model.

digital pen or on paper with a regular pen. This biomarker facilitates the development of medical decision support tools for PD identification and patient monitoring following diagnosis. The diagnosis of Parkinson's disease remains challenging, and improving diagnostic accuracy necessitates combining multiple biomarkers related to symptoms such as rigidity, bradykinesia, and tremor. Biomarkers that aid in patient screening can improve healthcare by enabling clinicians to focus on high-risk individuals, thereby shortening the

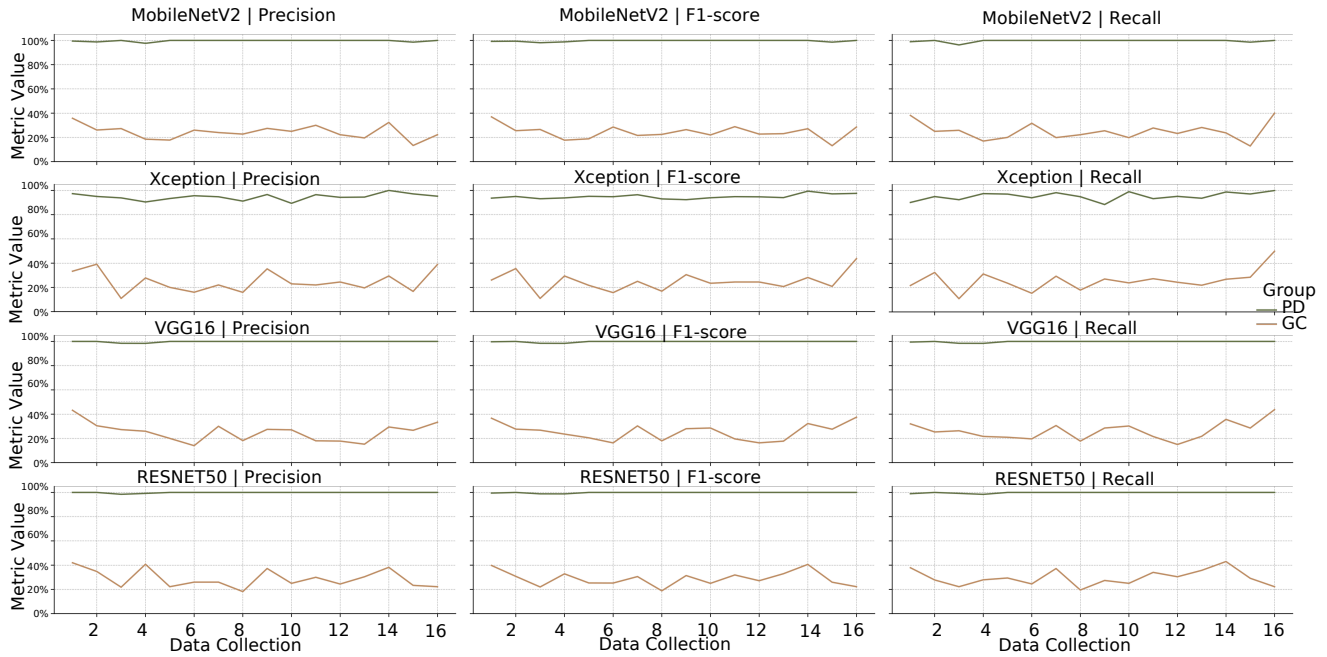


Figure 4.16 – F1-score and accuracy in different data collection sessions.

time to diagnosis and allowing earlier initiation of targeted treatment plans.

Our review began with the database created by Zham (35), which has been widely used in subsequent studies, including those by Gil Martin(13) and Kamram (21). Early studies utilizing spiral drawings employed a digital pen and tablet to evaluate various aspects of movement and drawing. Notably, Chakraborty (5) published a study in 2020 that was the first to collect spiral drawings using pen and paper, aiming to apply machine learning techniques to classify these drawings. Although pen-and-paper methods are less common, they offer practical advantages due to the accessibility of materials. Unlike previous investigations that relied on existing datasets, we developed an original dataset for this study. Additionally, most prior studies, except Zham (35), limited their analyses to binary classification of participants into PD and control groups.

In the present study, we stratified PD participants into three groups according to the Hoehn & Yahr (H&Y) scale (stages 1, 2, and 3), while also performing a binary classification between PD and the control group (CG). It is important to note that stages 4 and 5 were excluded due to the severe motor impairments that could compromise the execution of the protocol. Following Zham (35), we conducted analyses without employing data augmentation.

For clarity, we define two distinct concepts: Static Spiral Test (SST) and Dynamic Spiral Test (DST). SST considers only the Cartesian coordinates (x, y, z) of the spiral trajectory drawn by the participant, whereas DST incorporates temporal information such as drawing time, pressure, and duration, depending on the feature extraction method and authors' interests.

In this study, only SST data derived from scanned paper spirals were used. We

employed a convolutional neural network (CNN) architecture similar to that used by Huang et al. (17), targeting the four CNN models that showed the most promising results using the `net()` function available in MATLAB’s Deep Learning Toolbox. Unlike Huang et al. (13), who used a limited set of CNNs, we compared four models identified through preliminary analyses: ResNet50, VGG16, MobileNetV2, and Xception. Notably, results varied when comparing F1-score and accuracy (ACC) metrics.

In the binary classification scenario (PD vs. CG), ResNet50 achieved the highest F1-score, while Xception demonstrated superior accuracy, as shown in Figure 4.14. These results surpass those reported in recent studies employing spiral drawings and similar methods: Chandra et al (6). (4) reported an accuracy of 97.6% and an F1-score of 98.5%, whereas Nagamani et al. (25) obtained an accuracy of 86.67% and an F1-score of 87.5%. In the multiclass classification task using the H&Y scale, Xception, ResNet50, and MobileNetV2 performed best in terms of accuracy. Compared to one of the earliest H&Y-based classification studies (26), which reported 68.2% accuracy, our average accuracy was 74%. This study thus demonstrates that neural network models can effectively classify PD stages based on the H&Y scale.

Although only spiral drawings obtained from scanned paper were used, parameters such as pen pressure and drawing speed were not incorporated. However, Zham (35) indicated that these parameters do not significantly improve differentiation among disease severity levels based on F1-score. Moreover, we evaluated model generalization using two data-splitting approaches: intra-patient and inter-patient. The inter-patient approach, which tests the model’s ability to generalize to previously unseen patients, is considered more realistic. Traditional validation techniques, including holdout and cross-validation, were also employed to detect bias or overfitting, ensuring a robust evaluation. The intra-patient approach yielded higher performance, achieving a maximum accuracy of 94%.

