

**SÃO PAULO STATE UNIVERSITY**  
SCHOOL OF SCIENCES - CAMPUS OF BAURU  
DEPARTMENT OF COMPUTING  
POSTGRADUATE PROGRAM IN COMPUTER SCIENCE  
SÃO PAULO, BRAZIL

**RMIT UNIVERSITY**  
SCHOOL OF ENGINEERING  
COLLEGE OF SCIENCE, TECHNOLOGY, ENGINEERING AND MATHS  
MELBOURNE, AUSTRALIA

Investigation of AI-based image, video, and voice analysis to  
assess clinical symptoms

**Guilherme Camargo de Oliveira**

**São Paulo / SP**  
**May 2025**

**Guilherme Camargo de Oliveira**

**Investigation of AI-based image, video, and voice analysis to  
assess clinical symptoms**

Doctoral dissertation submitted in fulfillment of the requirements for the Joint Doctoral Program between São Paulo State University (UNESP), Faculty of Sciences, Bauru Campus, Postgraduate Program in Computer Science, and the Royal Melbourne Institute of Technology University (RMIT), Melbourne, Australia. Concentration Area: Computer Intelligence in Applied Computing.

Supervisors: Prof. Dr. João Paulo Papa (UNESP) and Prof. Dr. Dinesh Kumar (RMIT)

Co-supervisors: Dr. Leandro A. Passos (UNESP) and Dr. Quoc Ngo (RMIT)

**São Paulo  
May 2025**

O48i

Oliveira, Guilherme Camargo

Investigation of AI-based image, video, and voice analysis to assess clinical symptoms / Guilherme Camargo Oliveira. -- Bauru, 2025  
199 p.

Tese (doutorado) - Universidade Estadual Paulista (UNESP),  
Faculdade de Ciências, Bauru

Orientador: João Paulo Papa

Coorientador: Leandro Aparecido Passos Júnior

1. Inteligência Artificial. 2. Diagnósticos Não Invasivos. 3. Análise de Vídeo. 4. Análise de Voz. 5. Dados Sintéticos. I. Título.

**ATA DA DEFESA PÚBLICA DA TESE DE DOUTORADO DE GUILHERME CAMARGO DE OLIVEIRA, DISCENTE DO PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO, DA FACULDADE DE CIÊNCIAS - CÂMPUS DE BAURU.**

Aos 19 dias do mês de maio do ano de 2025, às 9h, por meio de Videoconferência, realizou-se a defesa de TESE DE DOUTORADO de GUILHERME CAMARGO DE OLIVEIRA, intitulada "**Investigation of AI-based image, video, and voice analysis to assess clinical symptoms**". A Comissão Examinadora foi constituída pelos seguintes membros: Prof. Dr. JOÃO PAULO PAPA (Orientador(a) - Participação Virtual) do(a) Departamento de Computação / UNESPCâmpus de Bauru, Prof. Dr. KELTON AUGUSTO PONTARA DA COSTA (Participação Virtual) do(a) FC / UNESP Bauru SP, Prof. Dr. ANDRÉ LUIS DEBIASO ROSSI (Participação Virtual) do(a) Departamento de Computação / UNESP - Campus de Bauru, Prof. Dr. HELIO PEDRINI (Participação Virtual) do(a) Instituto de Computação / Universidade Estadual de Campinas, Prof. Dr. RICARDO CERRI (Participação Virtual) do(a) Departamento de Computação / Universidade Federal de São Carlos. Após a exposição pelo doutorando e arguição pelos membros da Comissão Examinadora que participaram do ato, de forma presencial e/ou virtual, o discente recebeu o conceito final: APROVADO . Nada mais havendo, foi lavrada a presente ata, que após lida e aprovada, foi assinada pelo(a) Presidente(a) da Comissão Examinadora.

Prof. Dr. JOÃO PAULO PAPA



Documento assinado digitalmente  
JOAO PAULO PAPA  
Data: 17/07/2025 13:24:26-0300  
Verifique em <https://validar.iti.gov.br>

# ACKNOWLEDGEMENTS

---

---

I am deeply grateful for the guidance, support, and encouragement of numerous individuals who made this thesis possible.

First, I express my sincere gratitude to my supervisor, Prof. Dinesh Kant Kumar, who has guided and encouraged me throughout my doctoral journey. Beyond being a mentor, he has become a friend, offering not only his immense support and insightful advice but also his unwavering belief in my abilities, which have been instrumental in completing this work. I am particularly thankful to Prof. Kumar for providing me the opportunity to be a part of RMIT University through the Joint Ph.D. program between UNESP and RMIT University, as well as for the RMIT scholarship, which greatly facilitated my research.

I am also profoundly thankful to Prof. João Paulo Papa, my supervisor both during my undergraduate research and my Ph.D., for his mentorship and friendship. His support has allowed me to broaden my research experience internationally, and I am deeply grateful for the opportunities he has provided, both academically and personally.

My deepest gratitude goes to my family for their unwavering love and support. Their support was not just a part of my journey but a significant factor that kept me going. To my mother, Marcia Cristina Camargo de Oliveira, and my brother, Gabriel Camargo de Oliveira, thank you for your endless encouragement, patience, and understanding. I dedicate this work to the memory of my father, Sergio Batista de Oliveira, who passed away during my Ph.D. His belief in me has been a constant source of inspiration and strength.

I would like to offer my heartfelt thanks to my co-authors and colleagues: Dr. Quoc Ngo, Dr. Nemuel D. Pah, Dr. Leandro Passos, Vikash Shaw, Leonardo de Oliveira, Arissa Yoshida, and Nicolás Gomes. Their collaboration, insightful discussions, and camaraderie have not just been a part of this journey but have greatly enriched my research, making it more comprehensive and insightful. I also extend my heartfelt gratitude to my partner, Alexandra Rodrigues, for her unwavering support and encouragement throughout this journey.

I extend my deepest gratitude to everyone who has contributed to my journey in one way or another.

# ABSTRACT

The rapid advancement of artificial intelligence (AI) is transforming healthcare, offering the potential to enhance diagnostic accuracy, streamline clinical workflows, and personalize treatment plans. However, the comprehensive application and integration of AI technologies in healthcare face challenges, particularly in enhancing non-invasive screening methods. This thesis investigates the application of AI-assisted tools across three key modalities—video, voice, and image—to improve clinical decision-making and patient outcomes through non-invasive methods. Focusing on neurological conditions such as Parkinson’s disease, stroke, and Amyotrophic Lateral Sclerosis (ALS), as well as ophthalmology and wound care, the research is guided by three main questions. While the first two research questions leverage video and voice analysis to detect subtle neurological symptoms—addressing key challenges of non-invasive diagnostics such as subjective clinical assessments, delayed timeliness, and limited patient monitoring—the third question aims to enhance AI non-invasive methods in ophthalmology and wound care by overcoming data scarcity and advancing image translation techniques. The three questions are: **(I) How can AI-assisted facial expression analysis enhance the detection and understanding of neurological conditions such as Parkinson’s disease, stroke, and ALS?** The study demonstrated that AI-assisted facial expression models could detect subtle symptoms of these disorders, achieving 83% accuracy in identifying hypomimia associated with Parkinson’s disease. Similar techniques effectively detected facial weaknesses in Post-Stroke and ALS patients, highlighting the value of AI-driven video analysis for non-invasive assessments. This approach offers a groundbreaking non-invasive way to identify subtle symptoms that might otherwise go unnoticed. Additionally, an AI-driven stroke app can assist in screening cases with just a smile in emergency departments, highlighting the potential of video analysis for rapid and non-invasive assessments. **(II) In what ways can AI-based voice analysis tools improve the remote assessment of Parkinson’s disease severity and support ongoing monitoring?** This study integrates ensemble Diadochokinetic analysis to identify severity of Parkinson, leverages formant-based vocal tract length measurements from phonemes to detect subtle changes, and utilizes an large language models (LLM) as an agent for real-time patient feedback. Together, these components offer a scalable, non-invasive solution for improved early detection and continuous management of Parkinson’s disease. **(III) How do AI-powered synthetic imaging techniques contribute to the detection and diagnosis of medical conditions like age-related macular degeneration and venous leg ulcers?** In imaging, deep learning models such as StyleGAN-2 achieved 85% accuracy in detecting age-related macular degeneration, outperforming human experts. Additionally, AI-generated thermal imaging achieved promising results for chronic wound assessment with an SSIM score of 0.84, although further validation is necessary. In conclusion, this thesis underscores the transformative potential of AI in healthcare, providing non-invasive solutions that improve early detection, facilitate remote monitoring, and enhance

diagnostic precision. Future efforts must address demographic biases, ensure ethical data use, and work with regulatory bodies to integrate these tools into clinical practice, advancing towards more accessible and effective healthcare solutions.

**Keywords:** Artificial Intelligence, Non-Invasive Diagnostics, AI-Assisted Tools, Video Analysis, Voice Analysis, Data Augmentation, Synthetic Data, Neurological Conditions, Parkinson's Disease, Stroke, Amyotrophic Lateral Sclerosis, Ophthalmology, Wound Care, Machine Learning, Remote Monitoring.

# LIST OF FIGURES

---

---

Figure 1 – Thesis Contribution. . . . .	21
Figure 2 – The number of PubMed publications related to facial expression in Parkinson’s disease from 1967 to September 2024. It includes emotion recognition, hypomimia detection, and diagnosis by facial expression. . . . .	34
Figure 3 – Flow chart of the literature selection process. . . . .	38
Figure 4 – The essential steps in identifying Parkinson’s Disease through facial expressions. . . . .	43
Figure 5 – Generator architecture. . . . .	66
Figure 6 – Discriminator architecture. . . . .	67
Figure 7 – Steps of Test-Time Augmentation . . . . .	68
Figure 8 – Relationship of Accuracy and size of synthetic training data generated using the five algorithms. . . . .	69
Figure 9 – Standard Deviation in TTA . . . . .	69
Figure 10 – Performance of data augmentation. . . . .	70
Figure 11 – Example of Py-Feat extraction: (a) original image, (b) facial box and landmarks, (c) action units. . . . .	76
Figure 12 – Feature extraction flow. . . . .	76
Figure 13 – System architecture. . . . .	86
Figure 14 – Preprocessing and feature extraction for the ALS vs. healthy controls subset. . . . .	90
Figure 15 – Preprocessing and feature extraction for the post-stroke vs. healthy controls subset. . . . .	90
Figure 16 – Overview of the Facial Point Graph model approach. . . . .	91
Figure 17 – The distribution of severity of speech symptoms of Parkinson’s disease among the participants in this study: UPDRS-speech average score of 1.34 ( $\pm$ 0.82) . . . . .	99
Figure 18 – Example of a Logistic Regression pipeline fitted to “ta-ta-ta”: phonation features are extracted to classify Normal stage vs. not Normal stage. . . . .	101
Figure 19 – Inference of the ensemble approach. . . . .	105
Figure 20 – Confusion Matrix of Ensemble approach . . . . .	106
Figure 21 – The statistical distribution of apparent AVTL(F <sub>1</sub> ), AVTL(F <sub>2</sub> ), AVTL(F <sub>3</sub> ), and AVTL(F <sub>4</sub> ) or the sustained phoneme /a/ extracted from UCI dataset. (a) The distribution of all participants (b) The distribution of male participants, (c) The distribution of female participants. The effect size (ES) was displayed for the cases with a p-value of less than 0.05 and ES . . . . .	111

Figure 22 – The statistical distribution of apparent AVTL( $F_1$ ), AVTL( $F_2$ ), AVTL( $F_3$ ), and AVTL( $F_4$ ) or the sustained phoneme /a/ extracted from ITA dataset. (a) The distribution of all participants (b) The distribution of male participants, (c) The distribution of female participants. The effect size (ES) was displayed for the cases with a p-value of less than 0.05 and ES . . . . .	112
Figure 23 – Bot architecture: (a) The agent is responsible for starting the “PD Assessment” conversational flow; (b) The “PD Assessment” flow involves recording speech and running the machine learning model; (c) The result is returned to the agent, which then transmits it to the user in a user-friendly manner. . . . .	117
Figure 24 – Screenshot of NestNeuro. . . . .	119
Figure 25 – Example of a thermal image: the ulcer is clearly highlighted. . . . .	122
Figure 26 – Preprocessing step: (a) original RGB photo and (b) its postprocessed version. . . . .	124
Figure 27 – Comparison of a venous leg ulcer image: (a) shows the original RGB photo, while (b) and (c) display the thermal photo before and after LoRA Dreambooth fine-tuning, respectively. . . . .	126
Figure 28 – Results of ControlNet conditioned on Soft edges, featuring (a) RGB image, (b) captured thermal image, and (c) Stable Diffusion generated thermal image. . . . .	126
Figure 29 – Outcomes of ControlNet conditioned on recolor: (a) the RGB image, (b) the Stable Diffusion-generated image with the wound highlighted in strong yellow for easier segmentation, and (c) the wound segmented using "Segment Anything". . . . .	128
Figure 30 – Retina fundus image positive to age-related macular degeneration identified by the presence of drusen. The image was extracted from the iChallenge-AMD dataset (FU <i>et al.</i> , 2020). . . . .	130
Figure 31 – Sample image extracted from ODIR-2019 dataset and its corresponding transformations: (a) original image, (b) background removal using Hough Circle Transform and resizing, and (c) central cropping. . . . .	133
Figure 32 – Number of images per dataset to compose the test set: (a) images positive to AMD and (b) non-AMD images. . . . .	134
Figure 33 – Examples of (a) real retina images extracted from the training dataset, and (b) synthetic images generated by StyleGAN2-ADA. . . . .	137
Figure 34 – Examples of synthetic and real images for AMD and Non_AMD. (a) real, positive AMD, (b) synthetic, positive AMD, (c) real, Non-AMD and (d) synthetic, non-AMD. . . . .	138
Figure 35 – Accuracy over the test set for different percentages of synthetic image for augmentation purposes (a). Accuracy over the test set concerning ResNet18, AlexNet, and SqueezeNet architectures (b). . . . .	138
Figure 36 – Screenshots. . . . .	139

Figure 37 – Screenshot showing, from left to right, the websites of *The Conversation*, *Forbes*, and *RMIT News*, which feature AI-based stroke detection research. . 182

# LIST OF TABLES

---

---

Table 1 – Summary of Research Studies. . . . .	27
Table 2 – Approaches for facial expression analysis to identify PD (Part 1): <i>References, Expression, and Data</i> . . . . .	44
Table 3 – Approaches for facial expression analysis to identify PD (Part 2): <i>References, Expression, and Data</i> (Continued). . . . .	45
Table 4 – Approaches for facial expression analysis to identify PD (Feature, Model, and Metrics) — Part 1. . . . .	46
Table 5 – Approaches for facial expression analysis to identify PD (Feature, Model, and Metrics) — Part 2 (Continued). . . . .	47
Table 6 – Approaches for facial expression analysis to identify PD (Feature, Model, and Metrics) — Part 3 (Continued). . . . .	48
Table 7 – Comparison of some results using Jin et al. (JIN <i>et al.</i> , 2020) approaches in a different work. . . . .	51
Table 8 – Amount of images per training/test sets. . . . .	63
Table 9 – Grid-search space configuration . . . . .	64
Table 10 – Confusion matrix for Logistic Regression. . . . .	68
Table 11 – Confusion matrix for Logistic Regression when introducing 1,600 synthetically augmented data. . . . .	68
Table 12 – Confusion matrix for Logistic Regression when introducing 1,600 synthetically augmented data and TTA. . . . .	68
Table 13 – Comparison of Model Types in Prior Work. . . . .	72
Table 14 – Comparison of Accuracy in Prior Work. . . . .	72
Table 15 – Comparison of Percentage of PD Data in Prior Work. . . . .	72
Table 16 – Demographic and clinical information for the three participant groups, including the duration in months from the onset of symptoms for ALS or from the occurrence of a stroke for PS. . . . .	75
Table 17 – Toronto Neuroface Dataset Distribution. . . . .	76
Table 18 – List of AU, action descriptors, and underlying facial muscles. . . . .	77
Table 19 – Configuration of Grid-Search Space for HC vs. ALS. . . . .	78
Table 20 – Configuration of Grid-Search Space for HC vs. PS. . . . .	78
Table 21 – Logistic Regression - Comparative Performance in HC vs. ALS. . . . .	79
Table 22 – Logistic Regression - Performance in HC vs. PS. . . . .	80
Table 23 – Logistic Regression Analysis - HC vs ALS. . . . .	81

Table 24 – Logistic Regression Results - HC vs PS. . . . .	81
Table 25 – Comparison with prior works. . . . .	84
Table 26 – Comparison with prior works. . . . .	85
Table 27 – Number of repetitions for each subtask in the ALS vs. healthy controls subset	89
Table 28 – Number of repetitions for each subtask in the post-stroke vs. healthy controls subset . . . . .	89
Table 29 – Facial Point Graph results for each subtask in the ALS vs. healthy controls subset . . . . .	93
Table 30 – Facial Point Graph results for each subtask in the post-stroke vs. healthy controls subset . . . . .	93
Table 31 – Normal vs. Not Normal . . . . .	103
Table 32 – Slight vs. Not Slight . . . . .	103
Table 33 – Mild vs. Not Mild . . . . .	103
Table 34 – Moderate vs. Not Moderate . . . . .	103
Table 35 – Classification Report. . . . .	104
Table 36 – Descriptors selected using RFE. . . . .	105
Table 37 – Demographics of the datasets . . . . .	111
Table 38 – Comparison of model performances. . . . .	118
Table 39 – Mean FID values for image quality assessment (the best result is highlighted in bold). . . . .	137
Table 40 – Synthetic versus real images by humans experts. . . . .	138
Table 41 – Comparison between human experts and deep models to classify AMD and real Non-AMD images. . . . .	139
Table 42 – Confusion matrix in my test set. . . . .	140
Table 43 – Confusion matrix in STARE. . . . .	140
Table 44 – Broadcast Links . . . . .	183

# LIST OF ABBREVIATIONS AND ACRONYMS

---

---

AI	Artificial Intelligence
ALS	Amyotrophic Lateral Sclerosis
AMD	Age-related Macular Degeneration
AU	Action Unit
AVTL	Apparent Vocal Tract Length
CGAN	Conditional Generative Adversarial Network
CNN	Convolutional Neural Network
CVD	Chronic Venous Disease
DDK	Diadochokinetic Task
E-Health	Electronic Health
ECG	Electrocardiogram
F0	Fundamental Frequency
FACS	Facial Action Coding System
FID	Fréchet Inception Distance
FPG	Facial Point Graphs
GAT	Graph Attention Network
GNN	Graph Neural Network
GPU	Graphics Processing Unit
HC	Healthy Control
HOG	Histogram of Oriented Gradients
IMU	Inertial Measurement Unit
IoT	Internet of Things
KNN	K-Nearest Neighbors
LLM	Large Language Model
LoRA	Low-Rank Adaptation
LOSO-CV	Leave-One-Subject-Out Cross-Validation
LR	Logistic Regression
MCC	Matthews Correlation Coefficient
MDS-UPDRS	Movement Disorder Society Unified Parkinson's Disease Rating Scale
MFCC	Mel-Frequency Cepstral Coefficient
ML	Machine Learning

NHMRC	National Health and Medical Research Council
NLP	Natural Language Processing
OCT	Optical Coherence Tomography
PD	Parkinson's Disease
PS	Post-Stroke
PwPD	People with Parkinson's Disease
Py-Feat	Python Facial Expression Analysis Toolbox
RF	Random Forest
RFE	Recursive Feature Elimination
SAM	Segment Anything Model
SAMD	Software as a Medical Device
SMOTE	Synthetic Minority Over-sampling Technique
SSIM	Structural Similarity Index Measure
SVM	Support Vector Machine
TTA	Test-Time Augmentation
UPDRS	Unified Parkinson's Disease Rating Scale
VAE	Variational Auto-Encoder
VLU	Venous Leg Ulcer
VTL	Vocal Tract Length
WHO	World Health Organization

# CONTENTS

---

---

<b>1</b>	<b>INTRODUCTION</b>	<b>19</b>
<b>1.1</b>	<b>Objectives</b>	<b>21</b>
<b>1.1.1</b>	<b><i>Research Questions</i></b>	<b>21</b>
<b>1.1.2</b>	<b><i>Hypothesis</i></b>	<b>21</b>
<b>1.2</b>	<b>Thesis Structure</b>	<b>22</b>
<b>1.2.1</b>	<b><i>Publications</i></b>	<b>25</b>
<b>2</b>	<b>LITERATURE REVIEW</b>	<b>29</b>
<b>2.1</b>	<b>Video analysis in Healthcare</b>	<b>29</b>
<b>2.1.1</b>	<b><i>Facial expression for ALS Detection</i></b>	<b>30</b>
<b>2.1.2</b>	<b><i>Facial expression for Post-Stroke Detection</i></b>	<b>31</b>
<b>2.1.3</b>	<b><i>Facial expression for Parkinson Detection</i></b>	<b>32</b>
<b>2.1.3.1</b>	<b><i>Methods</i></b>	<b>34</b>
<b>2.1.3.1.1</b>	Relevant works by 2019	34
<b>2.1.3.1.2</b>	Literature search	36
<b>2.1.3.1.3</b>	Study selection	37
<b>2.1.3.1.4</b>	Data extraction	37
<b>2.1.3.2</b>	<b><i>Results</i></b>	<b>37</b>
<b>2.1.3.3</b>	<b><i>Discussion</i></b>	<b>43</b>
<b>2.1.3.3.1</b>	General Observations	43
<b>2.1.3.3.2</b>	Traditional Machine Learning	43
<b>2.1.3.3.3</b>	Deep Learning Approaches	49
<b>2.1.3.3.4</b>	Facial Video Datasets for PD	50
<b>2.1.3.3.5</b>	Technology Readiness for Clinical Translation	52
<b>2.1.3.4</b>	<b><i>Conclusion</i></b>	<b>54</b>
<b>2.1.4</b>	<b><i>CRedit authorship contribution statement</i></b>	<b>55</b>
<b>2.2</b>	<b>Voice analysis in Healthcare</b>	<b>55</b>
<b>2.2.1</b>	<b><i>Diadochokinetic tasks</i></b>	<b>55</b>
<b>2.2.2</b>	<b><i>Vocal Tract length</i></b>	<b>56</b>
<b>2.3</b>	<b>Image analysis in Healthcare</b>	<b>57</b>
<b>2.3.1</b>	<b><i>Venous Leg Ulcer</i></b>	<b>58</b>
<b>2.3.2</b>	<b><i>Age-related macular degeneration</i></b>	<b>59</b>

<b>3</b>	<b>VIDEO ANALYSIS FOR NEUROLOGICAL CONDITIONS</b>	<b>60</b>
<b>3.1</b>	<b>Data Augmentation for Facial Expression</b>	<b>61</b>
<b>3.1.1</b>	<b>Methods</b>	<b>62</b>
3.1.1.1	<i>Dataset</i>	62
3.1.1.2	<i>Hyperparameter Optimization</i>	63
3.1.1.3	<i>Augmentation</i>	64
3.1.1.4	<i>Test-Time Augmentation</i>	65
<b>3.1.2</b>	<b>Results</b>	<b>65</b>
3.1.2.1	<i>Model Tuning Assessment</i>	65
3.1.2.2	<i>Data Augmentation Assessment</i>	68
3.1.2.3	<i>Test-Time Augmentation Assessment</i>	68
<b>3.1.3</b>	<b>Discussion</b>	<b>69</b>
3.1.3.1	<i>Novelty of this study</i>	71
3.1.3.2	<i>Limitations</i>	72
<b>3.1.4</b>	<b>Conclusion</b>	<b>73</b>
<b>3.1.5</b>	<b>CRediT authorship contribution statement</b>	<b>73</b>
<b>3.2</b>	<b>Action Units for Facial Expression</b>	<b>74</b>
<b>3.2.1</b>	<b>Methods</b>	<b>75</b>
3.2.1.1	<i>Dataset</i>	75
3.2.1.2	<i>Feature Extraction</i>	75
3.2.1.3	<i>Classification</i>	77
<b>3.2.2</b>	<b>Results</b>	<b>78</b>
3.2.2.1	<i>Feature importance</i>	79
<b>3.2.3</b>	<b>Discussion</b>	<b>82</b>
3.2.3.1	<i>Novelty of this study</i>	83
3.2.3.2	<i>HC vs ALS</i>	83
3.2.3.3	<i>HC vs PS</i>	85
<b>3.2.4</b>	<b>Conclusion</b>	<b>86</b>
<b>3.2.5</b>	<b>CRediT authorship contribution statement</b>	<b>87</b>
<b>3.3</b>	<b>Graph Neural Networks for Facial Expression Analysis</b>	<b>87</b>
<b>3.3.1</b>	<b>Methods</b>	<b>88</b>
3.3.1.1	<i>Dataset</i>	88
3.3.1.2	<i>Preprocessing</i>	90
3.3.1.3	<i>Feature Extraction</i>	90
3.3.1.4	<i>Proposed Model</i>	91
3.3.1.5	<i>Classification</i>	92
<b>3.3.2</b>	<b>Results</b>	<b>92</b>
<b>3.3.3</b>	<b>Discussion</b>	<b>94</b>
3.3.3.1	<i>Novelty of this Study</i>	94

3.3.4	<b>Conclusion</b>	95
3.3.5	<b>CRediT authorship contribution statement</b>	95
4	<b>VOICE ANALYSIS FOR MONITORING</b>	96
4.1	<b>Diadochokinetic tasks</b>	96
4.1.1	<b>Methods</b>	98
4.1.1.1	<i>Dataset</i>	98
4.1.1.2	<i>Feature Extraction</i>	99
4.1.1.3	<i>Classification</i>	100
4.1.1.4	<i>Ensemble Modeling</i>	102
4.1.2	<b>Results</b>	102
4.1.3	<b>Discussion</b>	106
4.1.3.1	<i>Novelty of this study</i>	108
4.1.3.2	<i>Limitation</i>	108
4.1.4	<b>Conclusion</b>	109
4.1.5	<b>CRediT authorship contribution statement</b>	109
4.2	<b>The Apparent Vocal Tract length</b>	109
4.2.1	<b>Methods</b>	110
4.2.2	<b>Results</b>	111
4.2.3	<b>Discussion</b>	112
4.2.4	<b>Conclusion</b>	113
4.2.5	<b>CRediT authorship contribution statement</b>	114
4.3	<b>Voice Screening for Chatbot</b>	114
4.3.1	<b>Methods</b>	115
4.3.1.1	<i>Dataset</i>	115
4.3.1.2	<i>Feature Extraction</i>	115
4.3.1.3	<i>Classification and Evaluation</i>	116
4.3.1.4	<i>Chat-bot</i>	116
4.3.2	<b>Experimental Results</b>	117
4.3.2.1	<i>Classification</i>	117
4.3.2.2	<i>Chat-bot</i>	118
4.3.3	<b>Discussion</b>	118
4.3.3.1	<i>Novelty of this study</i>	120
4.3.4	<b>Conclusion</b>	120
4.3.5	<b>CRediT authorship contribution statement</b>	120
5	<b>AI-POWERED SYNTHETIC IMAGING</b>	121
5.1	<b>Stable Diffusion for Thermal Image Estimation</b>	121
5.1.1	<b>Methods</b>	123
5.1.1.1	<i>Dataset</i>	123

---

5.1.1.2	<i>Preprocessing</i>	123
5.1.1.3	<i>Generation</i>	123
5.1.1.4	<i>Prompt</i>	125
5.1.1.5	<i>Evaluation Measures</i>	125
<b>5.1.2</b>	<b>Results</b>	<b>125</b>
<b>5.1.3</b>	<b>Discussion</b>	<b>127</b>
5.1.3.1	<i>Novelty of this study</i>	129
<b>5.1.4</b>	<b>Conclusion</b>	<b>129</b>
<b>5.1.5</b>	<b>CRediT authorship contribution statement</b>	<b>130</b>
<b>5.2</b>	<b>StyleGAN-2 for Synthetic Eye Fundus Image Generation</b>	<b>130</b>
<b>5.2.1</b>	<b>Methods</b>	<b>131</b>
5.2.1.1	<i>Dataset</i>	131
5.2.1.2	<i>Evaluation Measures</i>	134
5.2.1.3	<i>Experimental Setup</i>	134
<b>5.2.2</b>	<b>Results</b>	<b>136</b>
5.2.2.1	<i>Synthetic Image Assessment</i>	136
5.2.2.2	<i>Distinguishing between Synthetic and Real Images</i>	137
5.2.2.3	<i>Data Augmentation Assessment</i>	138
5.2.2.4	<i>Comparison between Human Experts and Deep Models</i>	139
5.2.2.5	<i>Web Application</i>	139
5.2.2.6	<i>Model Generalization</i>	140
<b>5.2.3</b>	<b>Discussion</b>	<b>140</b>
5.2.3.1	<i>Novelty of this study</i>	142
<b>5.2.4</b>	<b>Conclusion</b>	<b>142</b>
<b>5.2.5</b>	<b>Data availability</b>	<b>143</b>
<b>5.2.6</b>	<b>Code availability</b>	<b>143</b>
<b>5.2.7</b>	<b>CRediT authorship contribution statement</b>	<b>143</b>
<b>6</b>	<b>CONCLUSION</b>	<b>145</b>
<b>6.1</b>	<b>Video Analysis for Neurological Conditions</b>	<b>145</b>
<b>6.2</b>	<b>Voice Analysis for Remote Monitoring</b>	<b>146</b>
<b>6.3</b>	<b>AI-Powered Synthetic Imaging</b>	<b>147</b>
<b>6.4</b>	<b>Summary of Innovations and Work Implications</b>	<b>148</b>
<b>6.4.1</b>	<b><i>Ethical Implications, Challenges, and Risks</i></b>	<b>148</b>
<b>6.5</b>	<b>Future Work</b>	<b>149</b>
	<b>BIBLIOGRAPHY</b>	<b>150</b>
	<b>APPENDIX A MEDIA COVERAGE</b>	<b>181</b>

**APPENDIX B            ETHICS APPROVAL LETTER . . . . . 186**

---

# INTRODUCTION

---

With advancements in artificial intelligence (AI), healthcare has emerged as one of the most significantly impacted sectors. AI-driven techniques, particularly in medical image analysis, have shown great promise in enhancing diagnostic accuracy, improving treatment planning, and optimizing patient outcomes by addressing challenges like data scarcity and interpretability in clinical workflows (PILLAI, 2021). In non-invasive diagnostics, AI plays a crucial role, particularly in the early detection and management of diseases, as seen in assessing coronary artery disease, where AI aids in offering reliable, low-risk diagnostic solutions (DOOLUB *et al.*, 2023). Additionally, AI holds the potential to bridge healthcare delivery gaps, especially in resource-poor settings, where healthcare worker shortages are prevalent, by enabling innovative tools and systems to extend care to underserved populations (WAHL *et al.*, 2018).

The application of AI in healthcare has been a rapidly expanding field, with numerous studies emphasizing its transformative potential in improving patient care (KASULA, 2024). For example, in radiology, AI has enhanced the accuracy of image interpretation by aiding in identifying complex patterns (LOEHFELM, 2021). In dermatology, AI has demonstrated substantial capabilities in detecting skin cancer (LIOPYRIS *et al.*, 2022). AI algorithms have been successfully employed in cardiology for arrhythmia detection and predicting cardiac events (FEENY *et al.*, 2020). AI has also shown promise in neurology, where AI-driven gait analysis algorithms have effectively monitored motor symptoms in Parkinson's disease, although challenges remain in achieving widespread clinical validation (BIASE *et al.*, 2020). Similarly, in the context of cognitive disorders, AI tools analyzing speech patterns have demonstrated potential in detecting early signs of Alzheimer's disease, highlighting its role in early, non-invasive diagnostic methods (LUZ *et al.*, 2021). Despite these advancements, significant challenges remain, such as data security, ethical concerns, and ensuring AI systems are reliable across diverse patient demographics. Addressing these challenges is crucial for successfully integrating AI into clinical practice (KHAN *et al.*, 2023).

While significant progress has been made in utilizing AI for medical diagnostics, several

challenges remain. Current AI systems often face integration limitations within healthcare settings due to concerns about infrastructure compatibility, ethical considerations, and a general lack of AI literacy among healthcare professionals, hindering widespread adoption (ESMAEILZADEH, 2024). Ongoing research is focused on enhancing personalized patient communication, improving remote monitoring, and increasing treatment adherence through AI-driven tools (NOVA, 2023). There is also a pressing need for comprehensive studies that extend beyond task-specific diagnostics to broader applications in Telemedicine and E-Health (KADU; SINGH, 2021). This thesis will explore the impact of AI-assisted techniques across three key modalities—video, voice, and image—on clinical decision-making and patient outcomes, focusing on conditions like neurological disorders, ophthalmic diseases, and leg ulcers. The goal is to highlight AI's potential to transform modern healthcare.

One focus of this thesis is the potential of AI-assisted facial expression analysis in improving interactions between clinicians and patients, particularly for conditions like Parkinson's disease, Stroke, and Amyotrophic Lateral Sclerosis (ALS) (QIANG *et al.*, 2022). Current diagnostic practices often rely on subjective assessments, which can be inconsistent and time-consuming (IENCA; IGNATIADIS, 2020). By developing and implementing AI models that analyze facial expressions to detect signs of these neurological conditions, this study seeks to provide clinicians with a reliable, objective tool that enhances diagnostic accuracy and facilitates timely interventions. This study will also explore the utility of mobile applications for first responders, enabling rapid identification and management of post-stroke patients in emergency settings.

Another key area of investigation is the role of AI-based voice analysis tools in remote monitoring and ongoing care for Parkinson's disease. As the disease progresses, continuous monitoring becomes crucial for effective management (ARORA *et al.*, 2015). Traditional methods are often invasive and inconvenient for patients. This thesis aims to develop AI models that can analyze voice recordings to assess disease severity, offering a non-invasive, accessible alternative for regular monitoring. Additionally, the integration of language models and chatbots will be explored to provide patients with real-time feedback and support, thereby improving adherence to treatment plans and enhancing overall care quality.

Lastly, this thesis investigates the application of AI-powered deep learning techniques for synthetic imaging in medical field. Specifically, this thesis will develop methods to create synthetic images for conditions such as age-related macular degeneration and venous leg ulcers. By using data augmentation techniques and advanced models like stable diffusion architectures, the study aims to improve the accuracy and reliability of diagnostic tools. The thesis will also develop web-based tools to make these advanced diagnostic techniques more accessible to clinicians and patients, ultimately enhancing the quality of clinician-patient interactions and supporting non-invasive screening efforts.

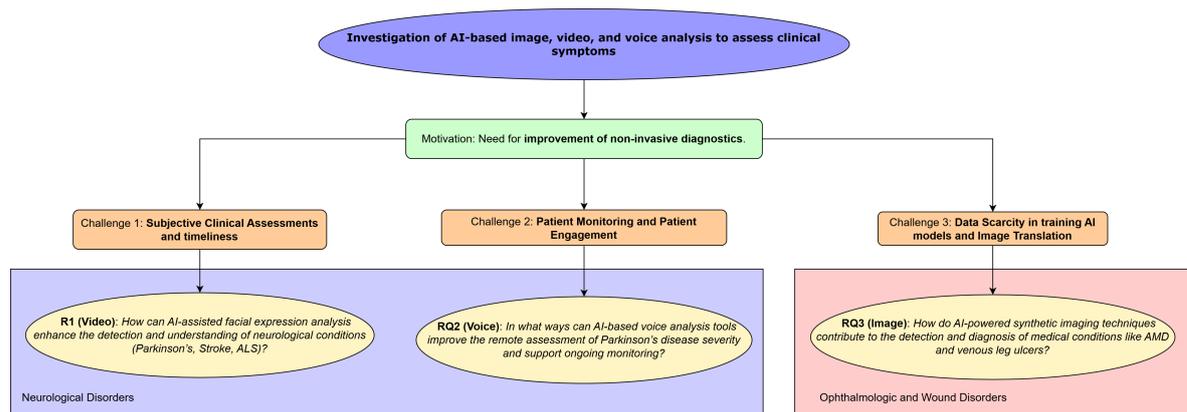


Figure 1 – Thesis Contribution.

## 1.1 Objectives

The objective of this thesis is to investigate and develop AI-assisted techniques across different modalities— video, voice, and image —to enhance diagnostic accuracy, improve clinical decision-making, and elevate patient outcomes in the fields of neurology, ophthalmology, and wound care. This research aims to bridge the gap between current AI capabilities and practical, non-invasive medical applications by focusing on conditions such as Parkinson’s disease, stroke, ALS, age-related macular degeneration, and venous leg ulcers.

### 1.1.1 Research Questions

This work seek to achieve the objectives of this work by answering the following research questions:

1. **(Video) How can AI-assisted facial expression analysis enhance the detection and understanding of neurological conditions such as Parkinson’s disease, stroke, and ALS?**
2. **(Voice) In what ways can AI-based voice analysis tools improve the remote assessment of Parkinson’s disease severity and support ongoing monitoring?**
3. **(Image) How do AI-powered synthetic imaging techniques contribute to the detection and diagnosis of medical conditions like age-related macular degeneration and venous leg ulcers?**

Figure 1 presents a diagram showing the main motivation and challenges behind each research question.

### 1.1.2 Hypothesis

This research has been conducted based on the following hypothesis:

**1. Facial Action Unit Analysis for Neurological Conditions:** AI-assisted analysis of facial action units can distinguish patients with neurodegenerative diseases—such as Parkinson’s disease, amyotrophic lateral sclerosis, and stroke—from healthy individuals. This system will lead to better understanding of facial expression impairments associated with these conditions.

**2. Voice Analysis Using Diadochokinetic (DDK) Tasks:** AI-based voice analysis utilizing Diadochokinetic (DDK) tasks provides an effective, non-invasive method for assessing disease severity in Parkinson’s patients. Employing an ensemble of different DDK tasks is hypothesized to reduce overfitting and enhance the accuracy of remote monitoring.

**3. Integration of Large Language Models in Voice Analysis:** Incorporating Large Language Models (LLMs) as agents to initiate AI-based voice analysis tools is expected to improve the remote identification of Parkinson’s disease symptoms. This integration has the potential to enhance patient engagement and adherence to treatment plans.

**4. Synthetic Imaging for Age-Related Macular Degeneration Detection:** Generating synthetic images through data augmentation techniques is expected to improve the performance of deep learning models in automatically detecting signs of age-related macular degeneration. This enhancement will lead to more accurate and reliable diagnostic tools.

**5. Image Translation for Venous Leg Ulcer Assessment:** Deep learning-based image translation methods is expected to generate synthetic thermal images from RGB images. This capability can enhance the assessment and monitoring of venous leg ulcers, providing a non-invasive and accessible diagnostic aid.

## 1.2 Thesis Structure

This thesis is organized into six main chapters, each addressing a specific AI-assisted technique applied to various medical diagnostics and patient care modalities, including video, voice, and image analysis. The following is a detailed outline of each chapter:

### ***Chapter 2: Literature Review***

This chapter presents a brief literature review along with one systematic literature review on the application of video, voice, and image analysis techniques within the healthcare field. Specifically, Section 2.1 focuses on video analysis in healthcare, highlighting facial expression analysis for ALS detection (Section 2.1.1), post-stroke detection (Section 2.1.2), and Parkinson’s disease detection (Section 2.1.3).

The systematic literature review presented in Section 2.1.3 systematically evaluates and synthesizes machine learning and deep learning techniques applied to Parkinson’s disease detection through emotional facial expressions. This review includes methods of literature search, study selection criteria, data extraction procedures, as well as the results obtained. It also covers

general observations, traditional machine learning approaches, deep learning methodologies, available facial video datasets for Parkinson's disease, and considerations regarding technology readiness for clinical translation.

The chapter also briefly covers voice applications in healthcare (Section 2.2), addressing tasks like diadochokinetic tests (Section 2.2.1) and vocal tract length assessment (Section 2.2.2). Additionally, image analysis techniques in healthcare are reviewed in Section 2.3, focusing particularly on diagnosing venous leg ulcers (Section 2.3.1) and age-related macular degeneration (Section 2.3.2).

### ***Chapter 3: Video Analysis for Neurological Conditions***

Chapter 3 explores the application of machine learning and deep learning techniques for analyzing facial expressions in patients with neurological conditions, such as Parkinson's disease, ALS, and post-stroke impairments. It begins with hypomimia identification in Parkinson's disease. A promising approach involved using action units, but one of the main limitations was the small dataset. To overcome this, a data augmentation strategy was introduced to enhance model performance. Building on this foundation, the action unit-based approach was extended to other conditions that affect facial movement, such as stroke and ALS. The chapter concludes by introducing Graph Neural Networks (GNNs) as a method for capturing facial patterns through graph-based representations of facial landmarks. This chapter is divided into three main sections:

- **Section 3.1: Improving Facial Action Unit analysis for Parkinson's Disease Hypomimia Detection**

This section refines facial action unit analysis using data augmentation on a tabular dataset to address small data limitations. These enhancements improve the accuracy and robustness of AI models in detecting subtle facial changes associated with Parkinson's disease.

- **Section 3.2: Facial Action Unit Analysis for Post-Stroke and ALS**

This section details the development of AI models capable of analyzing facial action units to objectively assess facial expression impairments in Post-Stroke and ALS. The analysis aims to enhance diagnostic accuracy for neurological conditions that affect facial expressions, offering paramedics a screening tool for timely intervention.

- **Section 3.3: Graph Neural Networks for Post-Stroke and ALS**

In this section, facial expressions affected by post-stroke impairments and ALS are analyzed using graph-based techniques. Facial landmarks are modeled as graphs and processed with Graph Neural Networks (GNNs) to capture spatial and dynamic patterns. As a co-author on this work, contributions included assisting in the project design, validating the experimental procedures, and reviewing the manuscript.

## ***Chapter 4: Voice Analysis for Remote Monitoring***

Chapter 4 focus in AI-driven voice analysis as a tool for evaluating disease severity in Parkinson’s disease. It explores innovative methods for leveraging vocal features to monitor patient health remotely. The chapter is divided into three main sections:

- **Section 4.1: Diadochokinetic (DDK) Task Analysis**

This section introduces a methodology for assessing Parkinson’s disease severity based on Diadochokinetic (DDK) tasks. An ensemble approach to different DDK tasks is discussed, demonstrating how it can be used to identify severity in the Parkinson’s disease group.

- **Section 4.2: Apparent Vocal Tract**

This section investigates how Parkinson’s disease alters vocal tract control. Using formant analysis of the phoneme /a/ from established datasets, the study compares the apparent vocal tract length between PD patients and healthy controls. Results indicate that PD is associated with an increased vocal tract length—especially notable among male participants—although limited sample size prevents deeper evaluation of confounding factors. As a co-author on this work, contributions included assisting in the project design and reviewing the manuscript.

- **Section 4.3: Integration of Large Language Models and Chatbots**

This section outlines the integration of a large language model (LLM)-powered agent with a machine learning model trained on the apparent vocal tract length feature discussed in Section 4.2. The agent leverages these voice analysis insights to assess Parkinson’s disease markers, while a chatbot interface provides real-time feedback and fosters enhanced patient engagement.

## ***Chapter 5: AI-Powered Synthetic Imaging***

Chapter 5 investigates the use of state-of-the-art generative models for creating synthetic medical images to improve diagnostic accuracy in the fields of ophthalmology and wound care. The chapter is divided into two main sections:

- **Section 5.1: Image Translation with Stable Diffusion for Wound Assessment**

This section explores the use of Stable Diffusion models to generate thermal images from RGB images for the assessment of venous leg ulcers.

- **Section 5.2: Data Augmentation with StyleGAN-2 for Ophthalmology**

This section details the application of StyleGAN-2 for generating synthetic images to enhance the detection of age-related macular degeneration. The study examines how synthetic data can improve the performance of deep learning models. Web-based system was developed.

## Chapter 6: Conclusion and Future Work

The final chapter summarizes the findings and contributions from the preceding chapters, highlighting the implications for clinical practice. The chapter emphasizes the transformative potential of AI in modern healthcare, particularly in enhancing diagnostic accuracy and facilitating patient-clinic interactions through non-invasive methods. Additionally, this chapter addresses the limitations of the current research and suggests avenues for future studies to further advance AI-assisted techniques in healthcare, with a focus on non-invasive diagnostics and improved patient outcomes.

### 1.2.1 Publications

This thesis resulted in **12 publications**, including **six first-author articles published in Q1 journals**, among them prestigious journals such as *ACM Computing Surveys* (Impact Factor: 23.8; ranked 1/143 in Computer Science Theory & Methods), *Computer Methods and Programs in Biomedicine*, *Biomedical Signal Processing and Control*, *Digital Biomarkers*, and *Computers in Biology and Medicine*. Additionally, **two first-author conference papers** were presented at the International Symposium on Computer-Based Medical Systems, alongside three collaborative conference papers and one collaborative journal publication.

Therefore, the contributions of these publications, listed below, reflect significant advancements in the application of video analysis, voice processing, and synthetic imaging technologies in healthcare, highlighting their clinical relevance and potential impact:

1. **Facial expression analysis in Parkinson's disease using machine learning: a review.** OLIVEIRA, G.C., Ngo, Q.C., Passos, L.A., Jodas, D.S., Papa, J.P. and Kumar, D., 2025. *ACM Computing Surveys*, 57(8), pp. 1–25.
2. **Tabular data augmentation for video-based detection of hypomimia in Parkinson's disease.** OLIVEIRA, G.C., Ngo, Q.C., Passos, L.A., Papa, J.P., Jodas, D.S. and Kumar, D., 2023. *Computer Methods and Programs in Biomedicine*, 240, p. 107713.
3. **Facial expressions to identify post-stroke: a pilot study.** OLIVEIRA, G.C., Ngo, Q.C., Passos, L.A., OLIVEIRA, L.S., Papa, J.P. and Kumar, D., 2024. *Computer Methods and Programs in Biomedicine*, 250, p. 108195.
4. **Video assessment to detect amyotrophic lateral sclerosis.** OLIVEIRA, G.C., Ngo, Q.C., Passos, L.A., OLIVEIRA, L.S., Stylianou, S., Papa, J.P. and Kumar, D., 2024. *Digital Biomarkers*, 8(1), pp. 171–180.
5. **Facial point graphs for stroke identification.** Gomes, N.B., Yoshida, A., de OLIVEIRA, G.C., Roder, M. and Papa, J.P., 2023, November. In: *Iberoamerican Congress on Pattern Recognition (CIARP)*, Cham: Springer Nature Switzerland, pp. 685–699.

6. **Facial point graphs for amyotrophic lateral sclerosis identification.** Gomes, N.B., Yoshida, A., Roder, M., OLIVEIRA, G.C. and Papa, J.P., 2024. In: *19th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2024) – Volume 3: VISAPP*, pp. 207–214.
7. **A pilot study for speech assessment to detect the severity of Parkinson’s disease: an ensemble approach.** OLIVEIRA, G.C., Pah, N.D., Ngo, Q.C., Yoshida, A., Gomes, N.B., Papa, J.P. and Kumar, D., 2025. *Computers in Biology and Medicine*, 185, p. 109565.
8. **NestNeuro: leveraging chatbots for vocal screening.** OLIVEIRA, G.C., Pah, N.D., Ngo, Q.C., Papa, J.P. and Kumar, D., 2024, June. In: *2024 IEEE 37th International Symposium on Computer-Based Medical Systems (CBMS)*, IEEE, pp. 182–185.
9. **The change of vocal tract length in people with Parkinson’s disease.** Pah, N.D., Motin, M.A., OLIVEIRA, G.C. and Kumar, D.K., 2023, July. In: *2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, IEEE, pp. 1–4.
10. **Robust deep learning for eye fundus images: bridging real and synthetic data for enhancing generalization.** OLIVEIRA, G.C., Rosa, G.H., Pedronette, D.C., Papa, J.P., Kumar, H., Passos, L.A. and Kumar, D., 2024. *Biomedical Signal Processing and Control*, 94, p. 106263.
11. **A stable diffusion approach for RGB to thermal image conversion for leg ulcer assessment.** OLIVEIRA, G.C., Ngo, Q.C., Papa, J.P. and Kumar, D., 2024, June. In: *2024 IEEE 37th International Symposium on Computer-Based Medical Systems (CBMS)*, IEEE, pp. 158–163.
12. **Screening major depressive disorder in patients with obstructive sleep apnea using single-lead ECG recording during sleep.** Shaw, V., Ngo, Q.C., Pah, N.D., OLIVEIRA, G.C., Khandoker, A.H., Mahapatra, P.K., Pankaj, D. and Kumar, D.K., 2024. *Health Informatics Journal*, 30(4), p. 14604582241300012.

Table 1 summarizes the research studies by organizing them under specific research questions, chapters, and published study titles. All works comprising each chapter of the thesis were primarily the result of my own research, with the exception of Section 3.3 and Section 4.2, in which I contributed as a co-author. In the study described in Section 3.3, I was involved in experimental validation, and manuscript review. In contrast, in the study presented in Section 4.2, I participated in project design and manuscript review. Additionally, I collaborated on another project that fell outside the scope of the thesis.

Research Question	Chapter	Study Title
	Chapter 2	<b>Facial Expression Analysis in Parkinson’s Disease Using Machine Learning: A Review</b>
RQ1	Chapter 3.1	<b>Tabular data augmentation for video-based detection of hypomimia in Parkinson’s disease</b>
	Chapter 3.2	<b>Facial expressions to identify post-stroke: A pilot study</b>
	Chapter 3.2	<b>Video Assessment to Detect Amyotrophic Lateral Sclerosis</b>
	Chapter 3.3	Facial point graphs for stroke identification
	Chapter 3.3	Facial Point Graphs for Amyotrophic Lateral Sclerosis Identification
RQ2	Chapter 4.1	<b>Speech Assessment for Detecting the Severity of Parkinson’s Disease: An Ensemble Approach</b>
	Chapter 4.2	The Change of Vocal Tract Length in People with Parkinson’s Disease
	Chapter 4.3	<b>NestNeuro: Leveraging Chatbots for Vocal Screening</b>
RQ3	Chapter 5.1	<b>Robust deep learning for eye fundus images: Bridging real and synthetic data for enhancing generalization</b>
	Chapter 5.2	<b>A Stable Diffusion Approach for RGB to Thermal Image Conversion for Leg Ulcer Assessment</b>
		Screening Major Depressive Disorder in Patients with Obstructive Sleep Apnea Using Single-Lead ECG Recording During Sleep

Table 1 – Summary of Research Studies.

## Appendix A: Media Coverage

Appendix A highlights the extensive media attention the thesis research has received, exemplified by the pilot study titled **Facial Expressions to Identify Post-Stroke**. This investigation has been featured by major media outlets including 9News, 7News, ABC, RMIT

News, *The Conversation*, and Brazilian *Forbes*, and has achieved an Altmetric score of 319. Among its notable contributions is an AI-powered smartphone application that leverages facial cues—particularly smiling—to rapidly detect signs of stroke, thereby potentially speeding up emergency treatment. This appendix also includes a concise summary of other related projects, including a study on **ALS detection through facial analysis** and a **systematic literature review** on using facial expression assessments for Parkinson’s disease.

### ***Appendix B: Ethics Approval Letter***

Appendix B addresses the vital issue of ethical compliance, presenting the Ethics Approval Letter and detailing the rigorous review process undertaken by an ethics committee. This section underscores adherence to the National Statement on Ethical Conduct in Human Research (NHMRC, 2007) and other relevant standards, as well as the proactive measures implemented to safeguard participants’ rights, well-being, and privacy. Readers are advised to consult the complete Ethics Approval Letter, provided at the end of this appendix, for a full account of the protocols and guidelines governing the research.

---

## LITERATURE REVIEW

---

This chapter provides a review of existing literature on video, voice, and image analysis techniques within healthcare. Specifically, it discusses the application of video analysis for detecting neurological diseases, voice analysis for monitoring disorders such as Parkinson's disease, and image analysis using synthetic data to diagnose medical conditions like venous leg ulcers and age-related macular degeneration (AMD).

Importantly, Section 2.1.3 features a systematic literature review titled "**Facial Expression Analysis in Parkinson's Disease Using Machine Learning: A Review**", published in the *ACM Computing Surveys* journal. This comprehensive review systematically evaluates and synthesizes current methodologies employing machine learning and deep learning approaches for analyzing emotional facial expressions to detect Parkinson's disease. The section examines existing literature, compares methods, discusses the associated challenges, and identifies research gaps to guide future studies in this specialized area.

### 2.1 Video analysis in Healthcare

Video analysis is becoming an important tool in healthcare because it allows for non-invasive, real-time monitoring and diagnosis of various medical conditions. The advent of advanced machine learning techniques, particularly deep learning, has facilitated the extraction of meaningful patterns from video data, which can be applied to a wide range of healthcare problems (FARHAD *et al.*, 2023; KOLARIK *et al.*, 2023). This subsection provides an overview of the applications of video analysis in healthcare, focusing on the detection and monitoring of neurological diseases such as Amyotrophic Lateral Sclerosis (ALS), stroke, and Parkinson's disease through facial expression analysis.

Video analysis in healthcare has diverse applications, including diagnostic, monitoring, and therapeutic uses. Diagnostic applications focus on identifying medical conditions through

visual symptoms like abnormal gait or facial expressions, thus accelerating diagnoses and easing clinical workloads (FARHAD *et al.*, 2023). Monitoring applications involve continuous observation of disease progression, particularly beneficial in chronic neurological conditions like Parkinson's disease, where video-based tracking of symptoms aids personalized treatment (YOLCU *et al.*, 2019). Therapeutic applications employ real-time video analysis, often combined with IoT and edge computing technologies, providing immediate feedback for rehabilitation exercises and enhancing remote patient care (RAJAVEL *et al.*, 2022; IKECHUKWU; WANG, 2024).

Facial expression analysis is a specific area of video analysis that is used in health-care (YOLCU *et al.*, 2019). The Facial Action Coding System (FACS), developed by Ekman and Friesen in the 1970s, provides a comprehensive framework for categorizing facial movements by their appearance on the face. FACS has been instrumental in the development of automated systems for facial expression recognition. FACS is a widely used system for describing facial movements in terms of action units (AUs). Each AU corresponds to a specific facial muscle movement. FACS provides a standardized method for recording and interpreting facial expressions, making it an invaluable tool in both psychological and medical research. With advancements in computer vision and machine learning, automated systems for facial expression recognition have been developed. These systems use algorithms to detect and interpret facial movements, often using FACS as a reference framework. Such systems can be used to identify emotional states, pain levels, or neurological impairments based on facial expressions.

### **2.1.1 Facial expression for ALS Detection**

The computerized analysis of facial expressions in ALS has gained significant attention in recent years as a means to enhance diagnostic accuracy and monitor disease progression. Bandini *et al.* (2018c) have made notable contributions by developing a marker-less video-based approach to assess facial movements in ALS patients. Using a depth sensor, they were able to capture and analyze kinematic features of lip movements during both speech and non-speech tasks. This work emphasizes the potential of non-invasive, video-based techniques in overcoming the limitations associated with traditional motion tracking systems, which are often expensive and less accessible in routine clinical practice.

Beyond motor deficits, ALS also impacts cognitive and emotional functions. Aho-Özhan *et al.* (2016) studied emotional processing in ALS patients, finding diminished recognition of negative emotions like disgust and fear, with altered brain activity in areas related to social emotions. These findings suggest that ALS affects both motor and emotional processing, with possible modulation by social interactions.

Similarly, Zimmerman *et al.* (2007) examined the cognitive deficits associated with frontal lobe dysfunction in ALS, particularly in patients with bulbar involvement who often exhibit emotional lability. Their study found significant impairments in the recognition of facial

emotions, independent of depressive or dementia symptoms. This reinforces the notion that ALS affects multiple cognitive domains, including emotional perception, and suggests that these deficits can occur even in the absence of other cognitive impairments.

[Oh et al. \(2016\)](#) expanded the understanding of emotion perception deficits in ALS by investigating a cohort of Korean patients. Their findings confirmed that ALS patients have significant difficulties in recognizing facial emotions, particularly those associated with sadness and fear. This study further supports the view that ALS is not solely a motor disorder but also involves cognitive and emotional dysfunctions, which are critical to consider in both diagnosis and management.

In 2024, one approach leveraged accessible video capture and standard machine learning workflows to refine ALS detection through facial symmetry and landmark analysis. For instance, [Suárez-Hernández<sup>1</sup> et al. \(2024\)](#) proposed to identify ALS via facial-symmetry analysis. Key facial landmarks were extracted and converted to spherical coordinates to minimize spatial errors. Supervised ML models trained on these features achieved performance metrics—with accuracy, specificity, and sensitivity reported at 66.7%, 50.0%, and 87.5%, respectively—highlighting the potential for earlier and more efficient diagnosis with accessible equipment.

### 2.1.2 Facial expression for Post-Stroke Detection

Stroke detection and assessment pose significant challenges, particularly in emergency settings where rapid diagnosis is critical. Stroke often affects facial symmetry and muscle control, making facial expression analysis a valuable tool for early detection. Early detection of stroke is crucial for effective treatment. Delays in diagnosis can result in significant brain damage and long-term disability. Rapid assessment tools are needed, especially in pre-hospital settings, to ensure timely intervention. Stroke can cause asymmetry in facial movements, such as drooping of one side of the face. Analyzing these asymmetries through video analysis can provide early indications of stroke.