The VGG16 model demonstrated promising results but also revealed certain limitations. The dataset’s class imbalance—with PD and control groups exhibiting substantially different image distributions—may have led the model to learn patient-specific patterns rather than generalizable features. Analysis by class showed that the model achieved higher accuracy (80%) in the minority H&Y3 class, suggesting it could capture relevant patterns despite data imbalance. However, the inter-patient evaluation highlighted the challenge of generalizing across patients, with Xception achieving approximately 63% accuracy. This model correctly identified most controls and H&Y1 cases but struggled with H&Y2 and especially H&Y3, with F1-scores of 20% and 6%, respectively. These findings align with Zham (35), which also emphasized the difficulty in distinguishing PD severity stages. Furthermore, the severe class imbalance—with 1,065 images for H&Y1 and only 105 for H&Y3—likely increased classification errors in minority classes, complicating the model’s learning process for these categories.

The analysis of Cohen’s Kappa index, presented in Tables 4.2 and 4.3, reveals the con-

sistency and reliability of the models in classifying between the studied groups. In the task of distinguishing between individuals in the control group and patients with Parkinson’s disease (CG vs. PD), all models exhibited high Kappa values (0.82 to 0.87), indicating substantial to almost perfect agreement according to the interpretation of Landis and Koch. Xception achieved the best performance (0.87), closely followed by ResNet50 (0.86) and VGG16 (0.84), suggesting high robustness in binary classification.

In the more complex task involving multiple classes (CG vs H&Y1-3), Kappa values were naturally lower (0.69 to 0.74), reflecting the greater difficulty in distinguishing the different stages of the Hoehn & Yahr scale. Nevertheless, ResNet50 (0.74) and Xception (0.73) maintained consistent performance, reinforcing their generalisation capacity even in more challenging classification scenarios. These results show that, although accuracy is important, the Kappa index provides a more balanced measure of chance and is particularly relevant in tasks with unbalanced classes.

4.10.2 Classification analysis using MDS-UPDRS

Table 4.1 presents the results obtained from the MobileNetV2, Xception, ResNet50, and VGG16 models across various classification tasks related to tremors, motor movements, and facial aspects. These results are expressed as F1-scores for each class, alongside the average F1-score for each task, facilitating a detailed comparison of model performance. Figures 4.14 4.13 complement the numerical data presented in Tables ??, 4.6 and 4.5 by providing graphical representations of model behavior.

Overall, model performance varies depending on the classification task, reflecting the unique challenges inherent to each type of analysis. ResNet50, as illustrated in Figure 4.14, consistently achieves superior results, particularly in tremor-related tasks such as kinetic hand tremor, postural tremor, tremor amplitude, and tremor persistence. For example, kinetic tremor attained an average F1-score of 0.70765, while postural tremor reached 0.70951, underscoring ResNet50’s ability to capture complex motor patterns. However, intermediate classes (e.g., stages 2 and 3) remain challenging for all models.

In coordination and movement tasks—including pinching, hand movements, and pronation-supination—the performance is more heterogeneous. The pinching task proved especially difficult, with all models scoring below 0.5; nevertheless, VGG16 performed best with an F1-score of 0.49986. This trend is evident in Figure 4, where VGG16’s radar chart highlights modest yet focused strengths. In hand movement classification, ResNet50 led with a score of 0.5961, although performance declined for more advanced classes. Similarly, ResNet50 outperformed other models in pronation-supination, with an average F1-score of 0.55605.

For non-motor tasks such as speech, facial expression, and rigidity, results varied. Speech classification was the most challenging, with average F1-scores ranging from 0.41864 (VGG16) to 0.50599 (MobileNetV2). In contrast, facial expression classifica-

tion was led by Xception, achieving an F1-score of 0.6066, as shown in Figure 4.12. Rigidity was best modeled by ResNet50, which attained an average F1-score of 0.69437, particularly excelling in classes with high stiffness levels.

Tasks involving lower limb coordination, such as toe tapping and leg agility, yielded modest results. MobileNetV2, depicted in Figure 4.11, achieved the highest toe tapping score (0.48819), although no model exceeded the 0.5 threshold. Leg agility classification scores ranged between 0.4 and 0.6, with no single model dominating.

Table 4.7 displays the mean Cohen’s kappa values for various motor tasks from Part III of the MDS-UPDRS, contrasting the classification performance of the MobileNetV2, Xception, ResNet50, and VGG16 architectures based on distinct features. Tasks pertaining to rigidity (3.3) and tremor (3.15, 3.16, 3.17, 3.18) demonstrated the highest levels of concordance between raters and models, with values exceeding 0.60 for the majority of networks; specifically, resting tremor in the hands (3.15) recorded a value of 0.71 with ResNet50, while rigidity (3.3) achieved 0.69 with both MobileNetV2 and Xception. In contrast, tasks requiring lower limb agility, such as “Leg Agility” (3.8), exhibited diminished agreement, especially with MobileNetV2 (0.34). No single model demonstrated consistent superiority across all tests; however, ResNet50 exhibited marginally greater performance in tremor-related items, MobileNetV2 excelled in rigidity, and Xception provided balanced outcomes across most variables.

Finally, Figure 4.13 summarizes the F1-scores across all models for the 18 classification tasks. This bar chart illustrates that MobileNetV2, Xception, and ResNet50 perform competitively in most categories, whereas VGG16 generally underperforms—except in select tasks such as facial expression. Notably, postural tremor and rigidity tasks appear more tractable, while toe tapping and leg agility remain the most challenging.

These findings are consistent with Milken (24), who highlights the potential of deep neural networks in biomedical applications. The strong performance of ResNet50 in tremor and rigidity tasks reinforces the value of deeper architectures in clinical contexts. Conversely, challenges observed in speech and grip-related tasks reflect high interindividual variability, a factor also discussed by Milken. Future research should focus on improving data quality, addressing dataset imbalance, prioritizing clinically relevant tasks, and exploring data fusion techniques to enhance neural network performance in biomedical domains.

4.10.3 Classification analysis for monitoring handwritten spirals

Training vs. test accuracy across DNNs

The analysis of training and test accuracies across the evaluated models (VGG16, Xception, ResNet50, and MobileNetV2) reveals notable differences in their generalization capabilities. As anticipated, training accuracy consistently exceeds test accuracy,

reflecting partial overfitting induced by model optimization, as illustrated in Figure 4.14.