[Kaewmahanin et al. \(2022\)](#) proposed using cosine similarity between the left and right sides of the face to detect facial asymmetry, achieving a high classification accuracy of 97.9% on the Toronto NeuroFace dataset, highlighting its potential for elderly stroke detection. [Ipapo et al. \(2023\)](#) further explored facial motion analysis in stroke and ALS patients, using facial landmarks to assess orofacial dysfunction severity with random forest classifiers, achieving results close to baseline models in various accuracy metrics.

[Naeini et al. \(2022\)](#) investigated orofacial assessment videos, comparing traditional landmark-based methods with a deep learning model, RepNet, for detecting periodicity. RepNet outperformed the traditional approach, effectively distinguishing between healthy individuals and ALS patients. [Parra-Dominguez, Sanchez-Yanez and Garcia-Capulin \(2021\)](#) focused on detecting facial paralysis using a photograph-based system, employing a multi-layer perceptron

classifier that demonstrated high accuracy on public datasets.

[Bandini et al. \(2018a\)](#) developed a video-based approach for assessing orofacial impairments in stroke survivors. By using depth sensors and face alignment algorithms, they achieved 87% accuracy in differentiating stroke patients from controls, supporting the feasibility of an objective assessment tool for clinical use.

Recent advances in computer-aided stroke detection underscore the potential of facial analysis as a rapid diagnostic tool in emergency settings. Stroke often manifests as asymmetrical facial movements and altered muscle control, making objective assessment of facial expressions critical for early intervention. In a pioneering study, [Alsharif et al. \(2024\)](#) demonstrated that deep learning techniques—specifically convolutional neural networks—can detect facial asymmetries indicative of stroke in pre-hospital scenarios, thus minimizing delays that may lead to significant brain damage and long-term disability. Complementary research by [Ranjan et al. \(2024\)](#) proposed a framework integrating computer vision and machine learning to classify facial asymmetry in stroke survivors, achieving promising accuracy and reinforcing the feasibility of objective, video-based assessments in both clinical and emergency environments.

In addition, [Koob et al. \(2024\)](#) investigated the behavioral and neuroanatomical correlates of facial emotion processing in post-stroke depression (PSD), revealing that stroke patients exhibit deficits in recognizing specific emotions—especially happy, sad, and fearful expressions. These deficits are linked to lesions in key emotion-processing regions, such as the inferior and middle frontal gyri, insula, and putamen, and may be further amplified by depressive symptoms.

### **2.1.3 Facial expression for Parkinson Detection**

Parkinson's disease (PD) is a neurodegenerative disease that affects over 1% of adults over 60 years old, with around 4% diagnosed among 50-year-olds or even younger ([REEVE; SIMCOX; TURNBULL, 2014](#)). There are no easy-to-perform tests or biomarkers for its diagnosis. The disease is multi-symptomatic and is diagnosed based on observations of the symptoms. While there is no cure for the condition ([LEE; YANKEE, 2021](#)), early detection of the disease can help manage the symptoms, and improve the quality of life of the patients.

PD is characterized by a complex set of symptoms, including motor and non-motor symptoms ([POSTUMA et al., 2015; BERARDELLI et al., 2013](#)). In the early stages, motor symptoms may include pain, stiffness or numbness in the limbs, speech impairment (dysarthria), and a reduction in control of facial expressions ([RIZEK; KUMAR; JOG, 2016](#)). Advanced stages may have symptoms such as resting tremor, bradykinesia, gait and speech difficulties, hypophonia, muscle dystrophy, postural deformities, and instability ([JANKOVIC, 2008a](#)). Non-motor symptoms include cognitive impairment, anxiety, drowsiness, speaking irregularities, olfactory dysfunction, sleep problems, constipation, aggression, confusion, and erectile dysfunction ([FULLARD; MORLEY; DUDA, 2017; MAHLKNECHT; SEPPI; POEWE, 2015; SEPPI et](#)

*al.*, 2019) (POEWE *et al.*, 2017).

Individuals with Parkinson's disease frequently exhibit a condition characterized by decreased facial expression, namely hypomimia (POEWE *et al.*, 2017; SVEINBJORNSDOTTIR, 2016). The term, also known as "facial masking" (GOETZ, 2011), characterizes a common symptom even in its early stage (JANKOVIC, 2008a). Patients with hypomimia have considerable difficulties adjusting their facial muscles to display emotional expressions (RICCIARDI *et al.*, 2020), resulting in diminished alterations to their facial expression. They may appear to be disinterested in their surroundings, leading to miscommunication, causing misunderstandings (HO *et al.*, 2020; RICCIARDI *et al.*, 2017), decreased social well-being, and depression (GUNNERY *et al.*, 2016). Experienced clinicians use hypomimia as a biomarker to identify PD but early signs may not be easy to detect.

Advances in artificial intelligence (AI) computer vision and machine learning (ML) have developed methods for machine-based assessment of different symptoms of people with neuro-motor disorders such as PD. Such techniques have been shown to be accurate and also suitable for telehealth applications (GHORAANI *et al.*, 2019; DEMROZI *et al.*, 2019). With the development of AI to recognise facial expressions, an extension for detecting hypomimia has been proposed. Sonawane and Sharma (SONAWANE; SHARMA, 2021) reviewed works in the context of automated facial expression analysis for PD patients, indicating the quick progress in this field. They observed that deep learning was still in the early in 2020. However, since then, there have been significant developments in the field. My work fills this gap and provides an updated account of this area, presenting recent progress while also highlighting the limitations.

- What are the most suitable machine learning approaches for differentiating between people with PD and healthy individuals using videos of facial expressions?
- What are the strengths and limitations of using data augmentation for facial video analysis in detecting PD?
- Is the research field advanced enough for translation into clinical practice? What are the limitations?

Detecting hypomimia can help detect PD in the early stage. It also can help understand the behavior of the individual. It does not require special equipment and can be performed using a camera or smartphone (KATSIKITIS; PILOWSKY, 1988; KAN *et al.*, 2002; PELL; LEONARD, 2005; MERGL *et al.*, 2005; MARSILI *et al.*, 2014) and computerized assessment has the potential for being used for population screening. While this was first proposed in 1988 by Katsikitis and Pilowsky (KATSIKITIS; PILOWSKY, 1988), research on PD identification from facial expressions is a very recent phenomenon, with the majority of the works being published after 2019.

### 2.1.3.1 Methods

This subsection reviews relevant literature on facial analysis in PD up to 2019 and describes the methodology used for a systematic review of studies published since then, focusing specifically on advancements in video-based analysis of facial movements and expressions for clinical applications.

Initial screening in the PubMed dataset led us to observe that the works were mainly organized into three main categories: (i) emotion recognition, (ii) hypomimia degree detection, and (iii) diagnosis by facial expression. Fig. 2 depicts the number of publications from 1967 to 2024.

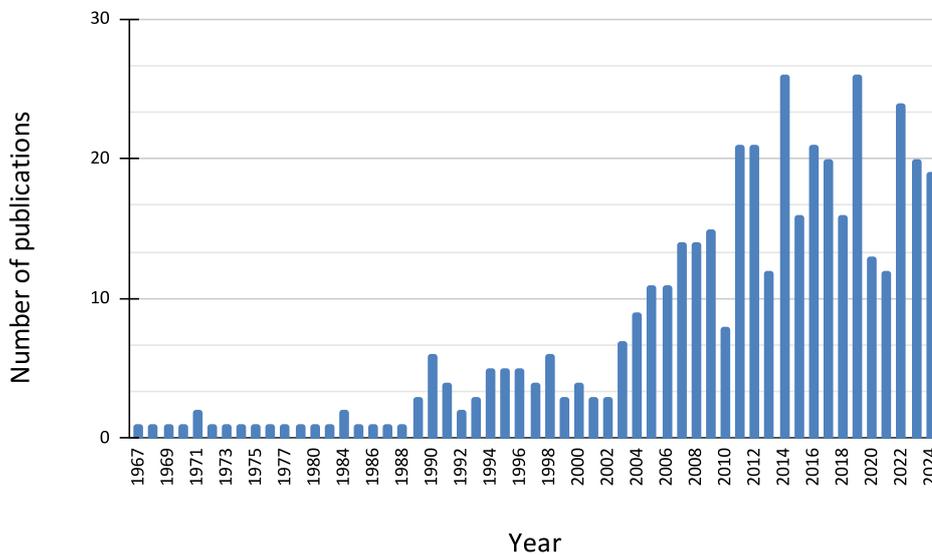


Figure 2 – The number of PubMed publications related to facial expression in Parkinson's disease from 1967 to September 2024. It includes emotion recognition, hypomimia detection, and diagnosis by facial expression.

#### 2.1.3.1.1 Relevant works by 2019

Undetected hypomimia in PD can lead to miscommunication and misunderstandings because the symptom is often associated with indifference or bad behaviour. Such a condition may lead to social isolation and stressful situations with family, friends, and caregivers (TICKLE-DEGNEN; LYONS, 2004). Studies have highlighted the challenges faced by PD patients due to hypomimia. For instance, Simons et al. (SIMONS *et al.*, 2004) investigated facial expressivity in people with PD and found that participants showed reduced spontaneous facial expressivity across various experimental situations. Similarly, Bowers et al. (BOWERS *et al.*, 2006) demonstrated that intentional facial expressions in PD patients are slowed and involve less movement, indicating that both voluntary and spontaneous expressions are impacted.

When the condition is diagnosed, it can lead to the people being informed of the situation and use of alternatives such as 'robot mediators' (ARKIN; SCHEUTZ; TICKLE-DEGNEN, 2014). Briggs et al. (BRIGGS; SCHEUTZ; TICKLE-DEGNEN, 2015) demonstrated the effectiveness of using robotic devices to interview PD patients to overcome some of the limitations due to poor facial expressions of the patients. Hence, management options are available but would be effective when hypomimia has been timely detected.

Early work in the field was conducted by Joshi et al. (JOSHI *et al.*, 2016), who presented a regression model-based technique to measure face expression. The architecture was trained using geometric feature descriptors of face landmarks on video sequences. The model extracts the characteristics by computing the distances between facial landmarks, which are further used for training the regression model. This investigation employed a data collection of 805 video clips from 117 individuals.

Almutiry et al. (ALMUTIRY *et al.*, 2016) conducted longitudinal research on the facial expressions of people with PD. The study investigated 8 individuals, 4 with PD, and 4 healthy for control. Patients were monitored five days per week, once every day, for six weeks, while the control group was monitored for five consecutive days. The participants were asked to make specific facial expressions when the video was recorded. The authors employed two traditional feature extraction methods to localize 27 facial features, i.e., the Active Appearance Model (AAM) and the Constrained Local Model (CLM). The results showed that people with PD have reduced control over their movement.

Bologna et al. (BOLOGNA *et al.*, 2016) noticed a decline in emotional expressiveness on 18 PD patients and 16 healthy individuals (control). The facial expressions were recorded using a 3D optical system, and specific markers were applied to the face. PD group individuals show a decrease in face motor activity during emotional expressiveness. Also, there was a general decrease in recognition of the following feelings in PD patients: disgust, despair, and fear. Additionally, the PD group demonstrated a decrease in the velocity and amplitude of all six basic emotional expressions.

Gunnery et al. (GUNNERY *et al.*, 2017) investigated the coordination of movements across areas of the face in 8 PD patients. The work employed the Facial Action Coding System to comprehensively measure spontaneous facial expression across 600 frames for a multiple case study of people with PD and created a correlation for the frequency and intensity of produced muscle activation across different areas of the face. The findings indicate a decrease in the number, duration, intensity, and coactivation of facial muscle action, while the degree of facial expression deficiency increased.

Another significant work was performed by Bandini et al. (BANDINI *et al.*, 2017), who proposed a video-based automated approach for analyzing facial expressions in PD patients. At the clinician's request and following the imitation of a visual cue on a screen, patients with PD and healthy individuals were instructed to display fundamental facial emotions. The

Euclidean distance of the facial model from a neutral baseline was calculated using an existing face tracker to assess changes in facial expressivity during the tasks. Furthermore, an automated facial expression recognition system was built to investigate how PD expressions varied from conventional expressions. Additionally, Bandini et al. (BANDINI *et al.*, 2017) and Sanchez et al. (ABRAHAM *et al.*, 2017) identified the following PD symptoms; Widened Palpebral Fissures (PF) on the side where symptoms occur, unintentional lip separation (mouth opening), decreased blink rate and nasolabial flattening.

Rajnoha et al. in 2018 verified that (RAJNOHA *et al.*, 2018) PD hypomimia could be recognized from static facial pictures. For this study, they recruited 50 people with PD and 50 age and gender-matched healthy individuals. Among the classifiers, the decision trees achieved the highest accuracy (67.33%). The findings showed that while automatic static face analysis can help with the PD hypomimia diagnosis, it is not as accurate as techniques based on video-recording processing.

Langevin et al. (LANGEVIN *et al.*, 2019) explored video-recording resources. They created PD Analysis with Remote Kinetic-tasks (PARK), which teaches and leads users through six motor activities and one audio task chosen from the standardized MDS-UPDRS rating scale, and then records their performance via camera. Their study attempted to replicate the clinician performing the tests on their patients.

Other notable contributions in that period include the work of Kang et al. (KANG *et al.*, 2019). They investigated whether the orofacial movement in people with PD was impaired and manifested in spontaneous or voluntary expressions. The activation of muscles was investigated using the East Asian Dynamic Facial Expression Stimuli database (LIM *et al.*, 2013), which comprises electromyography signals. They studied a sample of 20 people with PD and 20 healthy individuals and found restrictions in patients' ability to spontaneously express emotions, even though both groups had similar voluntary facial movements. The study also showed how the existence of a "masked face" impaired the patient's quality of life, altering social and psychological elements and raising their chance of developing depression-related symptoms.

#### 2.1.3.1.2 Literature search

A search was primarily conducted for English language publications from 2019 to September 2024 in five online databases; PubMed, IEEE Xplore, Scopus, ACM Digital Library, and Web of Science. The searched articles were imported into Rayyan (ELMAGARMID *et al.*, 2014), a systematic reviews web app, for removing duplicates and exploring and filtering searches for all eligible studies based on the screening of the title and abstract. "Parkinson's facial expression" was the common string for the search, combined with seven words: "detection", "classification", "identification", "recognition", "quantification", "measurement", or "video". This search resulted in 25 articles. The articles were further filtered depending on whether they contained "machine learning" or "deep learning". The following search expression was applied:

“(Parkins\*) AND (facial express\*) AND ((detect\*) OR (recog\*) OR (diag\*) OR (measur\*) OR (quantif\*))”.

Selected papers were then assessed and included based on the criterion described in the next subsection.

#### 2.1.3.1.3 Study selection

The inclusion criteria were: (i) studies reporting an outcome (i.e., quantitative analysis) of PD detection by facial expression, (ii) studies presenting methods to quantifying hypomimia, (iii) studies presenting a clear definition of PD, (iv) studies providing clear information about the database and number of images in the data sets, (v) studies describing the algorithms and procedures used in PD detection, and (vi) English-language publications only. Only full papers were included, while forms, comments, letters to the editors, and editorials were excluded. No constraints were placed on the journal or the length of the paper.

#### 2.1.3.1.4 Data extraction

A posterior analysis was performed to avoid redundancy, eliminating irrelevant and duplicate papers. The remaining articles were evaluated following the inclusion and exclusion criteria, i.e., title, year of publication, authors, study purpose, study type, number of individual participants, procedures, performance metric(s), outcomes, and conclusions. Following the same method, the most recent articles were hand searched, filtered for the current year (i.e., 2024), and submitted to the same inclusion criteria.

#### 2.1.3.2 Results

Initially, 1,315 potential articles (327 from PubMed, 82 from Scopus, 427 from Web of Science, 32 from IEEE Xplore, and 50 from ACM Digital Library) were found to be eligible for inclusion, ending up with 816 after the removal of replications. Further, 635 works were excluded, for they did not fit into the proposed period range. The next step was to filter the articles containing related keywords in the title or abstract, such as “facial expression”, “video”, “image”, “hypomimia”, “machine learning”, or “deep learning”, leading to 63 articles. Another 39 papers were excluded since they did not meet the selection criteria, i.e., articles that only mention PD (14), do not aim at distinguishing between healthy control (HC) group and PD by facial analysis (15), do not employ facial image or video as input (4), do not use machine learning or deep learning techniques (6). In the end, 24 articles met all criteria. One additional article of ArXiv was included after searching bibliographies, summing up to 25 articles using machine learning or deep learning techniques in facial analysis to identify PD. Figure 3 explains the aforementioned selection process.

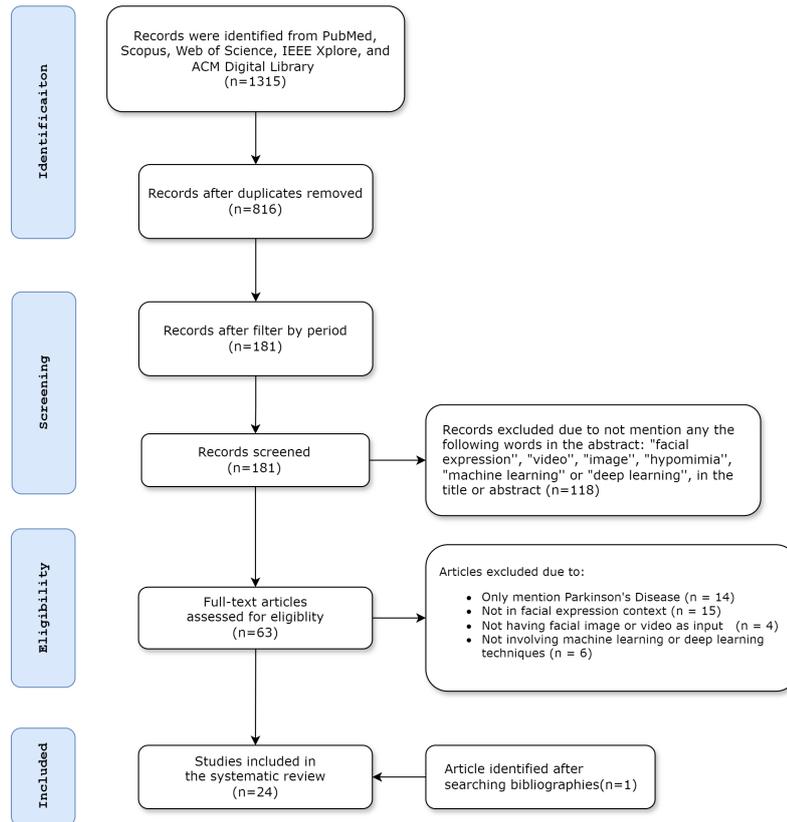


Figure 3 – Flow chart of the literature selection process.

Grammatikopoulou et al. (GRAMMATIKOPOULOU *et al.*, 2019) evaluated the facial expressions of 23 patients and 11 healthy controls individuals using images obtained directly from smartphones. Two geometric feature sets were extracted using the Google Face API and the Microsoft Face API. Afterward, the authors trained two linear regression models (one for each feature set) to estimate two distinct values of the Hypomimia Severity index, i.e., HSi1 and HSi2, which were used to classify healthy controls and PD patients. The authors report sensitivity and specificity values of 0.79 and 0.82 for HSi1, respectively, and 0.89 and 0.73 for HSi2, respectively.

In the following year, Gomez-Gomez et al. (GOMEZ-GOMEZ *et al.*, 2020) used a deep learning approach to model hypomimia in PD. The lack of huge databases of videos and images of PD individuals was the most significant obstacle they encountered when applying such approaches. Overall, their main contributions were a framework for using deep face architectures to predict hypomimia in PD patients; a comparison of PD diagnosis based on single photographs against image sequences while the patients are prompted various face emotions; to investigate different domain adaptation approaches to use current models, initially trained for Face Recognition or detecting FAUs, for automated classification between patients and healthy participants; and a novel method to use triplet-loss learning to enhance hypomimia detection.

In contrast to the previous article, Jin et al. (JIN *et al.*, 2020) gathered images of patients with PD and their matched controls' facial expressions. They extracted facial expression char-

acteristics, such as facial expression amplitude and shaking of tiny facial muscle groups, from the facial landmarks given by Face++<sup>1</sup> using relative coordinates and positional jitter. Standard machine learning and deep learning algorithms were used to help diagnose PD.

Xinyao et al. (HOU *et al.*, 2021) developed a model that performs masked face recognition of PD patients based on the geometric and texture features extracted from specific frames of the collected video, during which patients were poker-faced or smiling. The authors considered three geometric features, i.e., the angles of the mouth landmarks (angles before and after laughter), the overall deviation mouth angle, and the deviation angle of the left and right sides of the mouth. In contrast, the histogram of oriented gradient (DALAL; TRIGGS, 2005; DÉNIZ *et al.*, 2011) and local binary pattern features (OJALA; PIETIKAINEN; MAENPAA, 2002; AHONEN; HADID; PIETIKÄINEN, 2004) were used to describe facial textures. These two features were combined to establish a patient identification standard. The next step was to train the models, such as Random Forest, Support vector Machines, and k-nearest neighbors. Finally, they compared their outcomes against the technique proposed by Jin et al. (JIN *et al.*, 2020), showing an increase of 0.16 in the F1 value.

Sonawane and Sharma (SONAWANE; SHARMA, 2021) meant to review automated approaches and the application of machine learning in identifying emotional facial expressions in PD patients. The authors demonstrate that deep learning has not yet been effectively addressed in differentiating between healthy persons and patients. In addition, they conducted a pilot experiment that created a single CNN from scratch for detecting masked faces, demonstrating that deep learning-based models can be highly beneficial for diagnosis. On the testing images, the deep learning-based model yielded an accuracy of 85%.

In the same year, Abrami et al. (ABRAMI *et al.*, 2021) also trained a Convolutional Neural Network using two facial datasets, i.e., the YouTube Faces Database (WOLF; HASSNER; MAOZ, 2011), containing 3,425 videos of 1,595 people to represent the control database, and a dataset created by searching by the terms “Parkinson’s disease” and “interview” on YouTube to comprise people with PD. Further, the algorithm assigns a score between 0 and 1 for each frame in a new video, describing its likelihood of hypomimia, supposing that PD patients would produce a more significant hypomimia score than control individuals. Afterward, this trained model was evaluated in clinical interviews with 35 Parkinson’s disease patients in their on and off drug motor phases. The algorithm achieved an area under the receiver operating characteristic curve of 0.71 over the test set, comparable to a value of 0.75 for professional neurologists using the United Parkinson Disease Rating Scale-III Facial Expression score. Furthermore, the model’s accuracy in classifying on and off-drug states in clinical samples was 63%, compared to 46% clinical rater scores.

Like Abrami et al. (ABRAMI *et al.*, 2021), Jakubowski et al. (JAKUBOWSKI *et al.*, 2021) conducted a set of experiments on patients with Parkinson’s disease in the so-called ON

---

<sup>1</sup> <<https://www.faceplusplus.com/>>

phase, which reduces the symptoms of the disease with the action of drugs. The authors used face images acquired under visible light and infrared, suggesting that thermal imagery is less dependent on ambient lighting conditions. Moreover, they notice that the temperature distribution on the skin's surface may be a helpful biomarker in Parkinson's disease. Further, a convolutional network was trained using the facial images recorded in visible light, infrared, and a fusion of both spectral ranges, obtaining an F1 score of 0.941.

Ali et al. (ALI *et al.*, 2021) employed machine learning techniques to quantify the variance of facial muscle movements and used it to discriminate between people with and without Parkinson's disease. Participants with PD showed minor variation in cheek raiser (AU6), lip corner puller (AU12), and brow lower (AU4) than the control people, obtaining 95.6% accuracy with the Support Vector Machine classifier. In a similar work, Wu et al. (WU *et al.*, 2014) employed seven PD patients and eight control subjects, which were instructed to produce neutral facial expressions and expressions resembling amusement, sadness, anger, disgust, surprise, and fear. However, only disgust was used in the analysis because the participants rated its intensity as the highest while viewing recorded videos.

Guan (GUAN, 2021) proposed a Parkinson's disease expression disorder diagnosis and evaluation method based on Facial Action Coding System, extracted by OpenFace. In contrast to Ali et al. (ALI *et al.*, 2021), the intensity from each 17 AU was decomposed into a 36-dimensional feature vector, leading to a 612-dimensional representation per video file. Such representations are further preprocessed to remove missing values and replace outliers, as well as smoothing and filtering objectives.

Gomez et al. (GOMEZ *et al.*, 2021) investigated the performance of evoked facial movements for PD identification. They examined the use of static features, collected using a pre-trained Resnet50 for face verification, and dynamic features, produced by using 3D landmarks to determine distances between important facial areas and describe the movements associated with the AUs, for modeling hypomimia. They introduced a unique feature set of 17 parameters based on velocity, acceleration, and jerk, to define the expressiveness of evoked facial motions in video sequences. The study evaluated video recordings of people producing four distinct facial gestures, i.e., Happy, Angry, Surprised, and Wink. The results suggest that employing static features generated by pre-trained deep architecture increased the accuracy of PD detection by up to 77.36%, while combining static and dynamic features improves detection by up to 13.46% (from 75.00% to 88.46%).

Hu et al. (HU; ZHANG; HUANG, 2021) presented a technique that considered the six fundamental facial expressions (EKMAN; FRIESEN, 1971) and identification elements, i.e., anger, happiness, fear, surprise, disgust, and sadness. They proposed CycleGAN-based networks to synthesize the basic facial expression images of the "non-PD scenario" of PD patients. They trained six CycleGAN-based networks corresponding to each of the basic facial expressions, using public facial expression image datasets of non-PD patients. Further, the network generates

synthetic samples, which are combined with the real images to feed a ResNet-based network for classification purposes. A new triplet loss-based metric learning network is implemented to distinguish between PD patients and non-PD patients, increasing the accuracy of the PD diagnosis. The authors also made available a dataset composed of 95 PD patients for experimental analyses.

Like Grammatikopoulou et al. (GRAMMATIKOPOULOU *et al.*, 2019), Su et al. (SU *et al.*, 2021b) presented a facial expression-based automated detection technique for Parkinson's hypomimia, which comprises two main aspects, i.e., geometric features and texture features. Geometric features regard Facial expression factors (FEFs) and Facial expression change factors (FECFs), which are extracted with the help of facial key points. In contrast, to take into account temporal factors, they introduced the extended histogram of oriented gradients (HOG) algorithm (CHEN *et al.*, 2016), which combines temporal and spatial dimensions. Finally, these features are applied to four machine learning and detection algorithms, which performed best when combining geometric and texture characteristics, yielding an F1 score of 0.9997. The best F1 scores obtained using geometric and texture characteristics were 0.8286 and 0.9446, respectively. This research included 39 individuals (21 males and 18 females) for control and 47 people (26 males and 21 females) PD patients.

In the same year, Su et al. (SU *et al.*, 2021a) proposed the Semantic Feature-based Hypomimia Recognition network to recognize hypomimia given facial videos. They studied how to recognize semantic areas salient to hypomimia using deep learning models. Three techniques were developed. First, a face recognition network and sliding window are shown to search for video clips, including facial activity. Next, they developed a Semantic Feature Classifier (SF-C) to extract feature maps salient to hypomimia and compute semantic loss. Finally, they employed an encoder to produce semantic characteristics related to hypomimia. Besides that, a temporal optical flow is thereby incorporated to compute the dynamic changes. The two-stream architecture comprises both optical flow and spatial information.

In 2022, Pegolo et al. (PEGOLO *et al.*, 2022) implemented a face-tracking algorithm based on a Facial Action Coding system. They established simple measures, i.e., Facial Mobility Index (FMI), on the distances between pairs of geometric characteristics and presented a categorization system. The results imply that this score may be used to evaluate PD's degree of impairment and classify emotions. There were statistically significant changes for all emotions when distances were examined. Random Forest and  $k$ -NN produced the most accurate area under the curve measures. Meanwhile, Valenzuela (VALENZUELA *et al.*, 2022) employed spatio-temporal convolutional representations to learn facial movement patterns capable of discriminating between PD and control patients. The authors built a 3D convolutional architecture integrated with inception modules to achieve salient maps of face expression activations. The method obtained an average classification accuracy of 91.87% using 480 video sequences.

In the consecutive year, Gomez et al. (GOMEZ *et al.*, 2023) considered domain adaptation from face analysis to action unit recognition for hypomimia-based PD classification. The authors

employed single frames and sequences of images, transfer learning, and triplet-loss functions to improve the automatic subjects' classification. Meanwhile, Huang et al. (HUANG *et al.*, 2023) studied the masked face phenomenon caused by PD patients' emotional expression disorders and proposed an auto-PD diagnosis method based on synthesized face images. The authors trained a deep feature extractor and a facial expression classifier over a mixture of real facial expression images and synthesized samples of PD patients. Further, they also collected a new facial expression dataset of PD patients in collaboration with a hospital.

Similar work by Oliveira et al. (OLIVEIRA *et al.*, 2023) tackled the challenge posed by small and unbalanced datasets available for hypomimia detection. The work proposes augmenting and balancing such datasets, generating novel instances through a Conditional Generative Adversarial Network (CGAN). Such a procedure improved the classification efficacy, increasing the accuracy, specificity, and sensitivity. Notwithstanding, Xu et al. (XU *et al.*, 2023) explored a method for PD diagnosis by analyzing facial videos using deep learning. The authors compared facial expression videos from PD patients against healthy individuals from open-source resources. The video frames are preprocessed and used to feed a deep-learning architecture, which obtained a classification accuracy of 81.73%.

More recent studies conducted in 2024 also employ a similar approach (ZHOU *et al.*, 2024b), dealing with the limited training instances on PD patients by artificially generating facial expressions images through multi-domain adversarial learning and training a deep neural network prediction over artificial and real PD patients' and normal individuals' facial expression, confirming the advantages provided by the use of augmented image data for training.

Later on, Munsif et al. (MUNSIF *et al.*, 2024) proposed a deep learning-based, efficient, lightweight convolutional block attention module to aid the diagnosis of PD and other neurological disorders patients. The method collects data from real patients for further pre-processing for face detection and attention-enhanced feature extraction and refinement.

Among the current novelties, some recent works propose to attack the PD identification problem using facial expressions and audio information in a multimodal fashion. In this context, Zhou et al. (ZHOU *et al.*, 2024a) proposed the YouTubePD: a public multimodal dataset that includes in-the-wild videos, audio, and facial landmarks for PD analysis. In this context, Lv et al. (LV *et al.*, 2024) proposed an audio-visual fusion model that integrates visual features and audio Mel-spectrogram features using a Transformer-based cross-attention module to learn the complementarity between audio and visual cues. The method surpassed conventional approaches, achieving an accuracy rate of 92.68% in PD diagnosis.

Razzouki et al. (RAZZOUKI *et al.*, 2024) proposed a method to identify hypomimia in early-stage PD individuals using optical-flow-based video vision transformer. The authors encouraged patients and healthy participants to speak while recording typical facial muscle movements upon which the optical flow was computed. Such components and the RGB images feed a video vision transformer responsible for extracting feature representations for Random

Forest-based classification, obtaining classification scores up to 83% in terms of balanced accuracy. In similar work, Huang et al. (HUANG *et al.*, 2024) proposed a lightweight explainable 3D multi-head attention residual convolution network. The model feeds the video features extracted through a 3D attention-based convolution layer into LSTM and residual backbone networks, followed by a feature compression module to condense the learned contextual features. The model achieved state-of-the-art diagnosis performance.

### 2.1.3.3 Discussion

#### 2.1.3.3.1 General Observations

This subsection discusses some aspects of the solutions proposed in the reviewed articles. Tables 2 and 3 provide an overview of the references, facial expressions, and datasets utilized by each study, whereas Tables 4, 5 and 6 summarize the extracted features, types of models, and metric results.

#### 2.1.3.3.2 Traditional Machine Learning

Facial expression recognition systems typically follow a standard workflow involving several key steps, from data acquisition to emotion classification (Fig 4). This workflow includes image or video input, pre-processing, feature extraction, and classification. The choice of methods at each stage significantly impacts the system's performance.

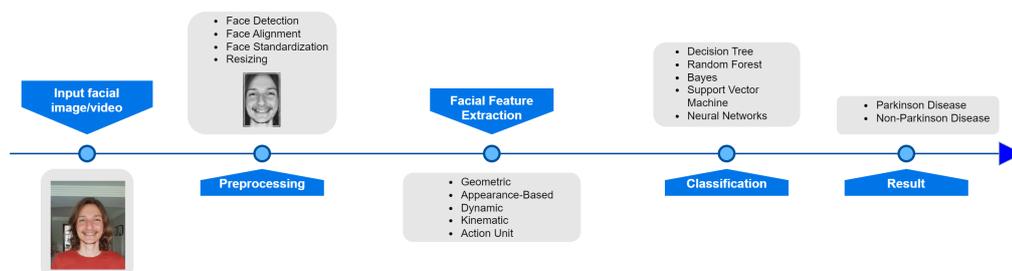


Figure 4 – The essential steps in identifying Parkinson's Disease through facial expressions.

In the pre-processing phase, the system detects and aligns faces to reduce variances caused by different postures, scaling, and lighting conditions. The subsequent feature extraction step is crucial, as it captures variations in facial expressions using various techniques. These features can be categorized into geometric features, appearance-based features, dynamic features, kinematic features, and AUs (KOTSIA; PITAS, 2006; GHIMIRE; LEE, 2013; YANG *et al.*, 2004; AHONEN; HADID; PIETIKAINEN, 2006).

Geometric features emphasize the spatial arrangement and structural relationships of facial components such as the eyes, eyebrows, nose, and mouth. By detecting key facial landmarks, systems measure distances, angles, and shapes formed by these points to capture the

Table 2 – Approaches for facial expression analysis to identify PD (Part 1): *References, Expression, and Data.*

References	Expression	Data
Grammatiko-poulou (GRAMMATIKOPOULOU <i>et al.</i> , 2019)	Expression not mentioned.	36 patients with Parkinson’s disease and healthy controls.
Gomez-Gomez <i>et al.</i> (GOMEZ-GOMEZ <i>et al.</i> , 2020)	Evoked emotional states (smiling, anger, and surprise) and coordinated face gestures (right eye wink, left eye wink).	24 healthy participants and 30 PD patients.
Jin <i>et al.</i> (JIN <i>et al.</i> , 2020)	Smile.	33 PD patients, 31 healthy control subjects, 176 records.
Xinyao <i>et al.</i> (HOU <i>et al.</i> , 2021)	Poker face and smiling.	140 videos of facial expressions from 70 PD patients and 70 matched controls.
Sonawane and Sharma (SONAWANE; SHARMA, 2021)	Expression not mentioned.	107 masked face, 104 normal face subjects.
Abrami <i>et al.</i> (ABRAMI <i>et al.</i> , 2021)	Expression not mentioned.	107 videos of self-identified PD patients from a YouTube search and 3425 control videos from the YouTube Faces Database.
Jakubowski <i>et al.</i> (JAKUBOWSKI <i>et al.</i> , 2021)	Smiling and Neutral face.	24 participants in the PD group and 24 participants in a health control group.
Ali <i>et al.</i> (ALI <i>et al.</i> , 2021)	Smiling, disgusted, surprised face and neutral face.	604 individuals, 61 with PD and 543 without PD in 1812 videos.
Guan (GUAN, 2021)	Expression not mentioned.	67 people with PD and 28 healthy control people.
Gomez <i>et al.</i> (GOMEZ <i>et al.</i> , 2021)	Happy, Angry, Surprise and Wink.	30 PD patients and 24 healthy participants.
Hu <i>et al.</i> (HU; ZHANG; HUANG, 2021)	Anger, Disgust, Fear, Surprise, Sadness, Happiness and Neutral.	95 PD patients.
Su <i>et al.</i> (SU <i>et al.</i> , 2021b)	Smile.	39 HC participants (21 males, 18 females) and 47 PD patients (26 males, 21 females).
Su <i>et al.</i> (SU <i>et al.</i> , 2021a)	Smile.	47 people with PD and 39 healthy control individuals.
Pegolo <i>et al.</i> (PEGOLO <i>et al.</i> , 2022)	Anger, disgust, fear, happiness, sadness, surprise and neutral.	50 PD subjects and 20 healthy control subjects.

Table 3 – Approaches for facial expression analysis to identify PD (Part 2): *References, Expression, and Data* (Continued).

References	Expression	Data
Valenzuela (VALENZUELA <i>et al.</i> , 2022)	Facial movements during the pronunciation of sustained vowels.	16 PD patients and 16 healthy control subjects.
Gomez <i>et al.</i> (GOMEZ <i>et al.</i> , 2023)	Neutral, onset-transition, apex, offset-transition, neutral.	30 PD patients and 24 healthy controls.
Huang <i>et al.</i> (HUANG <i>et al.</i> , 2023)	Anger, disgust, fear, happiness, sadness, and surprise.	95 PD patients and 64 HC.
Oliveira <i>et al.</i> (OLIVEIRA <i>et al.</i> , 2023)	Smiling, disgusted, and surprised.	37 PD patients and 480 HC.
Xu <i>et al.</i> (XU <i>et al.</i> , 2023)	General facial expressions captured in video interviews.	106 interview videos (51 PD patients, 55 healthy controls).
Zhou <i>et al.</i> (ZHOU <i>et al.</i> , 2024b)	Six basic emotions (happiness, surprise, anger, fear, disgust, sadness).	95 PD patients and 47 healthy control individuals.
Munsif <i>et al.</i> (MUNSIF <i>et al.</i> , 2024)	Facial expressions (not specifically detailed).	Data from PD and Alzheimer's patients, supplemented by RAVDESS and KDEF datasets.
Zhou <i>et al.</i> (ZHOU <i>et al.</i> , 2024a)	Expression not explicitly mentioned.	16 PD-positive and 89 healthy controls.
Lv <i>et al.</i> (LV <i>et al.</i> , 2024)	Lip movements, orofacial expressions.	130 PD patients and 90 HC people.
Razzouki <i>et al.</i> (RAZZOUKI <i>et al.</i> , 2024)	Spontaneous facial expressions during a monologue.	109 PD and 45 HC.
Huang <i>et al.</i> (HUANG <i>et al.</i> , 2024)	Happiness, anger, surprise, disgust, sadness, and fear.	300 PD patients and 77 healthy controls.

facial configuration associated with different expressions. For example, the degree of eyebrow-raising or mouth curvature can be quantified to distinguish between expressions like surprise and happiness. Studies such as Grammatikopoulou *et al.* (GRAMMATIKOPOULOU *et al.*, 2019) demonstrated the effectiveness of geometric features in classifying PD patients by using linear regression models to estimate Hypomimia Severity indices based on features extracted using the Google Face API and Microsoft Face API. Similarly, Jin *et al.* (JIN *et al.*, 2020) used facial landmarks from Face++ to measure facial expression amplitude and positional jitter in PD diagnosis. However, the accuracy of geometric features depends on landmark detection and facial alignment, making them sensitive to these preprocessing steps.

Appearance-based features capture texture information and subtle skin appearance changes resulting from facial muscle movements, such as wrinkles, furrows, and shading variations. Instead of relying on specific landmarks, these features analyze the pixel intensity patterns

Table 4 – Approaches for facial expression analysis to identify PD (Feature, Model, and Metrics) — Part 1.

References	Feature	Type of Model(s)	Metric Results
Grammatikopoulou (GRAMMATIKOPOULOU <i>et al.</i> , 2019)	Variance of mouth coordinates, average palpebral fissure asymmetry, average lip separation.	Linear regression.	Sensitivity/specificity: 0.79/0.82 (HSi1); 0.89/0.73 (HSi2).
Gomez-Gomez <i>et al.</i> (GOMEZ-GOMEZ <i>et al.</i> , 2020)	VGGFace2 adapted to an affective domain using EmotioNet and FAU detection.	VGG and ResNet models.	Accuracy improved from 78.4% to 87.3%.
Jin <i>et al.</i> (JIN <i>et al.</i> , 2020)	848 facial expression amplitude features and tremor features (65 features after compression).	LSTM Neural Network (for keypoint movements) and SVM (for tremor features).	LSTM: 86% precision, 75% F1; SVM: 99% precision, 99% F1.
Xinyao <i>et al.</i> (HOU <i>et al.</i> , 2021)	Geometric features (e.g., mouth angles) and texture features using HOG and LBP.	Random forest, SVM, and k-nearest neighbor.	Geometric features: 83% accuracy; Texture features: 86% accuracy.
Sonawane and Sharma (SONAWANE; SHARMA, 2021)	CNN-based feature extraction.	14-layer CNN with 2 final neurons (PD vs. non-PD).	85% accuracy (70%/30% split).
Abrami <i>et al.</i> (ABRAMI <i>et al.</i> , 2021)	VGG16 output of all video frames to generate a probability distribution.	VGG16 Net.	76% accuracy (off-drug), 67% (on-drug); compared to neurologist: 88%/70%.
Jakubowski <i>et al.</i> (JAKUBOWSKI <i>et al.</i> , 2021)	Fusion of visible light and infrared channels (neutral, smile, and mean).	Reduced AlexNet.	93.80% accuracy, F1 score of 94.10%.
Ali <i>et al.</i> (ALI <i>et al.</i> , 2021)	Variance of AU magnitude per frame (using OpenFace).	SVM trained on the variances of facial muscle movements.	95.6% accuracy (LOPO-CV).
Guan (GUAN, 2021)	612-dimensional eigenvector representation of the video.	Logistic Regression and KNN.	90.06% accuracy.
Gomez <i>et al.</i> (GOMEZ <i>et al.</i> , 2021)	2048 static features (ResNet50 pre-trained on VGGFace2) and 17 dynamic features (MediaPipe Face Mesh).	Two linear SVM models with Sum Rule fusion.	88.46% accuracy for PD detection, 44% for impairment estimation (5-fold CV).
Hu <i>et al.</i> (HU; ZHANG; HUANG, 2021)	ResNet18 features with a new triplet loss-based metric.	CycleGANs to generate the ResNet18 for PD classification.	93.8% accuracy in a 5-fold CV setting.

Table 5 – Approaches for facial expression analysis to identify PD (Feature, Model, and Metrics) — Part 2 (Continued).

References	Feature	Type of Model(s)	Metric Results
Su et al. (SU <i>et al.</i> , 2021b)	Geometric features and extended HOG texture features (with PCA reduction).	Pre-trained MTCNN for face detection and SVM for classification.	F1 score of 0.9997 (using fused features in 5-fold CV).
Su et al. (SU <i>et al.</i> , 2021a)	Deep learning-based feature extraction.	ResNet, VGG, and SFHR-NET.	F1 scores: ResNet=93.55%, VGG=95.28%, SFHR-NET=99.49%.
Pegolo et al. (PEGOLO <i>et al.</i> , 2022)	Distances between pairs of geometric features.	k-NN, Tree, Random Forest, Neural Network, Naive Bayes, CN2 rule inducer.	RF: AUC values 94.3–91.6, F1 scores 76.2–71.5.
Valenzuela (VALENZUELA <i>et al.</i> , 2022)	Spatio-temporal facial patterns via 3D convolutional layers with inception modules.	3D CNN based on the I3D (Inflated 3D ConvNet) architecture.	Accuracy: 91.87%, F1: 91.47%, AUC: 95.1%.
Gomez et al. (GOMEZ <i>et al.</i> , 2023)	ResNet50 features pre-trained on VGGFace2 and action unit features from EmotioNet.	ResNet50 with domain adaptation and Triplet loss.	87.3% accuracy for PD detection.
Huang et al. (HUANG <i>et al.</i> , 2023)	Deep features from real and synthesized images.	StarGAN for synthesis combined with EfficientNet-B7 and Swin transformer for classification.	100% accuracy in PD diagnosis; 70.08% for emotion recognition.
Oliveira et al. (OLIVEIRA <i>et al.</i> , 2023)	Variance of facial action units (AUs) using OpenFace for smile, disgust, and surprise.	Logistic Regression with CGAN and Test-Time Augmentation.	83% accuracy, 85.71% sensitivity, 82.47% specificity.
Xu et al. (XU <i>et al.</i> , 2023)	Spatiotemporal features from a sequence of video frames (ResNet-34 for spatial, LSTM for temporal).	Combined ResNet and LSTM model (ResLSTM).	81.73% accuracy, 96.84% precision, 72.24% recall, 82.75% F1.
Zhou et al. (ZHOU <i>et al.</i> , 2024b)	Deep features extracted via a ResNet-18-based network with MDFES-GAN for augmentation.	ResNet-18.	97.55% accuracy in PD diagnosis.
Munsif et al. (MUNSIF <i>et al.</i> , 2024)	Facial features using CNN refined with the Convolutional Block Attention Module (CBAM).	Lightweight CNN with CBAM.	73.2% accuracy, 73.4% precision, 73.5% recall.

Table 6 – Approaches for facial expression analysis to identify PD (Feature, Model, and Metrics) — Part 3 (Continued).

References	Feature	Type of Model(s)	Metric Results
Lv et al. (LV <i>et al.</i> , 2024)	Visual features from ShuffleNet-V2 and audio features as Mel-spectrograms.	Multimodal deep learning framework with a Transformer-based cross-attention mechanism.	Accuracy: 92.68%, F1: 94.23%, Sensitivity: 96.08%.
Razzouki et al. (RAZZOUKI <i>et al.</i> , 2024)	Optical Flow (OF) and RGB frames (via VideoMAE and ResNet34).	Random Forest classifier.	83% Balanced Accuracy, 84% AUC.
Huang et al. (HUANG <i>et al.</i> , 2024)	3D convolutional features combined with LSTM and attention mechanisms.	3D Multi-head Attention Residual Network (MARNet).	Accuracy: 0.88, Precision: 0.86, F1: 0.91, Recall: 0.98, Specificity: 0.65.

across facial regions (HOU *et al.*, 2021; SU *et al.*, 2021b). Techniques like Local Binary Patterns (LBP) encode local texture by comparing neighboring pixel intensities, while Gabor filters capture spatial frequency information at multiple orientations and scales. Histograms of Oriented Gradients (HOG) describe the distribution of edge directions within an image. Studies have combined appearance-based features such as LBP and HOG with geometric features to enhance patient identification accuracy, particularly in the context of PD diagnosis (HOU *et al.*, 2021; SU *et al.*, 2021b). For instance, Su et al. (SU *et al.*, 2021b) proposed an extended HOG algorithm incorporating temporal and spatial dimensions to improve the detection of Parkinson’s hypomimia. Appearance-based features can, however, be sensitive to lighting conditions and occlusions, requiring careful normalization and preprocessing to ensure robustness.

Dynamic features capture temporal changes in facial expressions by analyzing motion information over sequences of images or video frames, focusing on how expressions evolve rather than static facial configurations. Methods like optical flow compute motion vectors representing pixel or feature movement between consecutive frames, effectively capturing dynamic movements such as smiling or frowning. In the study by Gomez et al. (GOMEZ *et al.*, 2021), dynamic features were analyzed using 3D landmarks to capture temporal changes in facial expressions, which improved PD detection accuracy when combined with static features. Similarly, Su et al. (SU *et al.*, 2021a) proposed a Semantic Feature-based Hypomimia Recognition network, incorporating temporal optical flow to compute dynamic changes, resulting in more accurate hypomimia recognition. These examples illustrate that dynamic features are essential for recognizing expressions that unfold over time, like the gradual onset of anger or a quick surprise reaction, but require video data and increased computational resources to process temporal sequences effectively.

Kinematic features represent a more specialized subset of dynamic features, describing the motion properties of facial points or regions, such as displacement, velocity, acceleration, and

even higher-order derivatives. By tracking facial landmarks across frames, kinematic features quantify the speed and manner of facial component movements during expressions. For instance, Gomez et al. (GOMEZ *et al.*, 2021) developed a kinematic feature set based on velocity, acceleration, and jerk of facial movements to enhance PD detection models, providing a detailed analysis of expressiveness in video sequences. This level of fine-grained motion analysis is particularly valuable for detecting micro-expressions and subtle emotional cues that static features may not capture. However, applying kinematic features can require high-quality, high-frame-rate video data.

AUs are the building blocks of the FACS. Each AU represents a distinct facial movement and corresponds to activating a specific set of facial muscles. For example, AU6 (cheek raiser) involves the contraction of the orbicularis oculi, while AU12 (lip corner puller) is associated with activating the zygomaticus major muscle. By detecting and analyzing the presence and intensity of AUs, systems can interpret complex expressions as combinations of these fundamental muscle movements. In their research, Gomez-Gomez et al. (GOMEZ-GOMEZ *et al.*, 2020) utilized domain adaptation approaches to leverage models trained for AU detection to classify PD patients more accurately, improving hypomimia detection. Similarly, Oliveira et al. (OLIVEIRA *et al.*, 2023) focused on creating synthetic AUs to augment data and enhance the detection of hypomimia. Another study by Guan (GUAN, 2021) employed extensive pre-processing and smoothing techniques to extract intensities of 17 AUs using OpenFace, creating a high-dimensional feature vector that enhanced classification performance. These examples illustrate that AUs provide an anatomically grounded and interpretable framework for facial expression recognition, making them valuable in applications like psychological research, clinical diagnostics, and any domain requiring a detailed understanding of facial muscle activity.

#### 2.1.3.3.3 Deep Learning Approaches

Recent advancements in deep learning have provided considerable opportunities for enhancing facial expression analysis. By automatically learning hierarchical feature representations from raw data, deep learning models capture complex patterns and nuances in facial expressions that traditional methods might miss, improving classification accuracy for PD identification. Studies have utilized various advanced deep learning architectures, such as Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM), Generative Adversarial Networks (GANs), and others, in their facial expression analysis to identify PD more effectively.

CNNs have been extensively used in facial expression analysis for PD detection due to their powerful capability to learn spatial hierarchies from images, enabling them to extract complex patterns associated with PD. Researchers have applied various CNN architectures, ranging from custom-designed models to well-established architectures like ResNet, VGG, and AlexNet. For instance, Sonawane and Sharma (SONAWANE; SHARMA, 2021) employed CNN for PD identification using images. Similarly, models like VGG16 (ABRAMI *et al.*, 2021) and

adaptations of the VGGFace2 (GOMEZ-GOMEZ *et al.*, 2020) have been employed to capture detailed facial features from video frames, proving effective in distinguishing between PD and non-PD expressions. In other studies, pre-trained architectures such as ResNet50 (GOMEZ *et al.*, 2021; ZHOU *et al.*, 2024a) and ResNet18 (HU; ZHANG; HUANG, 2021) were fine-tuned to enhance PD detection accuracy, incorporating both static and dynamic facial features. These efforts underscore the flexibility of CNNs in modeling complex facial expressions and their pivotal role in PD facial analysis.

LSTM have shown promise in capturing the temporal dynamics of facial expressions, which is crucial for analyzing the progression and subtleties of PD symptoms over time. While CNNs excel at spatial feature extraction, LSTMs are well-suited for handling data sequences, making them ideal for modeling temporal variations in facial expressions. Jin *et al.* (JIN *et al.*, 2020) utilized LSTMs to analyze the temporal patterns of facial key points extracted from video data, allowing the model to learn PD-related facial movements such as reduced expressiveness or tremors. This integration of temporal modeling offers a more comprehensive understanding of PD symptoms, demonstrating how LSTMs capture the dynamic aspects of facial expression changes.

My literature search has not found any studies employing Large Language Models (LLMs) to analyze facial expressions in PD patients. However, the emergence of multimodal models that integrate visual and textual data suggests that LLMs could be extended to handle visual inputs. Furthermore, existing literature indicates that LLMs can act as agents in managing user interactions. For instance, Oliveira *et al.* (OLIVEIRA *et al.*, 2024b) developed a Telegram chatbot that uses LLMs to manage user interactions and trigger voice-based assessments for PD diagnosis. Their study demonstrated that this approach could be extended to video-based assessments, such as analyzing facial expressions via chat, making it a versatile tool for health screening.

#### 2.1.3.3.4 Facial Video Datasets for PD

The small amount of available data is a common issue that usually leads to other problems, such as lack of generalisation. Typical evidence of lack of generalization is overfitting, when the model performs very well in the training dataset while poor in the test set. Further, outliers may also represent a more challenging constraint in the target variable or the feature representations, affecting the training and model analysis. In this scenario, statistical tests, parametric models, bootstrapping, and other valuable statistics tools are essential to investigate the problem. For example, it is critical to use a margin of error rather than point estimates when working with a small dataset to avoid reaching incorrect conclusions. Models with limited data will have high confidence intervals; however, it is preferable to be aware of the range when making meaningful predictions than not to know.

Table 7 – Comparison of some results using Jin et al. (JIN *et al.*, 2020) approaches in a different work.

	Algorithm	Precision	Recall	F1-Score
Jin et al. (JIN <i>et al.</i> , 2020)	SVM	0.99	0.99	0.99
	RF	0.98	0.98	0.98
Hou et al. (HOU <i>et al.</i> , 2021)	SVM	0.78	0.7	0.74
	RF	0.6	0.9	0.72

As a result of working with a small dataset and not showing a margin of error, all results in Tables 4, 5 and 6 need to be interpreted cautiously. The approach of (JIN *et al.*, 2020) showed an outstanding result in precision, recall, and F1- value. However, Hou et al.(HOU *et al.*, 2021) conducted the same approach of (JIN *et al.*, 2020) in their dataset, showing different and worse results. As Table 7 shows, there is a significant difference between the results. The F1-Score when employing the SVM algorithm dropped sharply from 0.99 to 0.74, and F1-Score when using RF also decreased from 0.98 to 0.72. Besides that, precision and recall are also reduced using SVM or RF.

The progress in using machine learning for detecting PD faces has been significantly hindered by the absence of a unified, publicly available benchmark, which prevents comprehensive evaluation of existing PD analysis methods and the development of robust models. A notable attempt to address this issue is the introduction of the YouTubePD (ZHOU *et al.*, 2024a) dataset, comprised of videos featuring PD subjects on YouTube. The YouTubePD (ZHOU *et al.*, 2024a) dataset provides multimodal information, including in-the-wild videos, audio, and facial landmarks, with annotations from a clinical expert. It includes 283 videos featuring 16 public figures with confirmed PD diagnoses and 89 healthy control subjects. The two main limitations are the lack of diversity with only 16 PD subjects and, although the in-the-wild videos provide a more realistic and naturalistic setting, the uncontrolled conditions make it more challenging to interpret facial expressions and explain the findings. Zhou et al. (ZHOU *et al.*, 2024a) also address concerns related to the ethical implications of using publicly available videos for research purposes.

Despite this, many existing datasets do not provide facial expressions due to ethical constraints, as the facial image represents the patient’s identity (SONAWANE; SHARMA, 2021; ZHOU *et al.*, 2024a). Therefore, finding a public dataset for such a purpose becomes challenging. To enhance the effectiveness of automatic evaluation systems and promote their use in clinical practice, large public datasets containing face videos, images, and clinical information, e.g., diagnosis and clinical ratings, are required. Not only will the availability of the data stimulate the development of more precise procedures, but it will also bring together the efforts of several researchers working to tackle clinical problems. As a result, having a standardized dataset, all researchers would be pursuing the same goal.

To address the issue of limited datasets, data augmentation can be employed. Oliveira et al. (OLIVEIRA *et al.*, 2023) utilized Conditional Generative Adversarial Networks (CGAN) to generate synthetic data for augmenting a dataset in a study on the detection of hypomimia in

PD using video-based analysis, resulting in improved model accuracy. Similarly, Hu et al. (HU; ZHANG; HUANG, 2021) proposed a novel face-based approach for early diagnosis of PD, using CycleGAN to synthesize facial expression data to represent ‘non-PD scenarios’ for PD patients, which significantly enhanced the performance of their diagnostic model. Zhou et al. (ZHOU *et al.*, 2024b) also demonstrated the effectiveness of synthetic data generation using Multi-Domain Facial Expression Synthesis GAN (MDFES-GAN), generating augmented facial expression data to improve early PD diagnosis models.

However, selecting an appropriate technique for synthetic sample generation is crucial and can be complex, as these algorithms are sensitive to the nature of the data. In the case of image and video data, it is often beneficial to train generative models using datasets from diverse sources, exposing the model to a broader range of data distributions (FRÖHLICH *et al.*, 2022). This practice enhances the quality of synthetic data and improves the model’s generalization ability when applied to real-world data.

#### 2.1.3.3.5 Technology Readiness for Clinical Translation

Approaches that employ a conventional machine learning methodology require a time-consuming manual feature extraction procedure and domain knowledge. For example, the benefit of hand-crafted geometric features in facial analysis is their straightforward interpretation. Nevertheless, in the case of human faces, selecting suitable geometric characteristics might be challenging because they also include identifying information about the person. In this context, deep learning networks can extract both undefined and generalized features, i.e., extracting features automatically and then classifying them, requiring minimal domain expertise. Recent advances in automated facial expression analysis have concentrated on developing and training deep neural networks to achieve state-of-the-art results in face recognition (MASI *et al.*, 2018; WANG; DENG, 2021). However, the main disadvantage of this approach is that they need a considerable amount of data to train. In addition, the main issue surrounding deep learning medical models in practice is understanding how these models will interact with doctors’ decision-making processes. One of the constraints of deep learning architecture is that it is difficult to comprehend how models make particular decisions. In this context, various methods provide insight into these processes, namely explainable Artificial Intelligence (XAI) (ADADI; BERRADA, 2018; TJOA; GUAN, 2020; ARRIETA *et al.*, 2020) are warmly welcome.

A specific subfield of deep learning known as multimodal learning (JAEGLE *et al.*, 2021a; JAEGLE *et al.*, 2021b; AKBARI *et al.*, 2021; NAGRANI *et al.*, 2021) provided good insights regarding PD detection. Such algorithms combine features from different domains, integrating inputs from distinct domains for the same purpose, e.g., combining face images and voice signals, which is commonly employed for speech enhancement (PASSOS *et al.*, 2022; PASSOS; PAPA; ADEEL, 2022), but quite unexplored for PD detection tasks, except by a few works (FRÖHLICH *et al.*, 2022).