The VGG16 model demonstrates high training accuracy accompanied by a substantial decline in test accuracy, indicative of a pronounced tendency toward overfitting. In contrast, Xception and ResNet50 exhibit a more balanced relationship between training and test accuracies, suggesting improved generalization to unseen data. MobileNetV2 shows the smallest discrepancy between training and test accuracies, highlighting its robustness and adaptability. These characteristics position MobileNetV2 as particularly promising for real-world applications where generalization performance is paramount.

F1-score insights across classes and models

The F1-scores across different models and groups (Control, PD, H&Y1, H&Y2, and H&Y3) further highlight differences in model generalization. Both VGG16 and ResNet50 perform well in the Control group, where classification tasks involve relatively simpler patterns. However, pronounced performance disparities emerge within the PD-related groups, reflecting increased classification complexity.

Xception consistently achieves high F1-scores across all groups, demonstrating its capacity to effectively manage variability in disease progression. Although MobileNetV2 exhibits comparatively lower F1-scores in the Control and H&Y1 groups, it attains competitive performance in the more challenging H&Y2 and H&Y3 groups. These results suggest that MobileNetV2 may be particularly suitable for applications requiring robust performance under complex conditions, as illustrated in Figure 4.15.

Training accuracy trends over 16 collections

The training accuracy trends indicate consistent performance across the evaluated deep neural networks (DNNs) over the 16 data collection sessions. ResNet50 attains the highest training accuracy in most collections, demonstrating its strong capacity to fit the training data.

In contrast, MobileNetV2 exhibits slightly lower performance, particularly in intermediate collections, which may be attributed to its architecture being optimized for computational efficiency. Both VGG16 and Xception show stable and comparable training accuracies, suggesting their adequacy for the task.

Nevertheless, the minimal differences observed in training accuracy underscore the importance of considering additional evaluation metrics, such as recall, precision and F1-scores, to guide model selection, as illustrated in Figure 4.11.

Test accuracy trends across DNNs

In the test set, performance disparities among the models become more pronounced. Xception achieves superior test accuracy, particularly in later data collection sessions,

highlighting its robustness and adaptability to unseen data. In contrast, VGG16 exhibits greater variability, with occasional declines in accuracy, further emphasizing its susceptibility to overfitting.

MobileNetV2 consistently underperforms in all metrics, reinforcing the limitations of its architecture in capturing complex data patterns. Although ResNet50 excels in training accuracy, this advantage does not translate to improved test performance, suggesting that overfitting may underlie its elevated training results, as illustrated in Figure 4.11.

F1-score performance for CG vs. PD

The F1-score analysis, which balances precision and recall, reveals substantial differences in the models' abilities to distinguish between the control group (CG) and the Parkinson's disease group (PD). Xception achieves the highest F1-scores across both groups, particularly excelling in the PD group. This highlights its capability to capture subtle, class-specific patterns within the data. Although ResNet50 demonstrates high training accuracy, it delivers lower-than-expected F1-scores, likely due to elevated false positive or false negative rates. MobileNetV2 records the lowest F1-scores, reflecting architectural limitations in addressing the complexities of this classification task. VGG16 produces intermediate results, exhibiting consistent but moderate performance across groups, as illustrated in Figure 4.11.

Tables 4.8 and 4.9 display the mean values of the performance metrics—F1-score, precision, recall, and Cohen's Kappa index—achieved by four convolutional neural network architectures (MobileNetV2, Xception, VGG16, and ResNet50) on the training and test datasets, respectively. In the training set, all networks demonstrated exceptional performance, with F1-scores nearing 1, indicating minimal difficulty in assimilating the data patterns, and elevated Kappa index values (exceeding 0.94), signifying substantial concordance between predictions and actual classes. Conversely, the outcomes on the test set demonstrate much inferior performance, with F1-scores ranging from 0.24 to 0.29 and Kappa indices falling below 0.29, highlighting the models' challenges in generalising to novel data, particularly for MobileNetV2, which exhibited the poorest performance.

Tables 4.7 to 4.13 present the accuracy, F1-score, and Cohen's Kappa index for each collection, distinguishing the Control (CG) group in the training set from the Parkinson's (PD) group in the test set for each neural network. The Control group routinely exhibits exceptionally high accuracies and F1-scores (approaching or equal to 1), indicating the models' proficiency in accurately classifying this group. Conversely, the PD group has significantly lower values, with considerable variability across collections and numerous samples displaying negative Kappa indices, suggesting poorer agreement than random chance, potentially attributable to an unbalanced dataset, noise, or clinical changes.

Among the models, ResNet50 demonstrated superior performance for the PD group across several datasets, exhibiting positive and comparatively elevated Kappa indices (e.g.,

Collection14, with a Kappa of 0.7648), indicating enhanced robustness in identifying illness patterns. Models like MobileNetV2 and Xception exhibited increased instability, evidenced by negative Kappa values across many datasets, signifying inconsistency in predictions. In conclusion, despite exhibiting strong performance during training, the networks encounter considerable hurdles in generalising the recognition of data from Parkinson's cohort, indicating a necessity for model enhancement, an increase in both the quantity and quality of data, or the use of novel training methodologies.

Final considerations

The results emphasize that optimal model selection is contingent upon the specific application context and requirements. VGG16 and ResNet50 appear better suited for simpler classification tasks, such as distinguishing the control group. In contrast, Xception and MobileNetV2 exhibit greater versatility in managing the variability associated with Parkinson's disease progression, with Xception demonstrating particular robustness across all evaluated scenarios.

Future research should explore advanced techniques, including data augmentation and domain adaptation, to further improve model robustness. Evaluating these models on multimodal datasets that integrate clinical and imaging data may provide deeper insights into their practical applicability. Additionally, balancing computational efficiency with classification accuracy will be critical for successful deployment in clinical or portable settings.

4.11 Conclusion

4.11.1 Binary and multiclass classification analysis

This study demonstrates the feasibility of classifying groups based on spiral drawings created with pen and paper by employing various convolutional neural network (CNN) architectures. We developed a classification approach using CNNs to differentiate individuals with Parkinson's disease (PD) from healthy controls (HC) without neuromuscular impairments, yielding promising results. Among the models evaluated, Xception achieved the highest performance in classifying both PD versus HC and the different Hoehn & Yahr (H&Y) stages, as measured by accuracy and F1-score. These findings indicate that feature extraction followed by classification using the Xception model provides the most effective discrimination between groups.