One of Artificial Intelligence's main challenges for clinical problems concerns the achievement of reliable generalization across different populations. For example, suppose a facial expression model trained to identify PD in people from the Philippines, and I wanted to apply it to a hospital in the United States. Ethnic differences can affect facial muscle structure and typical expressions, which may impact the model's accuracy when applied to a different ethnic group. Therefore, adapting and testing the model for the new domain becomes necessary.

There is also a challenge of generalization regarding diagnostic technology. Suppose I have measured the model's performance on data collected from some countries over a few years, but smartphones are not standardized—some smartphones have higher video recording resolutions and different hardware specifications. Moreover, image preprocessing methods differ between manufacturers, which could make employing the smartphone method challenging. Before applying the PD facial expression model in a new hospital, it is necessary to ensure that the model can generalize to the hardware and preprocessing variations of the cameras employed there.

All the experiments discussed so far employ retrospective data, i.e., the algorithms were trained and tested using historically labeled data. However, to fully appreciate the utility of models in a practical aspect, they must be applied to prospective data. For example, consider training a model such that the dataset contains only frontal face images. Further, suppose that in practice, some pictures are often taken from the patient's side view of the face. In this context, such images should be removed, or the model must be tuned to identify PD patients under such circumstances.

Another challenge in deploying intelligent models concerns the requirement of appropriate metrics that represent the clinical application. Many previous studies consider the area under the receiver operating characteristics for the task. However, a more fair approach to compare the model's impact on patients at the hospital consists of decision curve analysis, which can assist in quantifying the advantage of employing a model to guide patient care and the traditional randomized clinical trial. Therefore, the experiments provided in this paper use these metrics to compare the outcomes between patients who received the intelligent model estimation and those who did not.

Analyzing the model's performance across subsets of the population, focusing on sensitive attributes such as patients' age, sex, and socioeconomic status, enables us to identify critical model blind spots or unintended biases in clinical settings. For example, a Parkinson's Disease facial expression classifier can achieve a performance comparable to a neurologist's when used on older people. However, its accuracy is reduced when applied to younger individuals, possibly due to age-related differences in facial muscle movements or expression patterns. Therefore, validation using a large dataset is necessary to confirm its effectiveness across different age groups.

Parkinson's disease includes further aspects beyond the assessment of the features of the

face. As a result, the final diagnosis is given by the clinical outcome (POSTUMA *et al.*, 2015) and not only by the expression evaluation. Otherwise, the test would become unfeasible, as a false negative can occur. For example, impaired movement of the legs and hands may appear before the loss of facial expression. The patient could have clear signs of Parkinson's in other body parts and just a slight loss of facial expression. Therefore, I can not make the final diagnosis by evaluating only the facial expression. However, since all symptoms are considered together to meet the Parkinson's test requirements and monitor the disease's progress, the automation of the facial disorder analysis and quantification of hypomimia becomes essential to accelerate the whole process and decrease the time to take action for treatment planning. Moreover, the lack of facial expression may also indicate other underlying neurological diseases, like Bell's palsy (BAUGH *et al.*, 2013), Alzheimer's disease (FÁDEL *et al.*, 2019; LOVELEEN *et al.*, 2023), and Post-Stroke (KAZANDJIAN; BOROD; BRICKMAN, 2007).

Finally, it is worth mentioning the difficulty of considering only facial expression as a biomarker for prognosis. The appearance of non-motor symptoms, such as loss of smell, constipation, sleep disturbances, and depression, are often considered the earliest signals of the disease. Lewy bodies (clumps of abnormally accumulated alpha-synuclein protein found within neurons in the brains of patients with Parkinson's disease) develop first in the olfactory bulb and lower brainstem, causing specific non-motor symptoms associated with these regions (MUELLER *et al.*, 2017). The Lewy pathology then progresses upward through the brainstem to the midbrain, where it affects the substantia nigra, leading to motor symptoms such as resting tremors, bradykinesia (slowness of movement), and rigidity. Therefore, this progression explains why non-motor symptoms frequently precede the motor manifestations of Parkinson's disease.

#### 2.1.3.4 Conclusion

Analyzing facial expressions through machine learning and deep learning has proven to be highly beneficial. In this scenario, it is possible to develop an affordable computer-aid diagnosis to help professionals to identify PD using simple devices like a camera. Moreover, the tool has the potential to reach different people across the world. This paper reviewed facial expressions to identify PD with deep learning and machine learning techniques, examining the constraints of clinical assessment improvement. Regarding the limitations, it could be seen that the small amount of labeled data is a recurrent problem that impacts poor generalization issues. Additionally, it presented the main concerns of diagnostic technologies and time-consuming manual feature extraction approaches.

There are many opportunities for future work in the field that will help translate the technology for clinical use. The scope to detect the severity of PD and differentiate it from other neurological diseases is essential. Exploring approaches to address small data limitations and increase generalization with the limited available data is also important. Finally, there is also the need to investigate the practical aspects of deploying the model in clinical practice and workflow.

### 2.1.4 CRediT authorship contribution statement

I am grateful for the support and collaboration that made this work possible. As the lead investigator, I was responsible for the research, design, methodology development, and writing this study. I thank Quoc C. Ngo for project administration, validation, and manuscript review; Leandro A. Passos for software validation and manuscript review; Danilo Jodas for review analysis; João P. Papa for securing funding and project oversight; and Dinesh Kumar for his conceptual guidance, supervision and manuscript review. Their contributions were essential to the success of this work.

## 2.2 Voice analysis in Healthcare

Voice analysis is increasingly important in healthcare, especially for diagnosing and monitoring neurological disorders like Parkinson's disease. Two key methods in this field are diadochokinetic (DDK) tasks and vocal tract length (VTL) analysis. These techniques provide ways to assess speech and vocal characteristics, helping doctors detect and track diseases.

One of the main advantages of voice analysis is its ability to detect subtle changes in speech that may not be noticeable in regular clinical assessments. These changes can include variations in pitch, speed, and clarity of speech, which are often early signs of neurological decline. By using advanced techniques such as machine learning, these vocal changes can be tracked and analyzed over time, helping doctors make more informed decisions about a patient's condition and treatment options.

Moreover, voice analysis supports remote healthcare, which is particularly valuable in areas with limited access to specialized medical services. Patients can record their voice at home, and the recordings can be analyzed by healthcare professionals to monitor disease progression or to screen for early signs of disorders. This approach not only improves patient outcomes through early intervention but also reduces the need for frequent in-person visits, making healthcare more efficient and accessible.

### 2.2.1 Diadochokinetic tasks

Diadochokinetic (DDK) tasks are simple tests where patients repeat syllables like “pa,” “ta,” and “ka.” These tasks help doctors assess how well a person can control their speech muscles. DDK tasks are useful for spotting early signs of speech problems, especially in conditions like Parkinson's disease.

Studies have indicated that DDK tasks are particularly useful in assessing various aspects of speech production, including phonation, articulation, and prosody (KODALI; KADIRI; ALKU, 2023; GARCIA *et al.*, 2017; LÓPEZ; OROZCO-ARROYAVE; GOSZTOLYA, 2019; CERNAK *et al.*, 2017). For example, the PC-GITA dataset (OROZCO-ARROYAVE *et al.*, 2014), which

includes speech recordings from individuals with PD, has been extensively used to analyze these features during DDK tasks. These tasks are valuable because they highlight variations in fundamental frequency, vocal fold vibration regularity, and prosodic elements, all of which are crucial in understanding the impact of PD on speech (GALAZ *et al.*, 2016; RUSZ *et al.*, 2011b; VÁSQUEZ-CORREA *et al.*, 2018b; MORO-VELAZQUEZ *et al.*, 2021).

Research has shown that individuals with PD often exhibit slower and more irregular DDK rates compared to healthy individuals. This is due to the neuromotor impairments that affect their ability to produce rapid, coordinated speech movements. The DDK task's simplicity and its focus on both consonants and vowels make it a versatile tool in clinical settings, as it is not dependent on the patient's language skills, unlike free speech tasks that may be influenced by reading abilities or language proficiency (PAH *et al.*, 2021).

Other studies also emphasize the significance of DDK tasks in evaluating speech production aspects such as phonation, articulation, and prosody. One study found that the diadochokinetic index of syllable production variability effectively detects articulatory inaccuracies at different stages of Parkinson's disease. This index revealed persistent articulatory deficits starting from stage 2, underscoring its potential as a clinical tool for early detection and monitoring of speech impairments in Parkinson's disease patients (SANHUEZA-GARRIDO; ROJAS-ZEPEDA; GARCÍA-FLORES, 2023).

Finally, a pilot study introduced a pitch diadochokinetic task to assess laryngeal performance in Parkinson's patients by measuring pitch range and slope during rapid transitions between low and high pitches (ANAND, 2023). Findings showed reduced pitch measures in Parkinson's patients, influenced by age and disease duration. Additionally, hypokinetic dysarthria, affecting up to 90% of PD patients, is marked by monotonous speech, imprecise consonant production, and reduced loudness. DDK tasks, combined with assessments like maximum phonation time and spontaneous speech analysis, remain crucial for evaluating phonation, articulation, respiration, resonance, and prosody in PD patients (ATALAR; OGUZ; GENC, 2023; ANAND, 2023).

## 2.2.2 Vocal Tract length

Vocal tract length (VTL) is a significant aspect of vocal analysis, studied in contexts like body size assessment (PISANSKI *et al.*, 2014), speech evolution (PISANSKI *et al.*, 2016a), and neurodegenerative diseases such as Parkinson's disease.

In exploring vocal control and its evolutionary implications, Pisanski *et al.* (2016a) argued that the modulation of nonverbal vocal features like formant and fundamental frequencies may offer insights into the evolution of speech. These vocal modulations could represent remnants of an intermediate system between human speech and the more limited vocalizations of other primates, suggesting that early vocal control abilities were foundational in the development of

articulated human speech.

When applied to neurodegenerative conditions, changes they have shown potential as biomarkers for diseases like PD. Apparent vocal tract length (AVTL), calculated from formant frequencies, refers to the estimated length of the vocal tract during speech production. [Pah, Motin and Kumar \(2022b\)](#) confirmed that AVTL measurements provided high accuracy in distinguishing PD patients from healthy individuals, supporting the use of VTL analysis in telehealth for remote diagnosis and monitoring. Further research by [Pah et al. \(2023a\)](#) examined the relationship between PD and AVTL, finding that PD might increase the AVTL, especially in men. This alteration in AVTL could be linked to the motor symptoms affecting speech in PD patients.

[Mittal and Sharma \(2021\)](#) emphasized the significance of acoustic parameters, including VTL, in detecting Parkinson's disease. Their study achieved a 97.2% accuracy in differentiating PD voices from healthy ones by analyzing jitter-to-pitch ratios and formant frequencies. These findings reinforce the importance of VTL and related acoustic features in the early detection of PD, demonstrating its potential for clinical application.

Finally, it has been shown to have clinical relevance. The cochlear implant study underscores VTL's perceptual importance in voice processing ([NAGELS et al., 2024](#)), supporting its utility in neurodegenerative assessments. However, mixed results in PD studies (e.g., subjective vs. objective measures) highlight the need for standardized protocols in VTL analysis ([ARIAS; CORTÉS; OLMO, 2024](#)).

## 2.3 Image analysis in Healthcare

The advent of deep learning and synthetic data generation has revolutionized image analysis in healthcare, addressing critical limitations such as small and imbalanced datasets. Synthetic data, particularly through techniques like Generative Adversarial Networks (GANs) and diffusion models, plays a pivotal role in enhancing the generalization and robustness of machine learning models in medical applications. This section highlights the importance of synthetic data in healthcare image analysis, focusing on two key areas: estimating thermal images from RGB data for leg ulcer assessment and using StyleGAN2 for generating synthetic retinal images to detect age-related macular degeneration (AMD).

Synthetic data is essential in healthcare for several reasons: it helps overcome the limitations of small datasets by generating additional data that can be used to train deep learning models, it preserves privacy by allowing data sharing and analysis without compromising patient confidentiality, it addresses the issue of imbalanced datasets common in medical imaging, and it improves the generalizability of models, making them more robust to real-world variations.

Generative AI models, including GANs, diffusion models, and autoencoders, are founda-

tional in creating synthetic data. GANs, like StyleGAN2, involve a generator and a discriminator in a competitive process to produce realistic images. This method is particularly effective for generating high-quality, detailed images, making it ideal for applications like retinal image synthesis for AMD detection. Diffusion models, such as Stable Diffusion, start with random noise and iteratively refine images through denoising.

### 2.3.1 Venous Leg Ulcer

Thermal imaging has gained attention as a valuable tool for assessing various medical conditions due to its ability to capture physiological changes such as microcirculation and inflammation. [Cwajda-Białasik et al. \(2020\)](#) highlighted the importance of monitoring temperature differences between the wound and periwound areas, showing that a decrease in these differences is a positive indicator of healing. They also found that increased periwound temperature often correlates with bacterial infections and venous inflammation. Similarly, [Dahlmanns et al. \(2021\)](#) identified distinct temperature patterns associated with different stages of chronic venous disease (CVD), demonstrating the potential of thermal imaging not only for classifying early stages of CVD.

In the context of venous leg ulcers (VLUs), thermal imaging has been explored for its potential to predict healing outcomes. [Monshipouri et al. \(2021\)](#) showed that textural analysis of thermal images could effectively differentiate between healing and non-healing VLUs as early as the second week of treatment, using principal component analysis to predict the likelihood of healing by the twelfth week. [Ngo, Ogrin and Kumar \(2022\)](#) extended this approach by developing a framework that utilizes a Bayesian neural network to analyze thermal images at the initial assessment (week 0), achieving a sensitivity of 78.57% and specificity of 60.00%. This early prediction capability offers a non-contact method for identifying VLUs at risk of delayed healing, potentially allowing for timely intervention.

However, the high cost and limited availability of thermal cameras in many clinical settings hinder widespread use, prompting the need for alternative methods. Recent advances in deep learning, particularly in stable diffusion techniques, have opened new possibilities for estimating thermal images from standard RGB photographs. This approach leverages deep learning models to create thermal images that can aid in the assessment of VLUs without the need for expensive thermal cameras.

Finally, a study demonstrated the high accuracy of thermal imaging in classifying burn wound depth, highlighting its potential to enhance patient care in burn units ([HAAN et al., 2024](#)). Another feasibility study combined thermal imaging with computer vision techniques to assess surgical incision healing, developing a YOLOV8-based machine learning model that accurately identified incision locations, segmented lesions, and classified healing stages ([LI et al., 2024](#)).

### 2.3.2 Age-related macular degeneration

Age-related macular degeneration (AMD) is a leading cause of vision impairment, affecting millions worldwide, especially in populations over 50 years of age. The progressive nature of AMD, which leads to central vision loss, poses a significant challenge in clinical diagnosis and treatment. AMD remains a leading cause of vision impairment, affecting millions worldwide, particularly in populations over 50 years of age. [Fleckenstein, Schmitz-Valckenberg and Chakravarthy \(2024\)](#) shows that AMD is expected to affect approximately 288 million people globally by 2040.

Different approaches have been utilized to address the challenges of generalizability, particularly within the domain of retinal imaging. [Brigato and Iocchi \(2021\)](#) emphasized the effectiveness of standard data augmentation techniques, such as cropping, rotation, and color jittering, especially when paired with low-complexity convolutional neural networks for small datasets. However, they also highlighted the necessity for more advanced data generation and augmentation pipelines, as simple spatial transformations may not provide sufficient data variability ([PEREZ; WANG, 2017](#)). Additionally, while pre-trained models offer some advantages, they often face limitations when applied to new learning domains, highlighting the need for more specialized solutions.

Generative Adversarial Networks (GANs) have emerged as a promising solution for synthetic image generation, providing advanced methods for data augmentation. GANs have demonstrated exceptional results across various medical imaging applications, including liver lesion classification ([FRID-ADAR et al., 2018](#)), and chest pathology recognition ([SALEHINEJAD et al., 2018](#)). In the field of Optical Coherence Tomography (OCT) imaging, super-resolution GANs like ESRGAN have been utilized to enhance image quality, significantly improving the detection of AMD ([THAKOOR et al., 2022](#)).

In retinal imaging, GANs have been used to generate synthetic data to improve disease classification. [Li et al. \(2021\)](#) and [Bellemo et al. \(2019\)](#) discussed the benefits of incorporating synthetic retinal images in training models, particularly in enhancing performance and reducing overfitting. [Burlina et al. \(2019\)](#) and [Ahn, Song and Shin \(2023\)](#) explored the potential of GANs like Progressive GAN and FundusGAN for AMD classification. However, issues related to demographic diversity and image quality control persist, underscoring the need for advanced image assessment models to enhance the effectiveness of GAN-generated data.

---

## VIDEO ANALYSIS FOR NEUROLOGICAL CONDITIONS

---

---

Facial expression analysis is a valuable tool for diagnosing and monitoring neurological disorders, including Parkinson’s disease (PD), amyotrophic lateral sclerosis (ALS), and post-stroke (PS) impairments. This chapter explores advanced machine learning and deep learning methodologies for analyzing facial expressions using video data. It begins with the hypomimia identification in Parkinson’s disease, highlighting the use of Action Units (AUs) analysis through the Facial Action Coding System (FACS) to capture subtle facial movements. Addressing challenges such as limited datasets, this chapter includes the development of data augmentation techniques to improve detection accuracy. Finally, the approach is extended to assess facial impairments in other neurological conditions, utilizing the Toronto Neuroface Dataset to enhance diagnostic reliability for both clinical and remote applications.

Section 3.1 has been published as **“Tabular Data Augmentation for Video-Based Detection of Hypomimia in Parkinson’s Disease”** in *Computer Methods and Programs in Biomedicine*. Section 3.2 presents one work that was split in two published articles: the first, **“Facial Expressions to Identify Post-Stroke: A Pilot Study”**, published in *Computer Methods and Programs in Biomedicine*, and the second, **“Video Assessment to Detect Amyotrophic Lateral Sclerosis”**, published in *Digital Biomarker*.

Section 3.3 presents work in which I contributed as a co-author. My role included assisting in validating the experimental procedures, and reviewing the manuscript. This work was also disseminated through two publications: **“Facial Point Graphs for Stroke Identification”**, presented at the *Iberoamerican Congress on Pattern Recognition*, and **“Facial Point Graphs for Amyotrophic Lateral Sclerosis Identification”**, presented at the *International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*.

## 3.1 Data Augmentation for Facial Expression

With the rapid progression of computerized video analysis and artificial intelligence, techniques have been developed for the computerized detection of hypomimia from videos (JOSHI *et al.*, 2016; BANDINI *et al.*, 2017; RAJNOHA *et al.*, 2018; PHOKAEVVARANGKUL *et al.*, 2020; SONAWANE; SHARMA, 2021; GOMEZ *et al.*, 2021; JAKUBOWSKI *et al.*, 2021; NOVOTNY *et al.*, 2022; CALVO-ARIZA; GÓMEZ-GÓMEZ; OROZCO-ARROYAVE, 2022; VALENZUELA *et al.*, 2022). Jin *et al.* (JIN *et al.*, 2020) work, used AI to extract facial expression features from videos of people with PD and report an F1 value of 99%. Hou *et al.* (HOU *et al.*, 2021) developed an AI model to recognize facial features of PD patients, achieving up to 88% accuracy using geometric and texture features. Gomez *et al.* (GOMEZ *et al.*, 2023) proposed different domain adaptation techniques in automatic face analysis and face action unit detection. However, one of the challenges in detecting the facial expressions associated with people with PD is the shortage of data. Machine learning with small datasets causes overfitting of the training set and yields poor generalization. Another issue is the unbalanced datasets, which generally lead to bias in favor of the majority class (KOTSIANTIS *et al.*, 2006). To overcome the issue of small and unbalanced datasets, generating synthetic data has been proposed with the aim to balance or augment the dataset (WANG; PEREZ *et al.*, 2017; MIKOŁAJCZYK; GROCHOWSKI, 2018; SHORTEN; KHOSHGOFTAAR, 2019), improving the model performance.

In the context of balancing the dataset, to separate the video features of people with PD from Control Matched (CM), Ali *et al.* (ALI *et al.*, 2021) balanced their dataset by upsampling the video features with Synthetic Minority Over-sampling Technique (SMOTE) (CHAWLA *et al.*, 2002) and reported excellent results of 95% accuracy. They increased the number of PD samples from 37 to 480 to match the CM. However, on closer inspection of their results, it is observed that they balanced the dataset with SMOTE before splitting it into training and test partitions. As a result, my work validated the accuracy when the synthetic data is only applied in the training set, avoiding data leakage to the test set. I proposed to split the dataset into training and test sets before generating synthetic data. Moreover, the experiments were conducted in a test set in which the prevalence of people with PD is about 7%, i.e., the test set was unbalanced to evaluate the potential bias of the model.

In contrast to balance, synthetic data can also be generated to augment the dataset. Classical modifications of augmenting data may include random translations, rotations, flips, and Gaussian noise addition, which do not impact the factors contributing to the data label. This strategy is helpful not only for small datasets but also for vast datasets such as ImageNet (DENG *et al.*, 2009a). An alternative to typical data augmentation is to generate synthetic data using Generative Adversarial Networks (GANs) (ANTONIOU; STORKEY; EDWARDS, 2017; WANG *et al.*, 2018; FRID-ADAR *et al.*, 2018; LUSTERMANS *et al.*, 2022). In this context, Antoniou *et al.* (ANTONIOU; STORKEY; EDWARDS, 2017) demonstrated an increase in accuracy of over 13% in the low-data regime experiments in Omniglot (from 69% to 82%), EMNIST (73.9%

to 76%), and VGG-Face (4.5% to 12%).

This work also performed and compared five methods for increasing the quantity of tabular data to enhance the detection of facial expressions of people with PD and those of healthy cohorts. It compared alternative methods to generate synthetic data and increase the number of PD instances in the dataset of facial features. The approaches used were: samples from a marginal and uniform distribution, Bayesian network (ZHANG *et al.*, 2017), Conditional Generative Adversarial Network (CGAN) (MIRZA; OSINDERO, 2014) and Conditional Tabular Generative Adversarial Network (XU *et al.*, 2019). Finally, I combined the best method found previously with Test-Time Augmentation (TTA) in order to improve the model inference.

Therefore, the main novelties of this work were:

- introducing approaches of data augmentation for tabular data applied to extracted facial features to improve the performance of identifying people with PD and healthy people by facial expression;
- combining synthetic data generated by CGAN to increase training data with TTA.

### 3.1.1 Methods

This subsection presents the methodology considered for the experiments, composed of theoretical background, grid-search-based cross-validation for hyperparameter optimization, data augmentation with synthetic data generation, and test-time data augmentation (WANG *et al.*, 2019).

#### 3.1.1.1 Dataset

The dataset obtained from<sup>1</sup> comprises features of the video analysis reported by Ali *et al.* (2021). The feature set includes the variance of the facial action unit (AU), which is the individual muscle movement components established systemically by Paul Ekman in FACS (EKMAN; FRIESEN, 1978).

This study has only investigated the facial action units of facial expressions which were in response to stimuli; spontaneous facial expressions have not been investigated. This has the advantage because it greatly reduces the complexity of the analysis. All patients were asked to make three facial expressions: smiling, disgusted, and surprised. Each facial expression was recorded in a different video. The videos were acquired utilizing the PARK<sup>2</sup> (Parkinson's Analysis with Remote Kinetic Tasks) (LANGEVIN *et al.*, 2019) online video recording tool and were analyzed using OpenFace (HAMMADI *et al.*, 2022) software, which automatically provides the facial AU values of each frame. Afterward, the AU variance was computed to

<sup>1</sup> available at <[https://github.com/mali7/PARK\\_facial\\_mimic](https://github.com/mali7/PARK_facial_mimic)>

<sup>2</sup> available at <[www.parktest.net](http://www.parktest.net)>

indicate how much facial muscle movement occurs while showing a facial expression. Notice that the authors selected a subset of three AUs to represent each facial expression: (i) smile is associated by AU01 (Inner Brow Raiser), AU06 (Cheek Raiser), and AU12 (Lip Corner Puller); (ii) disgusting is associated with AU04 (Brow Lowerer), AU07 (Eye Lid Tightener), and AU09 (Nose Wrinkler); and (iii) surprise is associated with AU01 (Inner Brow Raiser), AU02 (Outer Brow Raiser), and AU04 (Brow Lower).

Finally, the dataset was developed by concatenating the variances of the AUs mentioned above, in which each row corresponds to a patient. In my study, I divided the dataset into training and test sets using similar proportions, as seen in Table. 8.

Table 8 – Amount of images per training/test sets.

Type of images	Amount
PD (training)	30
Non-PD (training)	383
PD (test)	7
Non-PD (test)	97

### 3.1.1.2 Hyperparameter Optimization

The first experiment employed a grid search, i.e., a comprehensive search of parameters across a subset of the training algorithm’s space. Grid-search is a widespread method that scikit-learn (PEDREGOSA *et al.*, 2011a) includes as part of *GridSearchCV*<sup>3</sup>.

The pipeline employed in this work comprises three main steps: oversampling, constant normalization, and machine learning-based classification. SMOTE (CHAWLA *et al.*, 2002), ADASYN (HE *et al.*, 2008), O<sup>2</sup>PF (PASSOS *et al.*, 2022; PASSOS *et al.*, 2020)<sup>4</sup>, and Random Sampler (BRANCO; TORGO; RIBEIRO, 2016) were considered as oversampling approaches in the search space, and Table 9 shows the search space considered for each ML technique.

Four classifiers were evaluated for the task — Support Vector Machine (SVM) (HEARST *et al.*, 1998), Logistic Regression (COX, 1958), Random Forest (PAL, 2005), and *K*-Nearest Neighbors (ALTMAN, 1992) — because they represent distinct machine learning paradigms, offering a comprehensive performance comparison. SVM and Logistic Regression are well-suited for linear or kernel-based tasks. Random Forest is an ensemble method that mitigates overfitting through bagging and captures nonlinearities, and KNN relies on proximity in feature space to classify new instances. This diverse selection covers different learning biases and

<sup>3</sup> available at <[https://scikit-learn.org/stable/modules/generated/sklearn.model\\_selection.GridSearchCV.html](https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html)>

<sup>4</sup> Available at <https://github.com/Leandropassosjr/OpfImb>

decision boundaries, thereby improving my search for the best-performing model in detecting hypomimia associated with PD. Model evaluation was conducted using stratified  $K$ -fold cross-validation (REFAELZADEH; TANG; LIU, 2009) with  $k = 10$ , repeated three times with different randomizations, and performance was measured using the F1-score on the training set.

Table 9 – Grid-search space configuration

Machine Learning Model	Parameters	Values
<b>Support Vector Machine</b> (HEARST <i>et al.</i> , 1998)	kernel	[ <i>linear, radial</i> ]
	gamma	[0.1, 0.6, 1.1, 1.6]
	$C$	[0.1, 2.1, 4.1, 6.1, 8.1]
<b>Logistic Regression</b> (COX, 1958)	solver	[ <i>liblinear</i> ]
	penalty	[ <i>l1, l2</i> ]
	$C$	[0.01, 0.1, 1, 10, 100]
<b>Random Forest Classifier</b> (PAL, 2005)	number of estimators	[5, 50, 100, 250]
	max depth	[5, 8, 10]
<b><math>K</math>-Nearest Neighbors</b> (ALTMAN, 1992)	number of neighbors	[3, 5, 7, 9, 11]
	weight function	[ <i>uniform, distance</i> ]

The pipeline that provided the best results was ADASYN and Logistic Regression with  $C=0.01$ , and this was selected for comparison to the CGAN-based synthetic data augmented.

### 3.1.1.3 Augmentation

The data for both labels, PD and healthy, was augmented by the same level. However, before using this data to train the model, the data needed to be balanced, which was performed using the ADASYN oversampling technique. With this enhanced and balanced dataset, I started to prepare my augmentation methods to increase the training set. To generate synthetic data from the marginal, uniform distribution and Bayesian network (ZHANG *et al.*, 2017), I used the implementation of <sup>5</sup> setting the privacy rate to 1.0. CGAN was implemented by me in PyTorch and trained with 700 epochs, with a batch size of 64 and a latent dimension of 100. The ADAM (KINGMA; BA, 2017) optimizer was used to train the generator and the discriminator, with learning rates of 0.0002 and decay rates of 0.1 and 0.999, respectively. The figures describing the architectures of the CGAN Generator and Discriminator is in, Figure 5 and Figure 6, respectively. Finally, CTGAN<sup>6</sup> was trained with 500 epochs, with a batch size of 64 and a latent dimension of 128, and optimized with learning rates of 0.0002 and decay rates of 0.1 and 0.999,

<sup>5</sup> available at <[https://github.com/daanknoors/synthetic\\_data\\_generation](https://github.com/daanknoors/synthetic_data_generation)>

<sup>6</sup> available at <<https://github.com/sdv-dev/CTGAN>>

respectively. The experiments were conducted using the training set with an Nvidia RTX 2060 GPU.