4.11.2 Classification analysis using MDS-UPDRS

This study conducted a comprehensive evaluation of the MobileNetV2, Xception, ResNet50, and VGG16 architectures across 18 distinct tasks related to tremors, move-

ments, and other motor and facial characteristics. The results demonstrate that neural network performance varies significantly depending on the specific task, with ResNet50 emerging as the most robust model—particularly excelling in tremor and rigidity classification.

While MobileNetV2 and Xception achieved competitive results across several domains and VGG16 exhibited isolated strengths, the overall findings reinforce the superiority of deeper and more advanced architectures in capturing complex biomedical patterns. Nonetheless, certain tasks—especially speech, facial expression, and pinching—remain particularly challenging, underscoring the need for continued advancements in model architectures and dataset quality.

The analysis revealed a consistent trend across tasks: extreme classes, such as total absence or maximum severity of a condition, are classified with higher accuracy, whereas intermediate stages pose significant challenges. This pattern highlights the inherent difficulty of fine-grained biomedical classification and the necessity for models capable of discerning subtle variations.

Moreover, visual representations in Figures 4.14 through 4.13 provide a comprehensive overview of model behavior. Figures 4.14–4.12 (radar plots) illustrate the performance distribution of each architecture across tasks, while Figure 4.13 presents comparative performance through statistical bar charts. These figures corroborate the quantitative findings and offer valuable insights into the strengths and limitations of each model.

These results align with prior literature, including Milken (18), which emphasizes the strategic significance of deep learning in biomedical applications. The strong performance of ResNet50 in tremor- and rigidity-related tasks supports its potential clinical relevance, while variability observed in speech and grip tasks reflects interindividual differences noted in previous studies.

Future work should prioritize the development of more balanced and diverse datasets, the incorporation of advanced learning techniques such as data fusion, and the adaptation of models to better accommodate interindividual variability. Such improvements are essential to enhancing the reliability and practical applicability of deep learning solutions in healthcare.

4.11.3 Classification analysis for monitoring handwritten spirals

The comparative analysis of deep neural networks (VGG16, Xception, ResNet50, and MobileNetV2) for classifying spiral drawings into Parkinson’s disease (PD) and control groups highlights the strengths and limitations of each architecture. Xception emerges as the most robust model, consistently demonstrating superior generalization across both training and test datasets and yielding high F1-scores for both control and PD groups.

Its ability to capture complex data patterns renders it a strong candidate for real-world applications, particularly in scenarios requiring reliable discrimination across stages of PD

progression. ResNet50 exhibits excellent training performance but suffers from overfitting, limiting its effectiveness on unseen data. MobileNetV2 shows promise in contexts prioritizing computational efficiency and robustness, especially under more challenging PD-related conditions; however, its overall performance is constrained by its lightweight architecture. In contrast, VGG16, while consistent, underperforms relative to the other models, making it a less suitable choice for this classification task.

This study underscores the necessity of balancing precision, generalization capacity, and computational efficiency when selecting models for clinical applications. Practical deployment decisions should consider factors such as resource availability, dataset heterogeneity, and real-time processing requirements.

Future research should investigate strategies including regularization, fine-tuning, and ensemble learning to enhance model robustness and mitigate overfitting. Additionally, assessing these architectures using multimodal datasets and diverse experimental conditions will yield a more comprehensive understanding of their applicability for diagnosing and monitoring Parkinson's disease.

Future studies

This procedure was developed during the author’s master’s studies with the aim of establishing methodologies for alternative analyses of motor symptoms in Parkinson’s disease, using a consistent form—the spiral, whether drawn or performed as a movement—as the basis for evaluation. In addition to its practical contributions, this work also lays the groundwork for the expansion of an existing theoretical framework on the assessment and motor representation in neurodegenerative conditions.

Future research aims to further develop and refine this framework, incorporating new tools and analytical perspectives. In this context, the present protocol aims to help identify possible directions for future research that can contribute not only to methodological innovation but also to the theoretical advancement of models related to the analysis of motor symptoms in Parkinson’s disease.

5.1 General conclusions of the study

This study aimed to investigate the feasibility of employing spiral-based motor assessment techniques to analyze motor symptoms in Parkinson’s Disease (PD) through the application of machine learning algorithms, particularly convolutional neural networks (CNNs). The proposed methodology was evaluated across different classification scenarios, including distinguishing individuals with PD from healthy controls (HC), classifying disease stages based on the Hoehn & Yahr (H&Y) scale, and analyzing specific tasks related to the MDS-UPDRS motor score.

Regarding the **first specific objective**—to evaluate whether different machine learning architectures exhibit significantly different performance—the results supported this hypothesis. The Xception architecture consistently achieved superior performance across multiple tasks. In contrast, networks such as ResNet50, MobileNetV2, and VGG16 demonstrated variable results depending on the classification task and its inherent complexity.

For the **second objective**, which aimed to examine the relationship between clinical

scores (MDS-UPDRS) and spiral-based motor assessments, the findings revealed coherent performance patterns in tasks associated with motor symptoms such as tremor and rigidity. Although a formal statistical correlation analysis was not conducted, the model accuracies in symptom-specific tasks suggest a promising link between spiral-derived features and the clinical severity captured by MDS-UPDRS scores.

Concerning the **third objective**—to investigate the feasibility of using spiral drawings as a tool for monitoring PD progression—results from longitudinal data collections and generalization experiments suggest that spiral trajectories hold potential as digital biomarkers. The robustness of certain architectures, particularly Xception, across multiple sessions reinforces this perspective.

Despite the encouraging results, **several limitations** must be acknowledged. The most critical limitation lies in the *limited size and diversity of the sample*, which may constrain the generalizability of the findings to broader clinical populations. Additionally, the models demonstrated difficulty in accurately classifying intermediate disease stages, underscoring both the clinical complexity of PD and the challenges faced by deep learning models in nuanced biomedical tasks. Furthermore, the absence of external validation using data from other institutions or recording devices constitutes a significant limitation that future research should address.

In conclusion, the objectives set out in this study were successfully achieved. The findings demonstrate the feasibility and clinical relevance of using CNNs applied to hand-written spirals for the assessment and monitoring of Parkinson’s Disease. This research contributes not only to methodological development but also to the theoretical understanding of motor symptom representation in neurodegenerative disorders.

5.2 Image patterns

5.2.1 Handwritten spiral images

Participants will receive oral instructions regarding the experimental procedures. They will be instructed to remain seated in a comfortable chair, with their lower extremities resting on the floor and their torso supported in an upright position. Subsequently, participants will be provided with a document detailing the activities to be completed, which will be affixed to the table using tape. Video recordings of the writing process will be captured to facilitate subsequent analysis of hand movements. Recordings will be made using a stabilized cell phone camera.

The collected data will be digitized for analysis. The experimental tasks include:

- **Part B:** Replication of sinusoidal and spiral forms, alternating the orientation of the initial movement as indicated by an arrow in the pattern presented to the participant. Each shape will be replicated five times using the dominant hand.

- **Part D:** Reproduction of spirals within areas defined by reference spirals. This task will be performed with both the left (L) and right (R) hands, in clockwise (H) and counterclockwise (AH) directions. The spirals contain dots (.) indicating the starting and ending points of the curves.

To standardize the writing instrument and surface, a BIC BLUE CRYSTAL pen with a 1 mm ballpoint tip and white A4 paper will be used.

5.2.2 Digital tool using mouse for spiral drawings images

The same set of tests described in AM5 (Parts A to D) will be replicated in a virtual environment using a computer mouse. Participants will be instructed to sit in a comfortable chair, with their lower limbs resting on the floor and their torso upright and supported. To perform the task, participants will rest their dominant forearm on a table and position their hand to operate the mouse.


Prior to the test, participants will receive instructions on how to perform the activities using an application designed to capture mouse coordinates. The mouse employed will be a standard model, such as the Logitech M90 USB, featuring an ambidextrous design, plug-and-play functionality, dimensions of $10.6 \times 14.2 \times 4.4$ cm, and a weight of approximately 100 g. The task will be conducted using a monitor of at least 19 inches to maintain consistency with the measurements used in both real and virtual environments.

Mouse coordinate data and computer screen recordings will be collected and stored for subsequent analysis. The tasks to be performed include:

- **Part B:** Reproduction of sinusoidal and spiral shapes, alternating the direction of the initial movement as indicated by an arrow in the pattern presented to the participant. Each shape will be reproduced five times using the dominant hand.
- **Part D:** Reproduction of spirals within areas defined by reference spirals. This test will be conducted with both the left (L) and right (R) hands, in clockwise (H) and counterclockwise (AH) directions. The spirals contain dots (.) marking the starting and ending points of the curves.

5.2.3 Spiral drawings using laser

In this test, the participant is instructed to trace a spiral model displayed on a screen using a red laser pointer. The objective is to assess motor symptoms associated with Parkinson's disease by capturing the trajectory of the laser as it follows the spiral fixed to a surface. Movement data will be recorded using an inertial sensor placed on the dorsum of the hand, alongside two standard cameras (e.g., smartphone cameras) positioned behind the participant and laterally to the hand.



Parkinson Brasil
Base de Dados sobre Aspectos Motores e Não Motores da Doença de Parkinson no Brasil

ID participante

ID do local de coleta

ID do avaliador

____/____/____





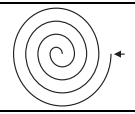
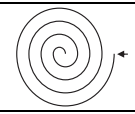
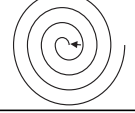
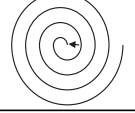
Data da avaliação

____:____:____

Horário da avaliação


RESULTADO DA AVALIAÇÃO
☐ Predominância de desenhos regulares
☐ Predominância de desenhos irregulares
☐ Observou-se lentidão na execução
☐ O participante não conseguiu executar
Mão utilizada: Direita ☐ Esquerda ☐

AM5: Parte B - Execução de desenhos em papel
¹Este teste deve ser realizado apenas com a mão dominante
²As setas indicam a origem do desenho

As setas indicam a origem do desenho

Part B.



Parkinson Brasil
Base de Dados sobre Aspectos Motores e Não Motores da Doença de Parkinson no Brasil

ID participante

ID do local de coleta

ID do avaliador

____/____/____

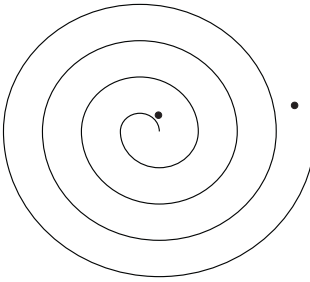
Data da avaliação

____:____:____

Horário da avaliação

RESULTADO DA AVALIAÇÃO
☐ Predominância de desenhos regulares
☐ Predominância de desenhos irregulares
☐ Observou-se lentidão na execução
☐ O participante não conseguiu executar
Sentido: Horário ☐ Anti-horário ☐
Mão utilizada: Direita ☐ Esquerda ☐

AM5: Parte D - Execução de desenhos em papel
¹Este teste deve ser realizado com ambas as mãos e ambos sentidos
²O participante deve fazer uma curva ligando os pontos indicados e dentro da região delimitada pela espiral

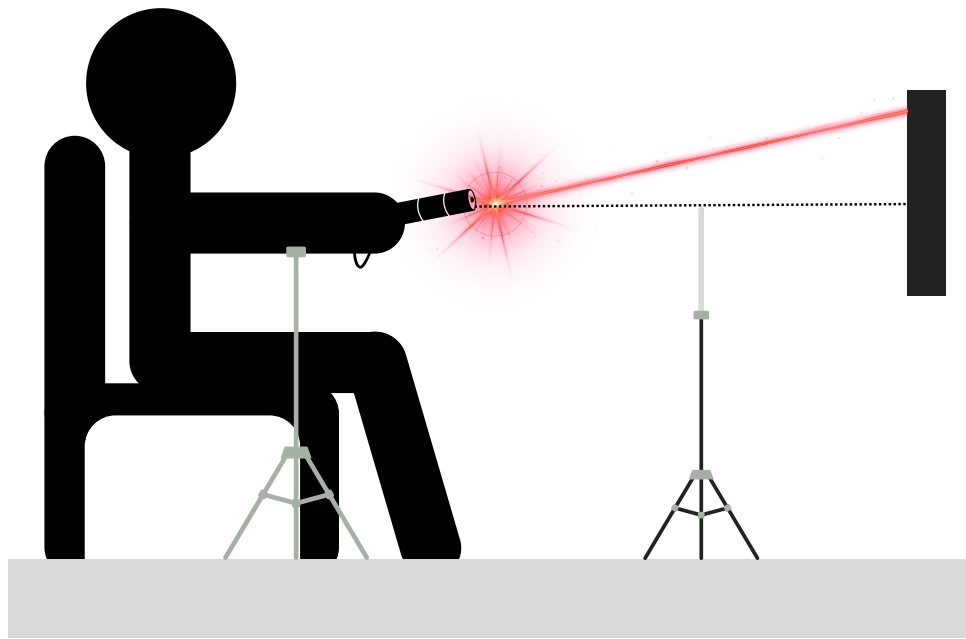


Part D.