#### 3.1.1.4 Test-Time Augmentation

The experiments presented in this subsection were performed to test the performance of the model. A problem like detecting disease conditions typically require high sensitivity. Test-Time Augmentation (TTA) technique was used because it improves the model's sensitivity performance. It predicts the label using a voting approach, based on the most frequently predicted class among a set of slightly modified classifiers. This approach is beneficial for cases in which the model is uncertain about the prediction. One straightforward method to generate the slightly modified versions of a classifier is adding a modest amount of Gaussian noise. A diagram of TTA is shown in Fig 7.

The Gaussian noise was generated using NumPy's (HARRIS *et al.*, 2020) `normal()` function, which produces a vector of random Gaussian values with zero mean and a standard deviation proportional to its distribution, which stands for values between 0.01 and 0.3 optimized through the grid-search using a step of 0.01 to maximize the accuracy.

### 3.1.2 Results

The experimental results are presented in three sub-subsections: (i) model tuning assessment, (ii) data augmentation assessment, and (iii) test-time augmentation assessment. The system's performance was evaluated using confusion matrices, and accuracy, sensitivity, and specificity were computed.

#### 3.1.2.1 Model Tuning Assessment

The best tuning approach was the oversampling method ADASYN and Logistic regression with  $C = 0.01$  and  $l2$  penalty. The pipeline achieved a sensitivity of roughly 86% and a specificity of about 61% in the test set, whose results are reported in the confusion matrix in Table 42. Such results show a high false-positive rate, demonstrating a weak performance for masked faces prediction. Overall, it obtained a classification accuracy of 62%.

Ali *et al.* (ALI *et al.*, 2021) achieved 95.6% accuracy. However, when performing k-fold cross-validation, their testing fold consists of a mix of synthetic and real data. This difference may be attributed because they balanced the entire dataset before performing the k-fold. As a result, the dataset has about 90% of synthetic data, and large similarity within this group may lead to the large similarity between the training and test fold. On the other hand, my method only generated synthetic data for training to ensure there was a balanced training dataset, while the test dataset need not be balanced. Thus, testing with unbalanced data represents a real-world scenario.

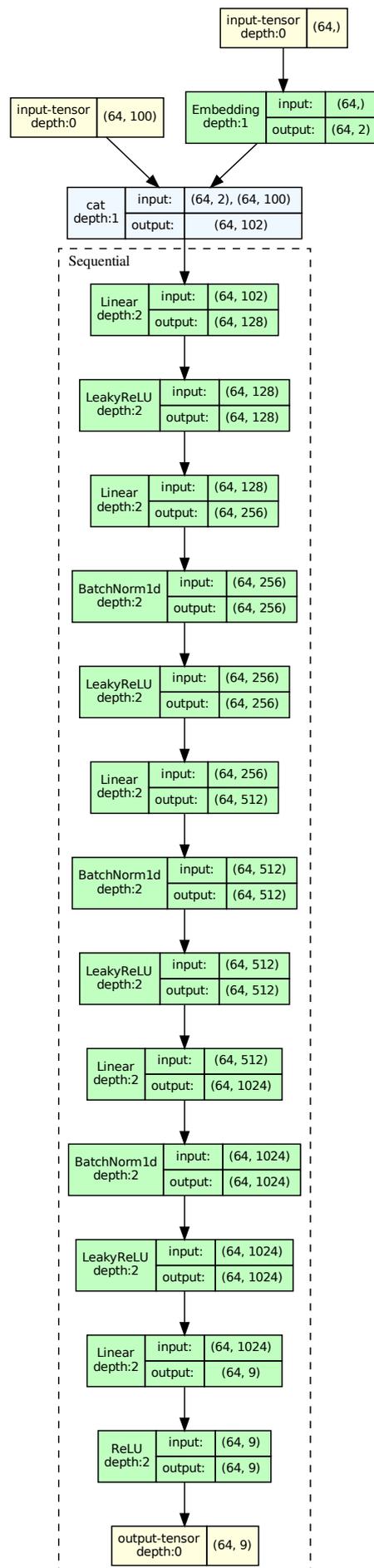


Figure 5 – Generator architecture.

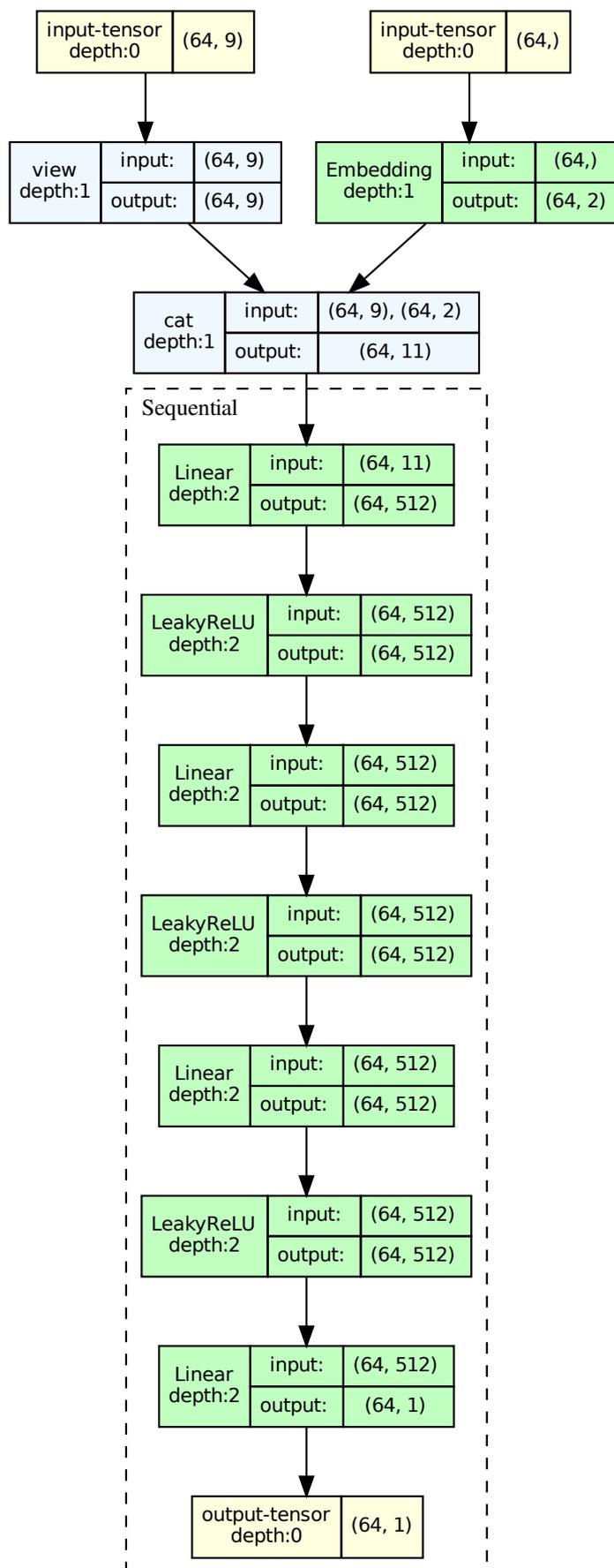


Figure 6 – Discriminator architecture.

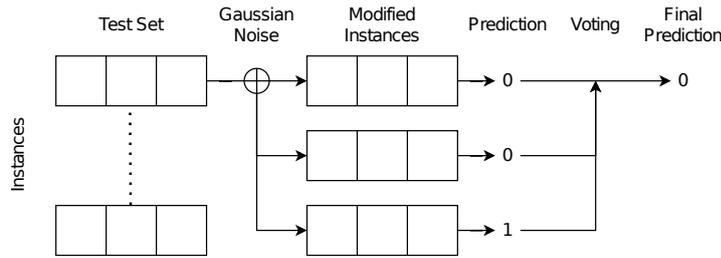


Figure 7 – Steps of Test-Time Augmentation

	PD	Non-PD
<b>PD</b>	6	1
<b>Non-PD</b>	38	59
<b>False Positives</b>		38
<b>False Negatives</b>		1
<b>Sensitivity</b>		85.71%
<b>Specificity</b>		60.82%
<b>Accuracy</b>		62.50%

Table 10 – Confusion matrix for Logistic Regression.

	PD	Non-PD
<b>PD</b>	6	1
<b>Non-PD</b>	27	70
<b>False Positives</b>		27
<b>False Negatives</b>		1
<b>Sensitivity</b>		85.71%
<b>Specificity</b>		72.16%
<b>Accuracy</b>		73.08%

Table 11 – Confusion matrix for Logistic Regression when introducing 1,600 synthetically augmented data.

	PD	Non-PD
<b>PD</b>	6	1
<b>Non-PD</b>	17	80
<b>False Positives</b>		17
<b>False Negatives</b>		1
<b>Sensitivity</b>		85.71%
<b>Specificity</b>		82.47%
<b>Accuracy</b>		82.69%

Table 12 – Confusion matrix for Logistic Regression when introducing 1,600 synthetically augmented data and TTA.

### 3.1.2.2 Data Augmentation Assessment

Figure 8 shows the test set accuracy when additional synthetic data generated using multiple algorithms were used in the training set. The best results were when CGAN was used to generate the synthetic data for training as seen in Figure 8(a). Overall, the accuracy improved when synthetic data was added to the training set, while recall/sensitivity remained constant. It is observed that there after an initial step improvement in the performance with an additional augmentation of 400 samples, the improvement was gradual and saturated on reaching a total of 1,600 augmented samples.

Table 11 provides the results of the experiments with 1,600 synthetic data. The number of false positives decreased from 0.61 to 0.72 when compared to Table 12. Consequently, specificity increased from 0.61 to 0.72. In general, there was a gradual growth in accuracy when employing data augmentation.

### 3.1.2.3 Test-Time Augmentation Assessment

Figure 9 illustrates the Logistic Regression's accuracy while varying the standard deviation between 0.01 and 0.3, considered to evaluate which value produces the best classification scenario, which was obtained with a standard deviation between 0.37 and 0.45.

Table 12 summarizes the evaluation of the best standard deviation, i.e., 0.45. According

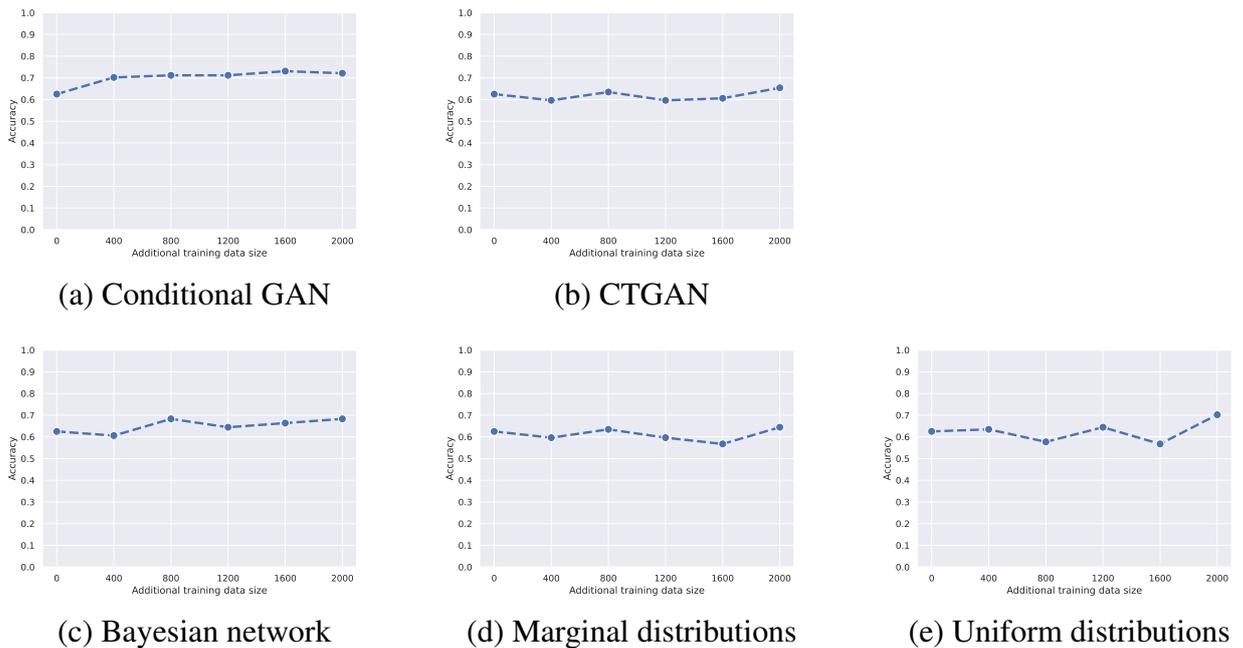


Figure 8 – Relationship of Accuracy and size of synthetic training data generated using the five algorithms.

to Tables 11 and 12, the false positive rate decreased from 27 to 17, consequently increasing specificity from 0.72 to 0.82. Additionally, the accuracy for identifying PD increased from 0.73 to 0.83, respectively.

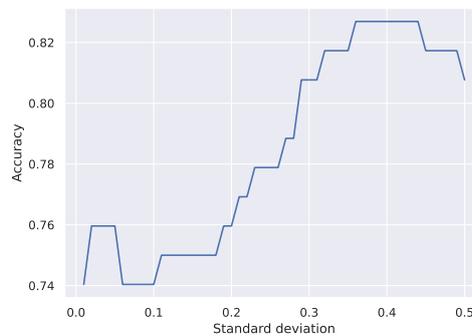


Figure 9 – Standard Deviation in TTA

### 3.1.3 Discussion

In summary, I first used grid search–based cross-validation to identify the best oversampling strategy, ADASYN, and machine learning approach, Logistic Regression. I then compared multiple synthetic data generators (Marginal, Uniform, Bayesian Network, CGAN, and CTGAN) to find the most effective for augmenting the training set. Finally, I applied Test-Time Augmentation by injecting Gaussian noise during inference, generating multiple slightly altered predictions and voting on the final classification. This method ensures that the model is trained on real and synthetic data while being evaluated only on real data for an unbiased performance assessment.

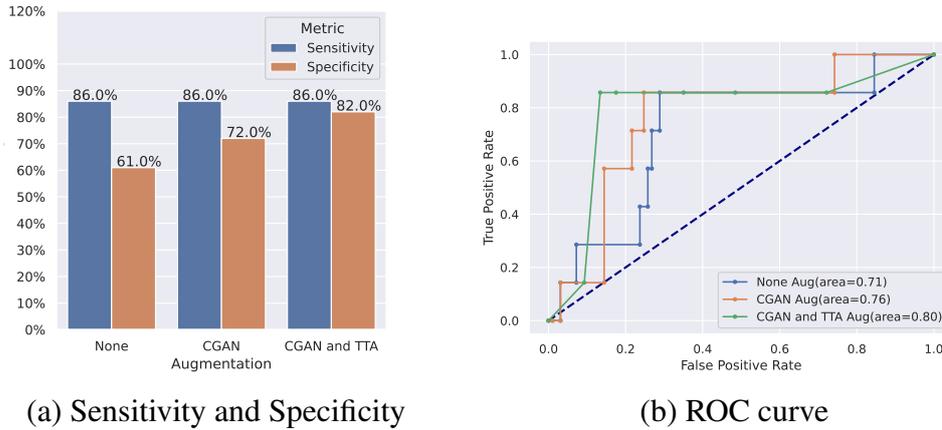


Figure 10 – Performance of data augmentation.

Table 6 provides a comparison of major works reported in the literature and my study. The table shows that in the earlier studies, even when these have been tested using a balanced dataset, it does not represent real-world conditions. The prevalence of PD has been reported to be more than 1% for people aged over 60 (TYSNES; STORSTEIN, 2017; BRAKEDAL *et al.*, 2022; MARRAS *et al.*, 2018). In the current study, I used the population-based prevalence of the disease for the age-matched group. Thus, while my results are comparable and better than most other studies, this comparison is not appropriate. Work by Ali *et al* (ALI *et al.*, 2021) is most suitable for the comparison because of the same dataset, and features. They provided the access to the dataset and the code. The accuracy reported by Ali *et al* (ALI *et al.*, 2021) was 95%. However, on closer inspection, I found that they had tested the model using a mix of real and synthetic data which can lead to a bias, especially when the number of real data points is very small. When their technique was tested where only the training data incorporates synthetic data while testing is real data only, the accuracy was around 62%. When I used my proposed method for developing additional data for training and tested using real data as it would be accepted in real-world situations, the accuracy improved to about 83%. The Sensitivity of the system remained at 86%, and the specificity increased from around 61% to nearly 82%, as shown in Fig 2(a). Work reported by Hao *et al* (HOU *et al.*, 2021) 2021 compared the performance of geometric and texture features of the videos of 140 people and report 88% results when using texture analysis. However, their dataset is not public, and I found that the description of the analysis was not sufficient for accurate comparison.

This work has shown that when testing with real data, the performance is much lower than when tested with synthetic data. This work has shown that when the training data consists of real and synthetic data generated using CGAN and TTA, the model has sufficient diversity and is suitable for being tested for real data.

### 3.1.3.1 Novelty of this study

This study presents a novel approach to addressing the challenge of small and unbalanced datasets in the video-based detection of hypomimia in PD. Unlike previous works that applied synthetic data to both training and test sets, potentially leading to data leakage and optimistic performance metrics, my study delineates the augmentation process to improve generalization without introducing bias. The key innovation lies in the integration of a CGAN for generating synthetic tabular data, which is then strategically combined with TTA. This dual approach significantly enhances classification accuracy while maintaining the integrity of real-world testing conditions. Notably, my methodology achieved an 83% accuracy rate while preserving a high sensitivity of 85.7% and improving specificity from 60.8% to 82.5%, a crucial factor for reducing false positives in clinical applications.

A comparison with other similar work by Ali et al ([ALI et al., 2021](#)) is shown in Table 6. The results show that when using the features proposed by Ali et al ([ALI et al., 2021](#)), and with data that uses synthetic data for both, training and testing as performed by them, the accuracy is 95%. However, when only the training data incorporates synthetic data while testing is with real data only, the accuracy is around 60%. My findings offer a novel perspective on how to improve the performance of video-based detection for hypomimia in people with PD, augmenting a tabular dataset. When I used my proposed method for developing additional data for training, and the real data for testing, the accuracy improved to about 83%. The Sensitivity of the system remained at 86%, and the specificity increased from around 61% to nearly 82%, as shown in Figure 2(a).

In the context of previous studies presented in Table 6, the results obtained in this work seem to be better than 3 works. However, the cited results cannot be fully comparable to ours due to the fact that all of my testing set has a prevalence of 7% patients with Parkinson's. The prevalence represents a real scenario in which the distribution is imbalanced. It is impossible to state unequivocally which of the methods will be characterized in practice by better generalization abilities without performing additional comparative studies.

Beyond the methodological advancements, this work also provides new scientific insights into how facial action unit variances contribute to hypomimia detection. By leveraging CGANs to generate additional training data, I demonstrated that synthetic augmentation can capture the subtle dynamics of facial muscle movements associated with PD, reinforcing the feasibility of AI-driven facial expression analysis for early-stage screening. Moreover, the introduction of TTA as a means of improving model underscores the importance of leveraging probabilistic variations during inference, improving the performance. These contributions collectively advance the field by demonstrating that a well-structured augmentation strategy is promising in mitigating the limitations of small datasets, ultimately paving the way for more scalable and clinically applicable AI-based PD detection systems.

Table 13 – Comparison of Model Types in Prior Work.

Ref.	Model Type
(RAJNOHA <i>et al.</i> , 2018) 2018	Handcrafted
(NOVOTNY <i>et al.</i> , 2022) 2022	Handcrafted
(CALVO-ARIZA; GÓMEZ-GÓMEZ; OROZCO-ARROYAVE, 2022) 2022	Handcrafted
(SONAWANE; SHARMA, 2021) 2021	Deep learning
(GOMEZ <i>et al.</i> , 2023) 2023	Deep learning
(HOU <i>et al.</i> , 2021) 2021	Handcrafted
(GOMEZ <i>et al.</i> , 2021) 2021	Handcrafted
(VALENZUELA <i>et al.</i> , 2022) 2022	Deep Learning
(JAKUBOWSKI <i>et al.</i> , 2021) 2021	Deep Learning
(JIN <i>et al.</i> , 2020) 2020	Handcrafted
(ALI <i>et al.</i> , 2021) 2021	Handcrafted (synthetic and real data in test)
(ALI <i>et al.</i> , 2021) 2021	Handcrafted (only real data in test)
ours	Handcrafted (CGAN aug)
ours	Handcrafted (CGAN aug and TTA)

Table 14 – Comparison of Accuracy in Prior Work.

Ref.	Accuracy
(RAJNOHA <i>et al.</i> , 2018) 2018	67%
(NOVOTNY <i>et al.</i> , 2022) 2022	78%
(CALVO-ARIZA; GÓMEZ-GÓMEZ; OROZCO-ARROYAVE, 2022) 2022	80%
(SONAWANE; SHARMA, 2021) 2021	85%
(GOMEZ <i>et al.</i> , 2023) 2023	87%
(HOU <i>et al.</i> , 2021) 2021	88%
(GOMEZ <i>et al.</i> , 2021) 2021	88%
(VALENZUELA <i>et al.</i> , 2022) 2022	92%
(JAKUBOWSKI <i>et al.</i> , 2021) 2021	94%
(JIN <i>et al.</i> , 2020) 2020	99%
(ALI <i>et al.</i> , 2021) 2021	95%
(ALI <i>et al.</i> , 2021) 2021	62%
ours	<b>73%</b>
ours	<b>83%</b>

Table 15 – Comparison of Percentage of PD Data in Prior Work.

Ref.	Percentage of PD Data
(RAJNOHA <i>et al.</i> , 2018) 2018	50%
(NOVOTNY <i>et al.</i> , 2022) 2022	55%
(CALVO-ARIZA; GÓMEZ-GÓMEZ; OROZCO-ARROYAVE, 2022) 2022	57%
(SONAWANE; SHARMA, 2021) 2021	52%
(GOMEZ <i>et al.</i> , 2023) 2023	55%
(HOU <i>et al.</i> , 2021) 2021	50%
(GOMEZ <i>et al.</i> , 2021) 2021	55%
(VALENZUELA <i>et al.</i> , 2022) 2022	50%
(JAKUBOWSKI <i>et al.</i> , 2021) 2021	50%
(JIN <i>et al.</i> , 2020) 2020	51%
(ALI <i>et al.</i> , 2021) 2021	50%
(ALI <i>et al.</i> , 2021) 2021	<b>7%</b>
ours	<b>7%</b>
ours	<b>7%</b>

### 3.1.3.2 Limitations

This study has two main limitations. First, I have only investigated the single, cross-sectional database of people with PD, and the dataset does not have ethnic and demographic variations, nor repeats for an individual. Thus, it does not allow the investigation of differences due to repetition and repeatability of the analysis. Second, I only had the facial action-unit data pre-extracted by Ali *et al.* (2021) available; the raw videos or additional information (such as lighting, camera angle, and participant details) are not available due to the ethical considerations for such data. Besides that, without frame-level data or contextual cues, I could not perform a

qualitative review of false positives and negatives to highlight systematic biases or subtle model failures.

While I have overcome the shortcomings of earlier works, my work is limited because it has been tested on a single dataset, and thus I cannot claim the generalizability of the model. It is essential to develop a new dataset with recordings of people of different ethnicity and demographics, repeated multiple times to study the patients over a period of time. That will check its suitability for being used for the detection of hypomimia and for monitoring the progression of PD. Another limitation is that I am not using a validation set because the data size is small. In this scenario, I am using the test set to find the best parameters of CGAN and TTA. However, for a preliminary augmentation study, I observed an increase in the performance of the machine learning model in distinguishing healthy people and people with PD based on facial expressions.

### **3.1.4 Conclusion**

This study is a progression towards creating robust affordable computing systems that can effectively analyze and measure facial expressions in individuals with PD who have various issues. The limitation of the small size and unbalanced dataset of PD facial expressions was overcome using CGAN-based data augmentation. Experimental results show a specificity of 0.72 and an accuracy in classifying non-PD individuals of 0.83. This has the potential of assisting clinicians with the assessment of videos to identify people with PD using videos obtained using a smartphone and may also be used for population screening. However, there are several limitations to my study that suggest areas for future research.

It is essential to investigate factors such as differences in ethnicity and demographics and to conduct longitudinal studies to check the effect of repeating the experiments. Besides that, future research could apply a different software to assess the head pose, instead of OpenFace. Finally, the augmentation method is promising to apply to different methods that use tabular data/features extracted from the face.

### **3.1.5 CRediT authorship contribution statement**

I am grateful for the support and collaboration that made this work possible. As the lead investigator, I was responsible for the research, design, methodology development, and writing this study. I thank Quoc C. Ngo for project administration, validation, and manuscript review; Leandro A. Passos for software validation and manuscript review; Leonardo S. Oliveira for review analysis; João P. Papa for securing funding and project oversight; and Dinesh Kumar for his conceptual guidance, supervision and manuscript review. Their contributions were essential to the success of this work.

## 3.2 Action Units for Facial Expression

Stroke is a severe neurological emergency characterized by a sudden loss of blood flow to the brain, potentially leading to irreversible brain damage, disability, or death if not promptly addressed. Time is of the essence in stroke cases; early detection and intervention are critical to improving patient outcomes. Research indicates that nearly 13% of stroke cases go undiagnosed in emergency departments (NEWMAN-TOKER *et al.*, 2014), and this rate increases dramatically in smaller hospitals, where 65% of patients without a documented neurology examination have missed strokes (ARCH *et al.*, 2016). In rural or regional centers, the percentage can be even higher (SHAKE; OOSTEMA; JONES, 2018). Since most strokes occur at home, the first responders are often paramedics or ambulance crews operating under less-than-ideal conditions (NGUYEN-HUYNH; JOHNSTON, 2007). This situation underscores the urgent need for accessible tools—ideally deployable on smartphones—that can assist paramedics in rapid and accurate stroke detection outside of hospital settings (BAT-ERDENE; SAVER, 2021).

ALS is another debilitating neurological condition, characterized by progressive muscle weakness and the inability to perform voluntary muscle contractions. This affects essential functions such as chewing, speaking, walking, and facial expressions. Early symptoms of ALS include reduced facial expressiveness (WIJESEKERA; LEIGH, 2009; KIERNAN *et al.*, 2011), but recognizing these subtle changes can be subjective and often leads to misdiagnosis due to individual differences in facial expressions (OH *et al.*, 2016; AHO-ÖZHAN *et al.*, 2016; CARELLI *et al.*, 2021). Therefore, there is a significant opportunity for computerized analysis of facial expressions to improve the reliability of early ALS diagnosis.

Facial expressions are a vital component of human communication and serve as important indicators of neurological health. Detecting subtle changes in facial expressions can be challenging, prompting interest in computerized systems like the FACS to enhance diagnostic accuracy. FACS categorizes facial movements using AUs, each corresponding to specific muscle actions, providing an objective and systematic method for analyzing facial expressions (FRIESEN; EKMAN, 1978).

This work introduces an AI-based approach that leverages AU analysis to perform binary classifications between healthy controls (HC) and individuals with either Post-Stroke (PS) or ALS, specifically focusing on distinguishing HC vs. PS and HC vs. ALS. Recent advancements have shown the potential of AU analysis in medical diagnostics, from detecting hypomimia in Parkinson's disease to differentiating neurological impairments (OLIVEIRA *et al.*, 2023). Building on these developments, the goal is to create a solution that can be deployed on smartphones, providing paramedics with a tool for real-time stroke detection in the field, and assisting neurologists in diagnosing ALS more accurately. This could significantly reduce the rate of missed diagnoses, leading to faster interventions and better patient outcomes.

To the best of my knowledge, this is the first study to report on using AUs from facial

videos for the binary classification of HC vs. PS and HC vs. ALS, highlighting the broader potential of AU analysis.

### 3.2.1 Methods

The methodology employed involved a systematic approach to computerized analysis of facial expressions using AUs. This process entailed capturing facial videos, extracting AUs, and applying machine learning models to classify facial expressions corresponding to healthy individuals and patients with neurological conditions.

#### 3.2.1.1 Dataset

The Toronto Neuroface dataset (BANDINI *et al.*, 2020) investigated in this study, provided by Toronto University, is a single center, cross-sectional dataset with 261 facial expression videos collected from 36 people. It comprises 11 HC, 14 PS, and 11 people with ALS. The videos were recorded when the participants were performing a set of speech and non-speech facial expression tasks. Table 16 reports the demographic and clinical summary of the participants, including their ages and the duration measured in months starting from when a stroke occurs. The literature review indicates that this is the only dataset available for investigating facial expressions in patients with neurological diseases.