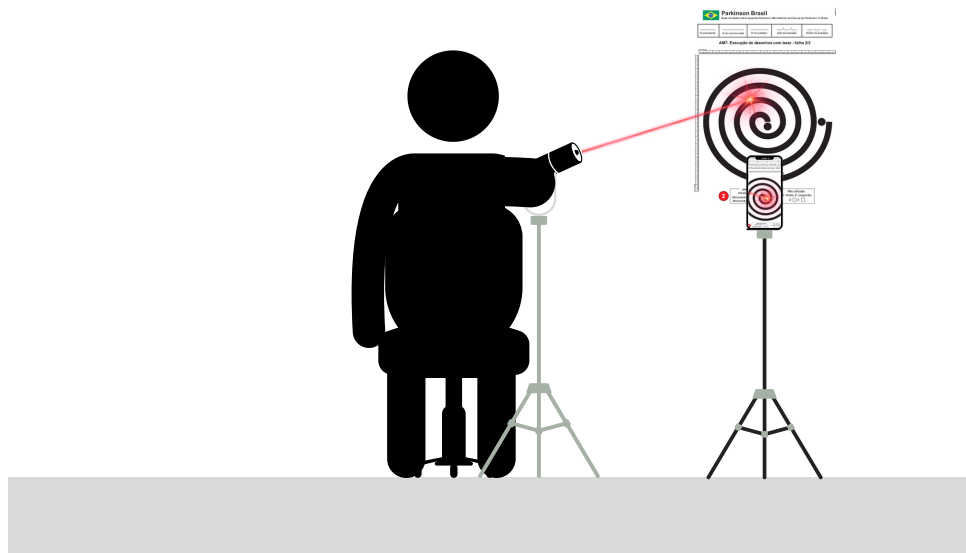
Figure 5.1 – Handwritten spiral protocol.

Participants will be seated comfortably in a chair with a spiral drawing affixed directly in front of them. They will hold the laser pointer with the dominant arm extended and supported, executing the tracing movement primarily through wrist motion. An armrest will be provided to facilitate isolation of wrist movement.

Figure 5.2 illustrates the participant’s positioning and the setup during the experi-



Front view



Side view

Figure 5.2 – Spiral drawings using laser protocol.

ment. Participants will be instructed to perform the task continuously at a natural and comfortable pace. The tracing protocol consists of three phases: (i) tracing the spiral contour from the center outward (increasing radius); (ii) tracing the contour in the reverse direction, from the outer edge inward (decreasing radius); and (iii) upon returning to the center, maintaining a static position for 20 seconds to record rest activity.

References

- 1 ALTHAM, Callum; ZHANG, H.; PEREIRA, E. Machine learning for the detection and diagnosis of cognitive impairment in Parkinson’s Disease: A systematic review. **PLoS ONE**, v. 19, n. 5, e0303644, 2024. DOI: [10.1371/journal.pone.0303644](https://doi.org/10.1371/journal.pone.0303644).
- 2 BLOEM, Bastiaan R.; OKUN, Michael S.; KLEIN, Christine. Parkinson’s Disease. **The Lancet**, v. 397, n. 10291, p. 2284–2303, 2021. DOI: [10.1016/S0140-6736\(21\)00218-X](https://doi.org/10.1016/S0140-6736(21)00218-X).
- 3 CANTÜRK, İsmail. Fuzzy recurrence plot-based analysis of dynamic and static spiral tests of Parkinson’s disease patients. **Neural Computing and Applications**, Springer Science and Business Media LLC, v. 33, n. 1, p. 349–360, 2020. DOI: [10.1007/s00521-020-05014-2](https://doi.org/10.1007/s00521-020-05014-2).
- 4 CHAKRABORTY, Sabyasachi; AICH, Satyabrata; JONG-SEONG-SIM; HAN, Eunyoung; PARK, Jinse; KIM, Hee-Cheol. Parkinson’s Disease Detection from Spiral and Wave Drawings using Convolutional Neural Networks: A Multistage Classifier Approach. In: 2020 22nd International Conference on Advanced Communication Technology (ICACT). IEEE, 2020. P. 298–303. DOI: [10.23919/icact48636.2020.9061497](https://doi.org/10.23919/icact48636.2020.9061497).
- 5 CHAKRABORTY, Saurabh; SEAL, Debpriya; MANDAL, Ashis Kumar; DHAR, Joydip; PAL, Susmita. Parkinson’s Disease Detection from Spiral and Wave Drawings using Convolutional Neural Networks: A Multistage Classifier Approach. **IEEE Access**, v. 7, p. 163356–163368, 2019. DOI: [10.1109/ACCESS.2019.2959033](https://doi.org/10.1109/ACCESS.2019.2959033).
- 6 CHANDRA, Jay; MUTHUPALANIAPPAN, Siva; SHANG, Zisheng; DENG, Richard; LIN, Raymond; TOLKOVA, Irina; BUTTS, Dignity; SUL, Daniel; MARZOUK, Sammer; BOSE, Soham; CHEN, Alexander; BHASKAR, Anushka; MANTENA, Sreekar; PRESS, Daniel Z. Screening of Parkinson’s Disease Using Geometric Features Extracted from Spiral Drawings. **Brain Sciences**, MDPI AG, v. 11, n. 10, p. 1297, 2021. DOI: [10.3390/brainsci11101297](https://doi.org/10.3390/brainsci11101297).
- 7 CHOLLET, François. Xception: Deep Learning with Depthwise Separable Convolutions. **arXiv preprint arXiv:1610.02357**, 2016.