Facial expression data consists of 9 specific facial expressions, repeated multiple times and related tasks that are shown in Table 17. These are: (i) pretending to kiss a baby (KISS), (ii) maximum opening of the jaw (OPEN), pretending to smile with tight lips (SPREAD), reproductions of the syllable */pal* (PA) as quickly as possible on a single breath, speaking of the word */patakal* (PATAKA) as fast as possible on a single breath, speaking the sentence “*Buy Bobby a Puppy*” (BBP) at a comfortable speaking rate and intensity, pretending to blow a candle (BLOW), repetitions of raising the eyebrows (BROW) and making a big smile (BIGSMILE). Participants were advised to take rests between activities to avoid weariness. Nonetheless, not all the participants could complete all the exercises.

Table 16 – Demographic and clinical information for the three participant groups, including the duration in months from the onset of symptoms for ALS or from the occurrence of a stroke for PS.

Group	Age (years)	Duration (months)	ALSFRS-R
HC	63.2 ± 14.3	–	–
ALS	61.5 ± 8.0	49.6 ± 31.6	34.8 ± 5.0
PS	64.7 ± 14.7	19.4 ± 34.2	–

#### 3.2.1.2 Feature Extraction

The AUs were extracted from individual recordings using the Python Facial Expression Analysis Toolbox (Py-Feat) (CHEONG *et al.*, 2023), which has been built by training an

Table 17 – Toronto Neuroface Dataset Distribution.

Task	ALS	HC	Stroke
<b>SPREAD</b>	11	11	14
<b>KISS</b>	11	11	14
<b>OPEN</b>	11	11	14
<b>PA</b>	10	11	14
<b>PATAKA</b>	10	11	14
<b>BBP</b>	9	11	14
<b>BLOW</b>	6	7	11
<b>BROW</b>	4	4	7
<b>BIGSMILE</b>	4	3	3

XGBoost (CHEN; GUESTRIN, 2016) classifier on a Histogram of Oriented Gradients extracted of the AUs and from five distinct datasets, i.e., BP4D (ZHANG *et al.*, 2014), DISFA (MAVADATI *et al.*, 2013), CK+ (LUCHEY *et al.*, 2010), UNBC-McMaster shoulder pain (LUCHEY *et al.*, 2011), and AFF-Wild2 (KOLLIAS; ZAFEIRIOU, 2018). The model outputs a continuous probability prediction for each AU in Table 18. Figure 11 shows a facial image’s facial landmarks, facial box, and action units extracted by Py-Feat.

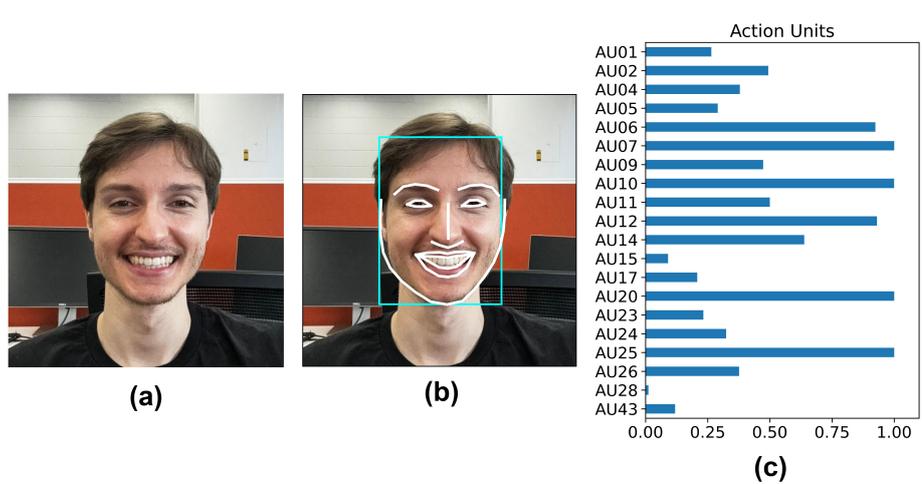


Figure 11 – Example of Py-Feat extraction: (a) original image, (b) facial box and landmarks, (c) action units.

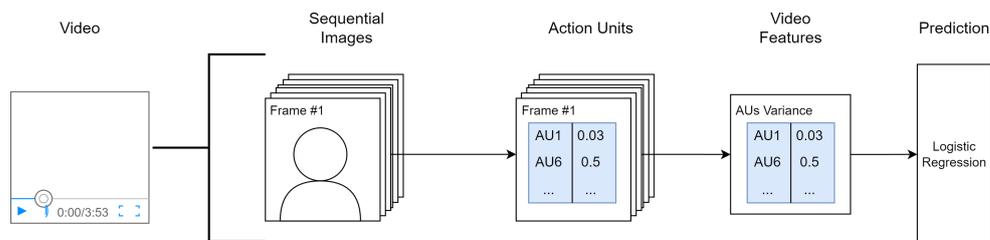


Figure 12 – Feature extraction flow.

Further, the following AUs’ statistical features are extracted from every video frame: mean, maximum, minimum, and standard deviation (Fig 12). The minimum and maximum

Table 18 – List of AU, action descriptors, and underlying facial muscles.

AU Number	Action Descriptor	Muscular Basis
1	Inner brow raiser	Frontalis (Pars Medialis)
2	Outer brow raiser	Frontalis (Pars Lateralis)
4	Brow lowerer	Depressor Glabellae, Depressor Supercilii, Corrugator Supercilii
5	Upper lid raiser	Levator Palpebrae Superioris
6	Cheek raiser	Orbicularis Oculi (Pars Orbitalis)
7	Lid tightener	Orbicularis Oculi (Pars Palpebralis)
9	Nose wrinkler	Levator Labii Superioris Alaeque Nasi
10	Upper lip raiser	Levator Labii Superioris
11	Nasolabial deepener	Zygomaticus Minor
12	Lip corner puller	Zygomaticus Major
14	Dimpler	Buccinator
15	Lip corner depressor	Depressor Anguli Oris
17	Chin raiser	Mentalis
20	Lip stretcher	Risorius
23	Lip tightener	Orbicularis Oris
24	Lip pressor	Orbicularis Oris
25	Lip part	Depressor Labii, Orbicularis Oris
26	Jaw drop	Maseter, Temporal and Internal Pterygoid relaxed
28	Lip suck	Orbicularis Oris
43	Eyes closed	Levator Palpebrae Superioris (relaxation)

values capture the extremes of facial muscle movement, which is critical in identifying muscle weakness or hyperactivity associated with ALS and stroke. Variance highlights inconsistencies in muscle control, while the mean offers a baseline for comparing normal and abnormal muscle activity levels. Standard deviation helps distinguish between normal facial expression variability on a scale different from variance. While variance and standard deviation are related, using them as features in logistic regression models can yield different results. These statistical tools provide diverse perspectives on facial muscle function.

### 3.2.1.3 Classification

The sample size of BROW and BIG SMILE were very small, and were not included in the study; balance even video sequences were used. Binary classification was performed to distinguish between HC individuals and ALS patients from each of the seven videos: KISS, OPEN, SPREAD, PA, PATAKA, BBP, and BLOW.

The classification pipeline used in this study consists of three steps: (i) standardization of features, where each feature is adjusted to have a mean of zero and a standard deviation equal one; (ii) removal of correlated features with Pearson correlation coefficients above 0.7. In this step, the first identified feature was retained, and all subsequent correlated features were removed; and (iii) employing a machine learning approach for classification using Logistic Regression (COX, 1958).

Logistic Regression is a method used for clinical prediction modeling, as highlighted by Christodoulou et al. (CHRISTODOULOU *et al.*, 2019). They noted that despite the increasing popularity of Machine Learning (ML) algorithms, such as artificial neural networks, support vector machines, and random forests, Logistic Regression remains highly effective, especially in the medical domain.

The classification pipeline and grid-search were carried out using the *Scikit-learn* (PEDREGOSA *et al.*, 2011b). However, *Scikit-learn* lacks comprehensive statistical analysis capabilities for Logistic Regression, leading to the use of the *Statsmodels* (SEABOLD; PERKTOLD, 2010) library for assessing feature importance. The search space and parameters used in *Scikit-learn* for Logistic Regression are detailed in Table 19 for the HC vs. ALS, while Table 20 shows those used for the HC vs. PS. The choice of grid search intervals for  $C$  differs between these comparisons to account for the distinct characteristics of each disease. For HC vs. ALS, a broader range [0.01, 0.1, 1, 10, 100] is used to identify the best regularization level, while for HC vs. PS, a finer grid [0.0, 0.3, 0.6, 0.9, 1.2, 1.5, ..., 100.2] allows for more precise tuning due to potentially subtler differences between the groups. For the HC vs. ALS analysis using *Statsmodels*, the  $\alpha$  parameter was set to 1, employing the  $L_1$  regularization method. When conducting the HC vs. PS analysis,  $\alpha$  was adjusted to approximately 0.3, using the  $L_1$  regularization method.

Table 19 – Configuration of Grid-Search Space for HC vs. ALS.

Parameters	Values
solver	[ <i>liblinear</i> ]
penalty	[ $l1$ , $l2$ , <i>elasticnet</i> ]
$C$	[0.01, 0.1, 1, 10, 100]

Table 20 – Configuration of Grid-Search Space for HC vs. PS.

Parameters	Values
solver	[ <i>liblinear</i> ]
penalty	[ $l1$ , $l2$ ]
$C$	[0.0, 0.3, 0.6, 0.9, 1.2, 1.5, ..., 100.2]

The assessment was carried out on a task-by-task basis, employing a "Leave-One-Out" k-fold cross-validation method, where k is equal to the total number of entries in the database. A distinct model was created for each instance in the training set and then evaluated. This method involves evaluating a model for every single instance, offering a more thorough and stronger evaluation of the model's effectiveness. The performance of the system was assessed based on accuracy, sensitivity, and specificity.

### 3.2.2 Results

This subsection presents the results obtained in the classification tasks. The experiments were conducted over the statistical information of 20 AUs computed for each task using the Logistic Regression classifier with hyperparameters optimized through a Grid-search approach to find the best model, i.e., the one that maximizes the accuracy.

Table 21 presents the results for HC vs. ALS across the seven facial expression tasks, with the highest accuracy of 0.91 achieved in the SPREAD task when using the minimum of

AUs. Similarly, Table 22 shows the best classification results for HC vs. PS, with the highest accuracy of 0.82 observed for the KISS and SPREAD tasks. Notably, there were considerable differences in performance across the various tasks and features in this comparison.

Table 21 – Logistic Regression - Comparative Performance in HC vs. ALS.

Task	Measure between frames	Acc	Sens	Spec
BBP	var	0.45	<b>1.00</b>	0.00
	mean	0.45	<b>1.00</b>	0.00
	min	0.45	<b>1.00</b>	0.00
	max	0.65	<b>0.78</b>	0.55
	std	0.45	<b>1.00</b>	0.00
PATAKA	var	0.48	<b>1.00</b>	0.00
	mean	0.57	0.60	0.55
	min	0.57	0.40	<b>0.73</b>
	max	0.62	0.50	<b>0.73</b>
	std	0.62	<b>0.80</b>	0.45
PA	var	0.48	<b>1.00</b>	0.00
	mean	0.48	<b>1.00</b>	0.00
	min	<b>0.71</b>	<b>0.80</b>	0.64
	max	0.48	<b>1.00</b>	0.00
	std	0.62	0.60	0.64
KISS	var	<b>0.82</b>	<b>0.73</b>	<b>0.91</b>
	mean	<b>0.73</b>	<b>0.73</b>	<b>0.73</b>
	min	<b>0.73</b>	<b>0.73</b>	<b>0.73</b>
	max	<b>0.82</b>	<b>0.73</b>	<b>0.91</b>
	std	<b>0.77</b>	0.64	<b>0.91</b>
OPEN	var	0.59	0.45	<b>0.73</b>
	mean	<b>0.77</b>	<b>0.82</b>	<b>0.73</b>
	min	<b>0.82</b>	<b>0.91</b>	<b>0.73</b>
	max	<b>0.73</b>	0.64	<b>0.82</b>
	std	0.59	0.45	<b>0.73</b>
SPREAD	var	<b>0.82</b>	<b>0.73</b>	<b>0.91</b>
	mean	<b>0.77</b>	<b>0.73</b>	<b>0.82</b>
	min	<b>0.91</b>	<b>1.00</b>	<b>0.82</b>
	max	0.68	<b>0.73</b>	0.64
	std	<b>0.82</b>	<b>0.73</b>	<b>0.91</b>
BLOW	var	0.62	0.50	<b>0.71</b>
	mean	0.46	<b>1.00</b>	0.00
	min	0.69	0.67	<b>0.71</b>
	max	0.62	0.67	0.57
	std	0.46	<b>1.00</b>	0.00

### 3.2.2.1 Feature importance

The predictive power of AUs in logistic regression is demonstrated through their coefficients, as seen in the SPREAD task for ALS patients. Table 23 highlights that AUs AU09, AU17, and AU11 have large coefficients for HC, indicating a strong predictive capacity for distinguishing HCs within the model. However, a high coefficient does not necessarily imply causation or universal relevance, as the significance of features can vary based on the dataset and context. The logistic regression analysis, conducted using the *Statsmodels* library, reveals a high Pseudo R-squared value of 0.8007, suggesting a well-fitting model based on 22 observations,

Table 22 – Logistic Regression - Performance in HC vs. PS.

Task	Measure between frames	Acc	Sens	Spec
<b>BBP</b>	var	0.59	0.64	0.55
	mean	0.64	0.45	<b>0.82</b>
	min	0.68	<b>0.73</b>	0.64
	max	<b>0.73</b>	<b>0.82</b>	0.64
	std	0.55	0.55	0.55
<b>PATAKA</b>	var	0.45	0.45	0.45
	mean	0.59	0.55	0.64
	min	<b>0.73</b>	0.64	<b>0.82</b>
	max	0.68	<b>0.73</b>	0.64
	std	0.45	0.45	0.45
<b>PA</b>	var	0.36	0.36	0.36
	mean	0.50	0.45	0.55
	min	<b>0.77</b>	<b>0.73</b>	<b>0.82</b>
	max	0.64	0.64	0.64
	std	0.41	0.36	0.45
<b>KISS</b>	var	<b>0.77</b>	<b>0.82</b>	<b>0.73</b>
	mean	<b>0.73</b>	<b>0.73</b>	<b>0.73</b>
	min	<b>0.73</b>	<b>0.73</b>	<b>0.73</b>
	max	0.68	<b>0.73</b>	0.64
	std	<b>0.82</b>	<b>0.91</b>	<b>0.73</b>
<b>OPEN</b>	var	0.68	<b>0.73</b>	0.64
	mean	0.64	0.55	<b>0.73</b>
	min	<b>0.77</b>	<b>0.73</b>	<b>0.82</b>
	max	0.64	0.64	0.64
	std	0.68	<b>0.73</b>	0.64
<b>SPREAD</b>	var	<b>0.77</b>	<b>0.73</b>	<b>0.82</b>
	mean	0.59	0.55	0.64
	min	0.68	0.64	<b>0.73</b>
	max	<b>0.82</b>	<b>0.73</b>	<b>0.91</b>
	std	<b>0.77</b>	<b>0.73</b>	<b>0.82</b>
<b>BLOW</b>	var	<b>0.73</b>	<b>0.88</b>	0.57
	mean	<b>0.73</b>	<b>0.75</b>	<b>0.71</b>
	min	<b>0.73</b>	<b>0.75</b>	<b>0.71</b>
	max	0.60	<b>0.75</b>	0.43
	std	0.60	<b>0.75</b>	0.43

with a statistically significant LLR p-value of 0.0001804. Notably, AU09 (Levator labii superioris alaquae nasi) stands out with a coefficient of -2.0042 and a p-value of 0.027, indicating a significant negative relationship with the dependent variable. In contrast, other predictors such as AU04 (Depressor Glabellae, Depressor Supercilli, Corrugator), AU11 (Zygomatic Minor), AU14 (Buccinator), AU17 (Mentalis), and AU25 (Depressor Labii, Relaxation of Mentalis, Orbicularis Oris) do not show statistical significance, suggesting a lesser impact in this model.

Table 23 – Logistic Regression Analysis - HC vs ALS.

<b>Dep. Variable:</b>	y	<b>No. Observations:</b>	22			
<b>Model:</b>	Logit	<b>Df Residuals:</b>	16			
<b>Method:</b>	MLE	<b>Df Model:</b>	5			
<b>Date:</b>	Tue, 13 Jun 2023	<b>Pseudo R-squ.:</b>	0.8007			
<b>Time:</b>	17:09:49	<b>Log-Likelihood:</b>	-3.0399			
<b>converged:</b>	True	<b>LL-Null:</b>	-15.249			
<b>Covariance Type:</b>	nonrobust	<b>LLR p-value:</b>	0.0001804			
	<b>coef</b>	<b>std err</b>	<b>z</b>	<b>P&gt;  z </b>	<b>[0.025</b>	<b>0.975]</b>
<b>AU04</b>	-0.1191	0.886	-0.135	0.893	-1.855	1.617
<b>AU09</b>	-2.0042	0.903	-2.219	0.027	-3.775	-0.234
<b>AU11</b>	-0.6289	1.231	-0.511	0.609	-3.041	1.783
<b>AU14</b>	0.1488	0.808	0.184	0.854	-1.435	1.733
<b>AU17</b>	-1.0058	0.999	-1.007	0.314	-2.964	0.953
<b>AU25</b>	-0.1091	1.078	-0.101	0.919	-2.221	2.003

In the case of stroke patients, grid search results show that the highest accuracy in differentiating between HCs and patients with PS was achieved in the KISS and SPREAD task. Table 24 summarizes the logistic regression results of KISS task, excluding all AUs with coefficients equal to zero. Notably, AUs AU07, AU23, and AU20 have high coefficients, indicating strong predictive power for PS (COX, 1958). However, only AU07, associated with the Orbicularis oculi and pars palpebralis muscles, has a p-value around 0.05, suggesting a statistically significant relationship with PS cases. The model exhibits a Pseudo R-squared value of approximately 0.89, indicating an excellent fit, and an LLR p-value near 0.0002, demonstrating the model's effectiveness in predicting the response variable.

Table 24 – Logistic Regression Results - HC vs PS.

<b>Dep. Variable:</b>	y	<b>No. Observations:</b>	25			
<b>Model:</b>	Logit	<b>Df Residuals:</b>	16			
<b>Method:</b>	MLE	<b>Df Model:</b>	8			
<b>Date:</b>	Sat, 12 Aug 2023	<b>Pseudo R-squ.:</b>	0.8908			
<b>Time:</b>	11:39:19	<b>Log-Likelihood:</b>	-1.8733			
<b>converged:</b>	True	<b>LL-Null:</b>	-17.148			
<b>Covariance Type:</b>	nonrobust	<b>LLR p-value:</b>	0.0001689			
	<b>coef</b>	<b>std err</b>	<b>z</b>	<b>P&gt;  z </b>	<b>[0.025</b>	<b>0.975]</b>
<b>AU23</b>	<b>3.0715</b>	1.779	1.726	0.084	-0.416	6.559
<b>AU07</b>	<b>2.6518</b>	1.388	1.910	<b>0.056</b>	-0.070	5.373
<b>AU20</b>	<b>1.6571</b>	1.531	1.082	0.279	-1.344	4.658
<b>AU25</b>	0.1495	1.771	0.084	0.933	-3.321	3.620
<b>AU02</b>	0.0779	1.390	0.056	0.955	-2.647	2.803
<b>AU05</b>	-0.0539	1.024	-0.053	0.958	-2.060	1.952
<b>AU09</b>	-0.5249	1.401	-0.375	0.708	-3.271	2.221
<b>AU06</b>	-1.1579	1.061	-1.091	0.275	-3.238	0.922
<b>AU04</b>	-1.9283	1.140	-1.692	0.091	-4.162	0.305

### 3.2.3 Discussion

This study has demonstrated the potential of using AUs derived from facial videos to identify neurological conditions, specifically ALS and post-stroke patients. Utilizing facial AUs allows for the assessment of clinically observable changes in facial muscle activity, which are often affected in neurological diseases due to muscle stiffness or weakness. Previous research has indicated that the variance of AUs can distinguish between healthy individuals and patients with conditions like Parkinson's disease (OLIVEIRA *et al.*, 2023; ALI *et al.*, 2021). My work extends these findings by applying AU-based analysis to ALS and post-stroke patients, highlighting its effectiveness in differentiating between healthy controls and affected individuals.

One significant advantage of my method is that it does not require manual video segmentation, calibration of camera-to-face distance, or the use of specialized equipment such as infrared cameras. Unlike prior studies that necessitated manual intervention or specialized imaging modalities (BANDINI *et al.*, 2018c; BANDINI *et al.*, 2018b), my approach solely relies on RGB videos captured with a single camera, making it suitable for deployment as a smartphone application. This accessibility enhances the practicality of the method for use by clinicians, first responders, and even patients in remote or underserved areas.

Moreover, my method focuses on the use of AUs, which are directly interpretable in terms of specific facial muscle groups (SCHIMMEL *et al.*, 2011; SCHIMMEL *et al.*, 2017). This interpretability is crucial for clinical relevance, as it allows for the identification of particular facial expressions and muscles that are most affected by neurological conditions. The findings underscore the value of studying facial expressions and associated symptoms in neurological diseases, potentially aiding in the development of protocols for screening and assessment.

However, there are limitations to my study. The small sample size and the exclusive association with a single hospital may introduce biases and limit the generalizability of the findings. Additionally, the lack of longitudinal data prevents evaluation of long-term consistency and repeatability of the results. Privacy concerns related to the use of facial videos in healthcare also pose challenges for data management and scalability. Addressing these issues is crucial for the success of AI-driven diagnostics to ensure they are unbiased and accurate.

Future research should focus on conducting comparative studies across different disease stages, incorporating larger and more diverse datasets, and including longitudinal assessments to track disease progression. Integrating patient-reported outcomes alongside clinical assessments could provide a more holistic understanding of the disease's impact on patients' daily lives. The development of remote smartphone-based video assessments offers promising avenues for improving accessibility and timely intervention, particularly in areas with limited access to specialized care (ABBAS *et al.*, 2021).

### 3.2.3.1 Novelty of this study

The present work advances the field of computerized facial expression analysis by introducing a unified, automated framework for differentiating between healthy individuals and those with either ALS or post-stroke conditions. By relying on RGB recordings—thereby eliminating the need for specialized hardware—my approach harnesses AUs to capture subtle orofacial movements. These AUs are then translated into interpretable statistical indices (mean, minimum, maximum, and standard deviation), quantifying facial muscle activity more directly under various tasks. Notably, for ALS detection, I achieved the highest classification accuracy of 0.91 using the SPREAD expression, whereas, for post-stroke identification, both KISS and SPREAD proved the most effective in isolating key muscular deficits. These results underscore the methodology’s feasibility for widespread clinical or first-responder use, given its low operational complexity and readiness for mobile-friendly deployment.

Furthermore, the study explores how distinct neurological conditions manifest in facial expressions. The interpretability of AUs establishes a solid link between motor neuron degeneration (as in ALS) or brain lesion impact (as in stroke) and the measurable weakening of specific facial muscles, especially around the mouth region. In contrast to traditional, observer-reliant evaluations, my method furnishes objective, fine-grained assessments that can detect mild symptoms. This opens the door to more consistent patient monitoring, timely intervention, and more precise rehabilitation strategies. By mixing accessibility, interpretability, and clinical pertinence, this pilot work sets an encouraging foundation for large-scale, multi-center validations of automated facial analysis in neurodegenerative and neuromuscular care.

### 3.2.3.2 HC vs ALS

Table 25 compares the accuracy values of my method against results reported in the literature by Bandini et al. (BANDINI *et al.*, 2018c) and Gomes et al. (GOMES *et al.*, 2024a) for the classification of HC versus ALS patients, considering distinct tasks such as BBP, PATAKA, PA, KISS, OPEN, SPREAD, and BLOW. Gomes et al. (GOMES *et al.*, 2024a) employed Delaunay triangulation with Graph Neural Networks (WU *et al.*, 2020), while Bandini et al. (BANDINI *et al.*, 2018c) and Gomes et al. (GOMES *et al.*, 2024a) combined kinematics with Logistic Regression and Support Vector Machines (SVM) (CORTES; VAPNIK, 1995) with linear and RBF kernels, focusing more on geometry. My method utilized AUs as features with classification using Logistic Regression, likely for facial expression analysis through muscle movement classification. Each approach varies in accuracy, reflecting their effectiveness in specific classification or prediction tasks.

One notable advantage of my method is that it does not require manual video segmentation or normalization using a REST subtask. While Gomes et al. (GOMES *et al.*, 2024a) adopted an approach similar to Bandini et al. (BANDINI *et al.*, 2018c), their results were significantly distinct, possibly due to the unavailability of videos from the REST subtask, manual cropping

Table 25 – Comparison with prior works.

Task	Study	Approach	Accuracy
BBP	Gomes et al. (GOMES <i>et al.</i> , 2024a)	Delaunay triangulation + Graph Neural Networks	50.0%
	Gomes et al. (GOMES <i>et al.</i> , 2024a)	Kinematics + SVM-lin	45.0%
	Bandini et al. (BANDINI <i>et al.</i> , 2018c)	Kinematics + Logistic Regression	89.0%
	<b>Ours</b>	Action Units + Logistic Regression	60.0%
PATAKA	Gomes et al. (GOMES <i>et al.</i> , 2024a)	Delaunay triangulation + Graph Neural Networks	66.6%
	Gomes et al. (GOMES <i>et al.</i> , 2024a)	Kinematics + Logistic Regression	42.0%
	Bandini et al. (BANDINI <i>et al.</i> , 2018c)	Kinematics + SVM-lin	82.0%
	<b>Ours</b>	Action Units + Logistic Regression	67.0%
PA	Gomes et al. (GOMES <i>et al.</i> , 2024a)	Delaunay triangulation + Graph Neural Networks	57.1%
	Gomes et al. (GOMES <i>et al.</i> , 2024a)	Kinematics + SVM-lin	33.0%
	Bandini et al. (BANDINI <i>et al.</i> , 2018c)	Kinematics + Logistic Regression	77.0%
	<b>Ours</b>	Action Units + Logistic Regression	62.0%
KISS	Gomes et al. (GOMES <i>et al.</i> , 2024a)	Delaunay triangulation + Graph Neural Networks	68.1%
	Gomes et al. (GOMES <i>et al.</i> , 2024a)	Kinematics + Logistic Regression	50.0%
	Bandini et al. (BANDINI <i>et al.</i> , 2018c)	Kinematics + SVM-lin	55.0%
	<b>Ours</b>	Action Units + Logistic Regression	82.0%
OPEN	Gomes et al. (GOMES <i>et al.</i> , 2024a)	Delaunay triangulation + Graph Neural Networks	81.8%
	Gomes et al. (GOMES <i>et al.</i> , 2024a)	Kinematics + SVM-lin	77.0%
	Bandini et al. (BANDINI <i>et al.</i> , 2018c)	Kinematics + SVM-RBF	72.0%
	<b>Ours</b>	Action Units + Logistic Regression	77.0%
SPREAD	Gomes et al. (GOMES <i>et al.</i> , 2024a)	Delaunay triangulation + Graph Neural Networks	81.8%
	Gomes et al. (GOMES <i>et al.</i> , 2024a)	Kinematics + Logistic Regression	68.0%
	Bandini et al. (BANDINI <i>et al.</i> , 2018c)	Kinematics + SVM-lin	82.0%
	<b>Ours</b>	Action Units + Logistic Regression	<b>91.0%</b>
BLOW	Gomes et al. (GOMES <i>et al.</i> , 2024a)	Delaunay triangulation + Graph Neural Networks	38.4%
	Gomes et al. (GOMES <i>et al.</i> , 2024a)	Kinematics + SVM-RBF	45.0%
	Bandini et al. (BANDINI <i>et al.</i> , 2018c)	Kinematics + SVM-lin	65.0%
	<b>Ours</b>	Action Units + Logistic Regression	69.0%

of frames, and the lack of three-dimensional depth features, which adds to computational and imaging complexity.

The tasks OPEN and SPREAD showed high mean accuracy of different approaches, suggesting that these tasks might be better at differentiating between ALS and HC. The variation in performance across tasks indicates that certain facial movements might be more indicative of ALS-related changes, and that the classifier’s performance is task-dependent. This warrants further investigation with larger datasets to validate these findings and explore the underlying reasons for these differences.

It is important to note that the model used for action unit extraction, Py-feat, was not trained or fine-tuned on a dataset containing faces of people with neurological diseases. As a result, the prediction of the action units is only estimated. This limitation highlights the need for models trained on more diverse datasets, including those with neurological conditions, to improve accuracy.