- 8 DARU, Gilsiley Henrique; LOCH, Gustavo Valentim; PIETEZAK, Daniel Felipe. Ap-
rimorando a classificação de descrições de produtos em português com a utilização
de técnicas da recuperação de informação: uma abordagem de agrupamento de de-
scrições. **Em Questão**, v. 30, p. 1–29, 2024. DOI: [10.1590/1808-5245.30.139205A](https://doi.org/10.1590/1808-5245.30.139205A).
- 9 DESIPIO, Rebecca E. **Parkinson's Disease Automated Hand Tremor Analy-
sis from Spiral Images**. 2023. Master of Science Thesis – Virginia Polytechnic In-
stitute and State University, Blacksburg, Virginia. Available at [http://hdl.handle.
net/10919/115350](http://hdl.handle.net/10919/115350).
- 10 FARHAH, Nesren. Utilizing deep learning models in an intelligent spiral drawing
classification system for Parkinson's disease classification. **Frontiers in Medicine**,
v. 11, p. 1453743, 2024. DOI: [10.3389/fmed.2024.1453743](https://doi.org/10.3389/fmed.2024.1453743).
- 11 FOLADOR, João. **Open-Source System for Managing Parkinson's Disease
Data and Detecting Tremor Symptoms in Handwritten Drawings Using
Histogram of Oriented Gradients**. Dec. 2021. PhD Thesis – Federal University
of Uberlândia (UFU), Uberlândia, Minas Gerais, Brazil. Available at [https://repo
sitorio.ufu.br/handle/123456789/31691](https://repositorio.ufu.br/handle/123456789/31691). DOI: [10.14393/ufu.te.2021.220](https://doi.org/10.14393/ufu.te.2021.220).
- 12 GERGER, Mustafa; GÜMÜŞÇÜ, Abdülkadir. Diagnosis of Parkinson's Disease Using
Spiral Test Based on Pattern Recognition. **Romanian Journal of Information
Science and Technology**, v. 25, n. 1, p. 100–113, 2022. Available from: [https:
://www.romjist.ro/full-texts/paper710.pdf](https://www.romjist.ro/full-texts/paper710.pdf).
- 13 GIL-MARTÍN, Manuel; MONTERO, Juan Manuel; SAN-SEGUNDO, Rubén. Parkin-
son's Disease Detection from Drawing Movements Using Convolutional Neural Net-
works. **Electronics**, MDPI AG, v. 8, n. 8, p. 907, 2019. DOI: [10.3390/electronics8
080907](https://doi.org/10.3390/electronics8080907).
- 14 GOETZ, Christopher G.; POEWE, Werner; RASCOL, Olivier; SAMPAIO, Cristina;
STEBBINS, Glenn T., et al. Movement Disorder Society Task Force report on the
Hoehn and Yahr staging scale: Status and recommendations. **Movement Disorders**,
v. 19, n. 9, p. 1020–1028, 2004. DOI: [10.1002/mds.20213](https://doi.org/10.1002/mds.20213).
- 15 HE, Kaiming; ZHANG, Xiangyu; REN, Shaoqing; SUN, Jian. Deep Residual Learn-
ing for Image Recognition. **arXiv preprint arXiv:1512.03385**, 2015.
- 16 HOWARD, Andrew G.; ZHU, Menglong; CHEN, Bo; KALENICHENKO, Dmitry;
WANG, Weijun, et al. MobileNets: Efficient Convolutional Neural Networks for Mo-
bile Vision Applications. **arXiv preprint arXiv:1704.04861**, 2017.
- 17 HUANG, Yingcong; CHATURVEDI, Kunal; NAYAN, Al-Akhir; HESAMIAN, Mo-
hammad Hesam; BRAYTEE, Ali; PRASAD, Mukesh. Early Parkinson's Disease
Diagnosis through Hand-Drawn Spiral and Wave Analysis Using Deep Learning

- Techniques. **Information**, MDPI AG, v. 15, n. 4, p. 220, 2024. DOI: [10.3390/info15040220](https://doi.org/10.3390/info15040220).
- 18 IMPEDOVO, Donato; PIRLO, Giuseppe; VESSIO, Gennaro. Dynamic Handwriting Analysis for Supporting Earlier Parkinson's Disease Diagnosis. **Information**, MDPI AG, v. 9, n. 10, p. 247, 2018. DOI: [10.3390/info9100247](https://doi.org/10.3390/info9100247).
 - 19 ISHII, Nobuyuki et al. Spiral drawing: Quantitative analysis and artificial-intelligence-based diagnosis using a smartphone. **Journal of the Neurological Sciences**, v. 411, p. 116723, 2020. DOI: [10.1016/j.jns.2020.116723](https://doi.org/10.1016/j.jns.2020.116723).
 - 20 KAMBLE, Megha; SHRIVASTAVA, Prashant; JAIN, Megha. Digitized spiral drawing classification for Parkinson's disease diagnosis. **Measurement: Sensors**, Elsevier BV, v. 16, p. 100047, 2021. DOI: [10.1016/j.measen.2021.100047](https://doi.org/10.1016/j.measen.2021.100047).
 - 21 KAMRAN, Iqra; NAZ, Saeeda; RAZZAK, Imran; IMRAN, Muhammad. Handwriting dynamics assessment using deep neural network for early identification of Parkinson's disease. **Future Generation Computer Systems**, Elsevier BV, v. 117, p. 234–244, 2021. DOI: [10.1016/j.future.2020.11.020](https://doi.org/10.1016/j.future.2020.11.020).
 - 22 KIMBER, Thomas E.; THOMPSON, Philip D. Upper limb tremor. **Medical Journal of Australia**, 2012. DOI: [10.5694/mja11.11565](https://doi.org/10.5694/mja11.11565).
 - 23 KO, J.; KIM, E. On ECOC as Binary Ensemble Classifiers. In: **MACHINE Learning and Data Mining in Pattern Recognition**. Springer Berlin Heidelberg, 2005. P. 1–10. DOI: [10.1007/11510888_1](https://doi.org/10.1007/11510888_1).
 - 24 MILKEN, Pedro Henrique Bernardes Caetano; SILVA SOUSA, Leandro Rodrigues da; SANTOS COUTO PAZ, Clarissa Cardoso dos; NASUTO, Slawomir J.; REED, Catherine M. Sweeney; OLIVEIRA ANDRADE, Adriano de. Spiral Analysis in Parkinson's Disease: A Bibliometric Analysis. In: SOARES, Alcimar Barbosa; LEONI, Renata Ferranti; CARDOSO, George Cunha (Eds.), **2. XXIX Brazilian Congress on Biomedical Engineering - Volume 2: Tissue Engineering, Clinical Engineering and Computational Modeling in Biomedical Engineering**. Cham, Switzerland: Springer Cham, Sept. 2025. v. 126. (IFMBE Proceedings, 2). Proceedings of CBEB 2024, September 2-6, 2024, Ribeirão Preto-SP, Brazil. ISSN: 1680-0737. Available at <https://link.springer.com/book/9783031949203>.
 - 25 NAGAMANI, T. Automatic Diagnosis of Parkinson's Disease using Handwriting Patterns. **Journal of Electrical Systems**, Science Research Society, v. 20, 7s, p. 1395–1405, 2024. DOI: [10.52783/jes.3712](https://doi.org/10.52783/jes.3712).
 - 26 NÓBREGA, Lígia Reis. **A Method to Assess Freezing of Gait in Parkinson's Disease with Inertial Sensors**. Dec. 2023. PhD Thesis – Federal University of Uberlândia (UFU), Uberlândia, Minas Gerais, Brazil. Available at <https://repositorio.ufu.br/handle/123456789/19792>. DOI: [10.14393/ufu.te.2023.654](https://doi.org/10.14393/ufu.te.2023.654).

-
- 27 POSTUMA, Ronald B. et al. MDS clinical diagnostic criteria for Parkinson's disease. **Movement Disorders**, Wiley, v. 30, n. 12, p. 1591–1601, 2015. DOI: [10.1002/mds.26424](https://doi.org/10.1002/mds.26424).
 - 28 ROSEBROCK, Adrian. **Detecting Parkinson's Disease with OpenCV, Computer Vision, and the Spiral/Wave Test**. 2019. <https://pyimagesearch.com/2019/04/29/detecting-parkinsons-disease-with-opencv-computer-vision-and-the-spiral-wave-test/>.
 - 29 SAUNDERS-PULLMAN, Rachel; DERBY, Carol; STANLEY, Kaili; FLOYD, Alicia; BRESSMAN, Susan, et al. Validity of Spiral Analysis in Early Parkinson's Disease. **Movement Disorders**, v. 23, n. 4, p. 531–537, 2008. DOI: [10.1002/mds.21874](https://doi.org/10.1002/mds.21874).
 - 30 SIMON, David K.; TANNER, Caroline M.; BRUNDIN, Patrik. **Parkinson Disease: Epidemiology, Pathology, Genetics, and Pathophysiology**. 2024. Artigo do UpToDate.
 - 31 SIMONYAN, Karen; ZISSERMAN, Andrew. Very Deep Convolutional Networks for Large-Scale Image Recognition. **arXiv preprint arXiv:1409.1556**, 2014.
 - 32 SUBRAMANIAN, Krithika. **Handwriting Based Parkinson's Disease Diagnosis**. n.d. https://www.cs.utexas.edu/~krisub/HBPDD_Research.pdf.
 - 33 TOLOSA, Eduardo; WENNING, Gregor; POEWE, Werner. The diagnosis of Parkinson's disease. **The Lancet Neurology**, v. 5, n. 1, p. 75–86, 2006. DOI: [10.1016/S1474-4422\(05\)70285-4](https://doi.org/10.1016/S1474-4422(05)70285-4).
 - 34 VIERA, Anthony J.; GARRETT, J. Michael. Inter-rater reliability: the kappa statistic. **Biochem Med (Zagreb)**, v. 22, n. 3, p. 276–282, 2012.
 - 35 ZHAM, Poonam; KUMAR, Dinesh K.; DABNICHKI, Peter; POOSAPADI ARJUNAN, Sridhar; RAGHAV, Sanjay. Distinguishing Different Stages of Parkinson's Disease Using Composite Index of Speed and Pen-Pressure of Sketching a Spiral. **Frontiers in Neurology**, Frontiers Media SA, v. 8, 2017. DOI: [10.3389/fneur.2017.00435](https://doi.org/10.3389/fneur.2017.00435).

Glossary

Archimedes' Spiral

It is a plane curve generated by a point moving away from or approaching a fixed point at a constant rate, while the radius vector of the fixed point rotates at a constant rate.

Dopamine

A compound present in the body as a neurotransmitter that is produced in the substantia nigra, ventral tegmental area, and hypothalamus of the brain.

Hoehn and Yahr scale

A scale for assessing the disability of individuals with PD capable of indicating their general condition quickly and effectively.

MDS-UPDRS

MDS-UPDRS is a revision of the Unified Parkinson's Disease Rating Scale (UPDRS) originally developed in the 1980s. Addresses several problematic areas of the original scale identified by an MDS review task force. The MDS-UPDRS was developed to evaluate various aspects of Parkinson's disease, including motor and non-motor experiences of daily living and motor complications. It includes a motor evaluation and characterizes the extent and burden of disease across various populations. The scale can be used in a clinical setting and in research.

Spiral Analysis

It is a clinical component useful for supporting or depressing the motor, detecting and monitoring motion disturbances.

UFU

Federal University of Uberlândia.

Appendix 1

A.1 Environment for data collection

The experimental protocol involved a paper-and-pen test, with the paper positioned on a table directly in front of the participant, as illustrated in Figure 9. The test comprised five spirals and five sinusoidal waves, which the participant was instructed to draw continuously from left to right without pausing until the entire sequence was completed (see Figure A.1).

This procedure was repeated monthly over a twelve-month period, enabling longitudinal monitoring of changes in drawing patterns. At the conclusion of each session, all completed test sheets were labeled and prepared for scanning.

During the initial session, the Movement Disorder Society-Unified Parkinson's Disease Rating Scale (MDS-UPDRS) Part III and the Hoehn & Yahr (H&Y) scale were administered following the drawing protocol.

Post-scanning, the images were processed using CorelDRAW software. Each spiral was individually isolated and standardized into separate image files. Minor retouching was applied to correct distortions without modifying or compromising the original drawings, ensuring the integrity of the data for subsequent machine learning analysis.



Figure A.1 – Example of the instrument used for data collection and a typical data sample (on the right) obtained from a person with Parkinson’s disease.

Appendix 2

This appendix compiles the supplementary materials generated during the course of this dissertation.

B.1 Codes for the analysis

The files include a subset of the anonymized data collected and the script developed for the analysis of inter-examiner agreement. The code, written in R, along with the data, is available at:

Henrique Bernardes Caetano, P. Andrade, A. (2025). Binary and Multiclass Classification codes. Zenodo. <https://doi.org/10.5281/zenodo.15643538>

B.2 Open access database

Researchers may utilize the open-access database available via the following DOI to facilitate the development of novel innovations aimed at improving the quality of life for individuals with Parkinson's disease. Contributions to the ongoing research conducted by the Centre for Innovation and Technology Assessment in Health (NIATS) are also encouraged. Access to the data reported in this study is restricted and managed through Zenodo. Interested parties must submit a request via Zenodo, which will be evaluated based on its justification before access is granted.

DOI = <https://doi.org/10.5281/zenodo.13891806>