Finally, there is a need to develop remote smartphone-based video assessments for ALS. Abbas et al. (ABBAS *et al.*, 2021) have shown that smartphone-based video assessments can provide an objective evaluation of motor abnormalities, demonstrating success in conditions like schizophrenia. The potential of using this technology for remote monitoring could make it

suitable for patients living in remote or underserved areas where access to specialized ALS care is limited.

### 3.2.3.3 HC vs PS

This study has examined the potential of using AUs obtained from facial videos to identify PS patients. I utilized the Toronto NeuroFace Dataset (BANDINI *et al.*, 2021), which comprises facial videos of PS patients and a healthy control group. The videos contain faces of participants performing nine predefined actions. The binary classification results show that, based only on the videos, the AI model differentiated between HC and PS with the best accuracy of 0.82.

I cannot directly compare this work to either Kaewmahanin *et al.* (KAEWMAHANIN *et al.*, 2022) or Parra-Dominguez *et al.* (PARRA-DOMINGUEZ; SANCHEZ-YANEZ; GARCIA-CAPULIN, 2021) because they extracted peaks of images and conducted 10-fold cross-validation, which might have resulted in data leakage between folds. Table 26 compares the performance of Bandini *et al.* (BANDINI *et al.*, 2018b) and my work in classifying HC versus PS. My approach is very similar to that of Bandini *et al.* (BANDINI *et al.*, 2018b)

Table 26 – Comparison with prior works.

Task	Approach	Accuracy
BBP	Ours	73.0%
	Bandini <i>et al.</i> (BANDINI <i>et al.</i> , 2018b)	87.0%
PATAKA	Ours	73.0%
	Bandini <i>et al.</i> (BANDINI <i>et al.</i> , 2018b)	73.9%
PA	Ours	77.0%
	Bandini <i>et al.</i> (BANDINI <i>et al.</i> , 2018b)	73.9%
KISS	Ours	82.0%
	Bandini <i>et al.</i> (BANDINI <i>et al.</i> , 2018b)	60.9%
OPEN	Ours	77.0%
	Bandini <i>et al.</i> (BANDINI <i>et al.</i> , 2018b)	78.3%
SPREAD	Ours	82.0%
	Bandini <i>et al.</i> (BANDINI <i>et al.</i> , 2018b)	82.6%
BLOW	Ours	73.0%
	Bandini <i>et al.</i> (BANDINI <i>et al.</i> , 2018b)	82.6%

Bandini *et al.* (BANDINI *et al.*, 2018b) used the same dataset but analyzed infrared along with RGB images, requiring special cameras not suitable for general-purpose smartphone-based imaging. Additionally, their work required manual segmentation of the videos and calibration of the distance between the camera and the face. My method does not require infrared imaging or manual detection of the start and end of video frames. Thus, my software achieves similar performance to the state-of-the-art reported by Bandini *et al.* (BANDINI *et al.*, 2018b), but without manual intervention and using only a single camera. This makes it suitable for use as an app on a smartphone. I have proposed Software as a Medical Device (SaMD) that could be used by first responders, with the front-end interface shown in Figure 13.

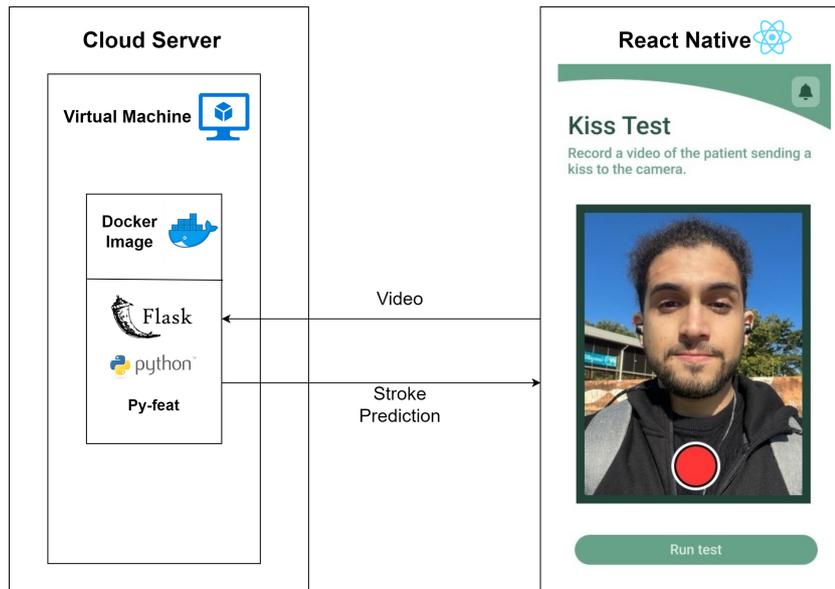


Figure 13 – System architecture.

The major advantage of using AUs for this application is that it provides an assessment based on clinically observable changes. Earlier studies have identified that asymmetry of the mouth is the most common change to facial expressions, while the eyebrow is least affected. My findings validate the significance of facial muscle activation around the mouth in relation to AU, and also emphasize the importance of the eyebrow muscles in this context. Thus, one outcome of this work is the potential to develop a protocol for evaluating a patient suspected of having suffered a stroke.

The proposed, yet untested, app interface and the model are shown in Figure 13. The front end has three stages: recording, transmission, and reporting. The video is recorded in the first stage; the data is uploaded automatically to a cloud-based server in the second stage, where the AI model analyzes it. In the final step, the report is displayed on the phone screen with three display options: (i) PS, (ii) Not PS, and (iii) Unable to classify. The app emphasizes simplicity of use with minimal interface buttons.

The tool is designed to be a supplementary aid for early screening by paramedics, enabling quicker preliminary assessments in emergencies. The loss of facial expressions is not unique to post-stroke, and there can be different etiologies leading to lower specificity. However, an accuracy of 82% and sensitivity of 91% are acceptable within the context of its supplementary role. The tool is not a replacement for comprehensive clinical diagnostics but serves as a first-level response.

### 3.2.4 Conclusion

This work presented a preliminary investigation into the potential of using AUs obtained from facial videos to identify facial weakness symptoms in patients with ALS and PS conditions.

The studies demonstrate that computerized analysis of videos of individuals performing facial expression tasks is a promising approach to assist clinicians in detecting symptoms associated with these neurological conditions.

Both studies were conducted on single datasets, necessitating additional research before the findings can be widely applied in clinical settings. Further investigations should focus on examining the impact of factors such as ethnicity, age, and gender to optimize these methods for more diverse patient populations, employing larger datasets to address these gaps. Moreover, future work should explore combining action units with kinematic analysis features or other techniques for extracting and correlating clinically significant information from facial action units and specific clinical symptoms, which may improve accuracy.

In the ALS classification, the model was cross-validated to identify the most suitable facial expressions, corresponding AUs, and statistical measures, achieving the highest accuracy of 0.91 when the "SPREAD" facial expression task was performed. The experiments were conducted on the Toronto Neuroface Dataset (BANDINI *et al.*, 2021), which comprises facial videos of ALS patients and healthy control group participants performing nine predefined facial expression.

In the post-stroke classification, an accuracy of 0.82 was achieved. The analysis revealed that the "Kiss" and "Spread" facial expressions are the most effective in differentiating between healthy and post-stroke individuals, indicating that muscles around the mouth are most affected. The system is designed for analysis of the videos, making it suitable for development as an app on a smartphone that first responders could potentially use.

### **3.2.5 CRediT authorship contribution statement**

I am grateful for the support and collaboration that made this work possible. As the lead investigator, I was responsible for the research, design, methodology development, and writing this study. I thank Quoc C. Ngo for project administration, validation, and manuscript review; Leandro A. Passos for software validation and manuscript review; Leonardo S. Oliveira for review analysis; João P. Papa for project oversight; and Dinesh Kumar for his conceptual guidance, supervision and manuscript review. Their contributions were essential to the success of this work.

## **3.3 Graph Neural Networks for Facial Expression Analysis**

Neurological disorders, although varied in their underlying causes, often share a common and debilitating consequence: the deterioration of orofacial functions. Stroke—a condition caused by an interruption in blood flow due to vessel blockage or rupture—results in insufficient delivery of oxygen and nutrients, leading to motor deficits, muscle atrophy, and facial paralysis (SCHIMMEL *et al.*, 2017; LAPCHAK; ZHANG, 2017; MULLEN; LOOMIS *et al.*,

2014). Similarly, Amyotrophic Lateral Sclerosis (ALS), a progressive neurodegenerative disorder with an average survival of two to four years after diagnosis (XU; YUAN, 2021; ARTHUR *et al.*, 2016), affects speech, swallowing, and emotional expression (YOLCU *et al.*, 2019). Both conditions substantially impact daily activities and social communication, thereby necessitating precise diagnostic methods and effective rehabilitation strategies.

Traditional assessment methods rely on handcrafted features extracted from facial images or videos—such as measures of facial symmetry, motion, action units, and shape—to detect impairments (WU; JI, 2019). Although these methods have demonstrated promise, they depend heavily on manual design and expert insights, which may limit their adaptability in dynamic clinical scenarios. Recent advancements in automated image and video analysis have created opportunities to learn discriminative features directly from the data.

This study introduces a novel framework that employs Graph Neural Networks (GNNs) to detect orofacial impairments in both stroke and ALS patients, developing a new framework "Facial Point Graphs". By mapping key facial landmarks into a non-Euclidean space, the approach captures intrinsic geometric relationships and dynamic changes, thereby enhancing feature extraction precision without the need for handcrafted features.

The main contributions of this work are:

1. The application of facial graph representations to assess impairments arising from stroke and ALS.
2. The development of a deep learning framework that automatically learns structural and motion features from facial expressions.

### 3.3.1 Methods

The methodology involves constructing Facial Point Graphs from video data to capture the geometric information of facial expressions. The process includes data preprocessing, feature extraction, graph construction, and classification using GNNs.

#### 3.3.1.1 Dataset

This study utilizes two distinct subsets derived from the Toronto NeuroFace dataset (BANDINI *et al.*, 2020). One subset targets orofacial assessments for identifying impairments in patients with ALS compared to healthy controls, while the other focuses on post-stroke conditions by comparing individuals who have experienced stroke with healthy controls. Each subset consists of videos recorded during speech and non-speech tasks that are commonly employed in clinical assessments of neuromuscular impairments.

*ALS vs. Healthy Controls Subset*

The ALS subset consists of 261 colored (RGB) videos from 36 participants, including 15 individuals diagnosed with ALS and 21 healthy controls. Each participant performed multiple repetitions of various prescribed subtasks routinely recommended in ALS clinical evaluations. After manually cropping the videos to isolate each participant’s task, a total of 921 video repetitions were obtained. Table 27 summarizes the number of repetitions for each subtask.

Table 27 – Number of repetitions for each subtask in the ALS vs. healthy controls subset

<b>Subtask</b>	<b>Description</b>	<b>ALS</b>	<b>Healthy Controls</b>
SPREAD	Spread the lips as if smiling	55	59
KISS	Pretend to kiss a baby	59	57
OPEN	Maximum opening of the jaw	54	55
BLOW	Pretend to blow a candle	31	39
BBP	Repetitions of the sentence “Buy Bobby a Puppy”	95	111
PA	Repetitions of the syllable /pa/ as fast as possible	100	110
PATAKA	Repetitions of the syllables /pataka/ as fast as possible	88	108

*Post-Stroke vs. Healthy Controls Subset*

The post-stroke subset contains orofacial videos of 22 participants (11 post-stroke patients and 11 healthy controls). Each participant performed multiple repetitions of speech and non-speech tasks that are recommended for post-stroke evaluations. After manually cropping the videos for each task, 2,509 valid video repetitions were obtained for analysis. Table 28 summarizes the repetitions for each subtask.

Table 28 – Number of repetitions for each subtask in the post-stroke vs. healthy controls subset

<b>Task Type</b>	<b>Subtask</b>	<b>Healthy Controls</b>	<b>Post-Stroke</b>
Speech	BBP (“Buy Bobby a Puppy”)	111	104
Speech	PA (/pa/ as fast as possible)	884	533
Speech	PATAKA (/pataka/ as fast as possible)	275	163
Non-speech	SPREAD (Pretending to smile with tight lips)	59	62
Non-speech	KISS (Pretend to kiss a baby)	57	62
Non-speech	OPEN (Maximum opening of the jaw)	55	61
Non-speech	BLOW (Pretend to blow a candle)	39	44

These distinct subsets allow separate analyses for patients with ALS and those with post-stroke impairments against healthy controls, providing a robust framework for developing and validating automated diagnostic methods.

### 3.3.1.2 Preprocessing

A unified preprocessing pipeline is implemented for both subsets. Initially, OpenFace 2.0 (BALTRUŠAITIS; ROBINSON; MORENCY, 2016) is used to automatically detect the primary face in each video. Leveraging its head pose estimation capabilities, the tool normalizes and crops the region of interest to produce consistent  $200 \times 200$  pixel images, thereby eliminating background elements. To emphasize the geometric and kinematic features essential for diagnosis, each cropped frame is converted to grayscale.

### 3.3.1.3 Feature Extraction

Facial landmarks are detected using the Facial Alignment Network (FAN) (BULAT; TZIMIROPOULOS, 2017), which identifies 68 landmarks following the MULTI-PIE 2D format. Because the dataset primarily presents frontal face images, landmark detection is performed in two dimensions.

For the ALS vs. healthy controls subset, the approach focuses on regions around the lips and jaw—where subtle movement changes may indicate early ALS. To achieve this focus, 26 specific landmarks are selected from the full set, as illustrated in Figure 14.

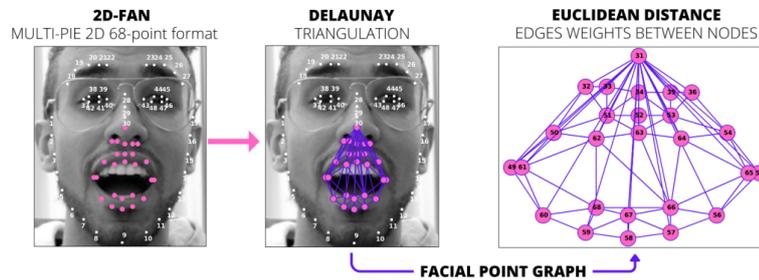


Figure 14 – Preprocessing and feature extraction for the ALS vs. healthy controls subset.

For the post-stroke vs. healthy controls subset, the goal is to capture facial symmetry and any imbalances caused by paralysis following stroke. In this configuration, 48 landmarks are used after excluding 20 that are less relevant for assessing symmetry. These landmarks span areas such as the forehead, eyes, and mouth to provide a comprehensive view of facial structure, as shown in Figure 15.

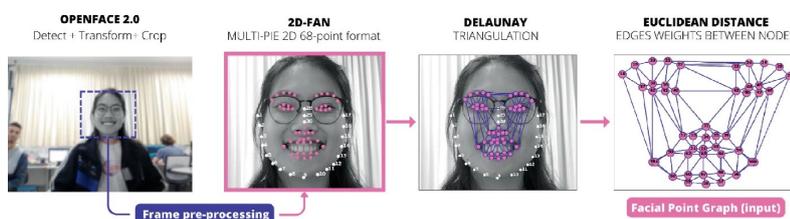


Figure 15 – Preprocessing and feature extraction for the post-stroke vs. healthy controls subset.

Delaunay triangulation (BORIS, 1934) is applied to connect the selected landmarks, forming a mesh that represents their spatial relationships. A central node corresponding to the nose tip (point 31) is manually connected to all other nodes to maintain a consistent spatial structure. The Euclidean distance between connected landmarks is computed and used as the weight for the corresponding edge.

The resulting landmark coordinates and spatial relationships form the Facial Point Graphs (FPGs) that serve as input for subsequent analysis.

#### 3.3.1.4 Proposed Model

The proposed model processes video repetitions to classify subjects based on their facial movements, as depicted in Figure 16. For each repetition, 15 equally spaced frames are extracted and converted into facial graphs.

##### *ALS vs. HC Configuration:*

Each frame becomes a graph with 26 nodes representing key landmarks in the lower facial region. Each node encodes a feature vector (the x and y coordinates of the landmark), and graph edges are weighted by the Euclidean distance between connected landmarks. The graph is processed through six Graph Attention Network (GAT) layers followed by two fully connected layers. An average pooling operation aggregates the node features into a single vector per frame. Frame-level predictions are combined using a majority vote to obtain a repetition-level decision, and subject-level classification is similarly determined via majority voting over all video repetitions.

##### *PS vs. HC Configuration:*

Here, each frame is converted into a graph with 48 nodes to capture a broader range of facial landmarks important for detecting asymmetry. The processing pipeline is similar, except that the pooling operation is implemented using max pooling, which better emphasizes the pronounced asymmetries typically observed following a stroke. As in the ALS configuration, frame-level outputs are combined through majority voting to obtain repetition-level and, subsequently, subject-level classifications.

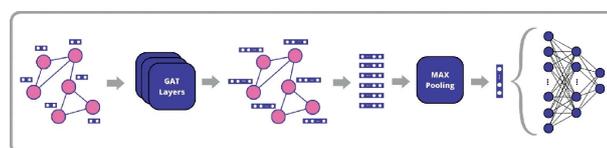


Figure 16 – Overview of the Facial Point Graph model approach.

### 3.3.1.5 Classification

A leave-one-subject-out cross-validation (LOSO-CV) strategy is adopted to evaluate the model's performance in discriminating between impaired patients and healthy controls. In each iteration, all video repetitions of one participant are reserved for testing while the remaining data is split into training and validation sets (with one subject randomly selected from each group for validation).

Two evaluation methods are employed:

**Repetition-Based Classification:** Each video repetition is treated as an independent sample. For each frame within a repetition, a facial graph is generated and processed through the network. A pooling operation aggregates the frame-level features into a single vector that represents the entire repetition. Performance metrics, including accuracy, sensitivity, and specificity, are computed based on these repetition-level predictions.

**Subject-Based Classification:** Subject-level classification is determined through majority voting over all repetitions associated with a subject, thereby mitigating the impact of occasional misclassifications at the repetition level.

Both evaluation approaches are applied independently to the ALS and post-stroke subsets, ensuring that the reported metrics reflect the model's generalization on unseen subjects.

## 3.3.2 Results

The model was evaluated separately on the ALS vs. healthy controls and the post-stroke vs. healthy controls subsets using both repetition-based and subject-based classifications. The performance metrics include accuracy, specificity, and sensitivity for each subtask.

### *Results for the ALS vs. Healthy Controls Subset*

Table 29 presents the results of the Facial Point Graph method for the ALS vs. healthy controls subset. For each task, results are reported for both repetition-level and subject-level classification. Subtasks emphasizing dynamic movements of the lower face (e.g., SPREAD and OPEN) yield higher performance, whereas tasks like KISS and BLOW show lower scores.

### *Results for the Post-Stroke vs. Healthy Controls Subset*

Table 30 shows the performance for the post-stroke subset. Both speech (BBP, PA, and PATAKA) and non-speech (SPREAD, KISS, OPEN, and BLOW) tasks are evaluated using repetition- and subject-based classifications.

Overall, the results indicate that the FPG approach effectively differentiates between impaired patients and healthy controls. Notably, tasks emphasizing lip and jaw dynamics perform

Table 29 – Facial Point Graph results for each subtask in the ALS vs. healthy controls subset

<b>Task</b>	<b>Classification</b>	<b>Accuracy</b>	<b>Specificity</b>	<b>Sensitivity</b>
SPREAD	Repetition	80.7%	79.6%	81.8%
	Subject	81.8%	81.8%	81.8%
KISS	Repetition	68.1%	80.7%	55.9%
	Subject	68.1%	81.8%	54.5%
OPEN	Repetition	77.0%	78.1%	75.9%
	Subject	81.8%	81.8%	81.8%
BLOW	Repetition	37.1%	51.2%	19.3%
	Subject	38.4%	57.1%	16.6%
BBP	Repetition	49.0%	63.0%	32.6%
	Subject	50.0%	63.6%	33.3%
PA	Repetition	64.2%	64.5%	64.0%
	Subject	57.1%	54.5%	60.0%
PATAKA	Repetition	67.3%	65.7%	69.3%
	Subject	66.6%	63.6%	70.0%

Table 30 – Facial Point Graph results for each subtask in the post-stroke vs. healthy controls subset

	<b>Classification</b>	<b>Accuracy</b>	<b>Specificity</b>	<b>Sensitivity</b>
BBP	Repetition	55.3%	58.6%	51.9%
	Subject	59.1%	63.6%	54.5%
PA	Repetition	69.3%	79.1%	53.1%
	Subject	81.8%	90.9%	72.7%
PATAKA	Repetition	45.7%	36.4%	54.5%
	Subject	52.9%	55.9%	50.0%
SPREAD	Repetition	50.0%	54.5%	45.5%
	Subject	52.9%	50.0%	54.5%
KISS	Repetition	64.7%	61.4%	67.7%
	Subject	72.73%	63.64%	81.82%
OPEN	Repetition	58.6%	38.2%	77.0%
	Subject	63.6%	45.5%	81.8%
BLOW	Repetition	65.1%	48.7%	79.5%
	Subject	73.3%	57.1%	87.5%

particularly well in the ALS subset, while repetitive motion tasks yield stable results in the post-stroke subset.

### 3.3.3 Discussion

The performance of the Facial Point Graph approach demonstrates its potential for capturing facial movement patterns associated with neuromuscular impairments. In the ALS vs. healthy controls subset, tasks that emphasize lower face dynamics (e.g., SPREAD and OPEN) achieved higher accuracies, consistent with clinical observations of bulbar involvement in ALS. Conversely, tasks such as KISS and BLOW, which involve more variable movements, exhibited lower performance.

In the post-stroke subset, the model achieved more stable results in tasks that involve repetitive and consistent motion patterns, such as PA and BBP. The results for tasks like KISS and OPEN suggest challenges in capturing subtle asymmetries arising from post-stroke paralysis. However, subject-level classification via majority voting improved overall differentiation.

These findings support the feasibility of using graph-based representations of facial landmarks for the detection of neuromuscular impairments. The differences in performance across subtasks emphasize the importance of carefully selecting and potentially refining tasks. Future work may include integrating temporal dynamics to further enhance classification accuracy and robustness, as well as expanding the dataset to encompass a wider range of facial activities.

#### 3.3.3.1 Novelty of this Study

This study introduces a novel framework that transforms facial expression analysis into a graph-based learning problem by representing facial landmarks as FPGs. Unlike conventional methods that rely on handcrafted features, this approach leverages the spatial and geometric relationships inherent in facial movements, enabling more accurate detection of subtle impairments. The use of GNNs, particularly GATs, facilitates the automatic extraction of both structural and motion-related features directly from the graph representations, enhancing adaptability to diverse clinical presentations.

Two specialized model configurations are designed to address the unique characteristics of ALS and post-stroke conditions. In the ALS setting, the model emphasizes the lower facial region to capture fine-grained dynamics using average pooling, while the post-stroke configuration incorporates a wider set of landmarks and employs max pooling to detect asymmetries due to facial paralysis. The use of leave-one-subject-out cross-validation and majority voting at both repetition and subject levels further ensures robust and generalizable classification performance across heterogeneous datasets.

### **3.3.4 Conclusion**

This study demonstrates that Facial Point Graphs, constructed from strategically selected facial landmarks, can capture both geometric and motion-related features needed to distinguish between patients with ALS or post-stroke impairments and healthy controls. The approach shows promising performance across various facial tasks while eliminating the need for handcrafted features. These results underscore the potential clinical applicability of the method for early detection and monitoring of neuromuscular impairments.

Future research directions include incorporating temporal dynamics, expanding the range of facial tasks, and increasing dataset diversity to improve the model's generalizability across different patient populations.

### **3.3.5 CRediT authorship contribution statement**

I am grateful to Nicolas Barbosa Gomes and Arissa Yoshida for inviting me to be a part of this work. My role included assisting in validating the experimental procedures, and reviewing the manuscript.

---

# VOICE ANALYSIS FOR MONITORING

---

---

This chapter explores AI-driven voice analysis techniques for monitoring Parkinson’s disease progression, focusing on the use of vocal features as indicators of disease severity. It first introduces a methodology based on Diadochokinetic (DDK) tasks, demonstrating how an ensemble approach can identify severity within groups in Parkinson’s disease. Additionally, the chapter examines the apparent vocal tract length in people with Parkinson and the integration of Large Language Models (LLMs) to facilitate AI-based voice assessments, along with the development of a chatbot designed to increase patient engagement and conduct voice assessments using ChatGPT.

Section 4.1 is a work titled “**Speech Assessment for Detecting the Severity of Parkinson’s Disease: An Ensemble Approach**”, which is published in the *Computers in Biology and Medicine* journal.

Section 4.2 presents work in which I contributed as a co-author. My role included assisting in the project design and reviewing the manuscript. This work published as “**The Change of Vocal Tract Length in People with Parkinson’s Disease**”, presented at *45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*.

Section 4.3 has been published as “**NestNeuro: Leveraging Chatbots for Vocal Screening**” in *IEEE 37th International Symposium on Computer Based Medical Systems (CBMS)*.

## 4.1 Diadochokinetic tasks

Parkinson’s disease (PD) is the second most prevalent neurodegenerative disorder, and its prevalence is estimated to be around 0.5%–1% in individuals aged 65–69 years which increases to 1%–3% in those aged 80 and older. People with PD (PwPD) experience a range of both motor and non-motor impairments ([HORNYKIEWICZ, 1998](#)). These include bradykinesia, tremors, difficulties with walking and balance, and speech impairment ([JANKOVIC, 2008b](#)). The

manifestation of the symptoms is complex, and there is no single definitive test for diagnosing PD. The Unified Parkinson's Disease Rating Scale (UPDRS) is a well-accepted metric that combines some of the important motor and non-motor signs and symptoms, thus reducing misdiagnosis (DISEASE, 2003; GOETZ *et al.*, 2008).

PwPD requires regular monitoring of the severity of the disease to ensure appropriate treatment, such as medicine dosage. However, this requires regular and long visits to specialty clinics, which can be challenging, particularly for individuals in remote areas. There is also the potential for bias in the diagnosis because of its subjective nature. Thus, an objective method that can assist the neurologist in assessing their patients remotely will reduce delay in regular monitoring of PwPD.

Literature reveals several computerized assessments of PD motor symptoms with good performance (MEI; DESROSIERS; FRASNELLI, 2021; OLIVEIRA *et al.*, 2023; RANA *et al.*, 2022). While handwriting (ZHAM *et al.*, 2018; ZHAM *et al.*, 2019; ZHAM *et al.*, 2019; ZHAM *et al.*, 2021), use of Inertial Measurement Unit (IMU) (CARAMIA *et al.*, 2018; ALI *et al.*, 2022) and posture analysis (PALMERINI *et al.*, 2011) are accurate, these require specific devices that are generally not available with most patients. Speech has the benefit that it can be recorded and transmitted using a smartphone and is thus suitable for remote assessment of PwPD (MOTIN *et al.*, 2022).

Computerized speech-based differentiating between PwPD and healthy people has been reported to be suitable for detection of the disease (PÉREZ-TORO *et al.*, 2019; SAKAR; SERBES; SAKAR, 2017; MORO-VELAZQUEZ *et al.*, 2019; RUSZ *et al.*, 2021; PAH *et al.*, 2023a; PAH; MOTIN; KUMAR, 2022b; HIREŠ *et al.*, 2022). Many works have reported high performance with accuracy in the range of 69.5-99.6%, sensitivity of 67.4-100.0%, and specificity of 67.2-97.1% (NGO *et al.*, 2022). However, the challenge becomes more complex when addressing the multiclass problem of differentiating among various severity levels of PwPD (BOCKLET *et al.*, 2013; ARIAS-LONDOÑO; GÓMEZ-GARCÍA, 2020; ARIAS-VERGARA *et al.*, 2018; VÁSQUEZ-CORREA *et al.*, 2018a; KODALI; KADIRI; ALKU, 2023).

Multiclass classification, i.e., distinguishing between multiple stages of Parkinson's, such as mild, moderate, and severe, becomes significantly more complex than a simple binary classification of healthy versus PwPD. This complexity might arise from the subtler variations in speech patterns associated with each stage, which often overlap and are less distinct than the differences between healthy individuals and those with any stage of PD. Addressing this will enable better monitoring of PwPD and for developing treatment plans for patient based on the progression of their condition. Therefore, I propose to investigate the severity difference while focusing solely on PwPD to reduce the significant influence of including healthy controls.

This study aimed to develop a computerised method to classify speech based on the disease severity among PwPD without including the healthy control group. The first step was to divide the recordings to four binary classes: (i) Normal stage vs. Not Normal Stage, (ii) Slight

Stage vs. Not Slight Stage, (iii) Mild Stage vs. Not Mild Stage, and (iv) Moderate Stage vs. Not Moderate Stage. The Normal Stage is the label for PwPD who do not show any voice symptoms. This fundamental step converts the multi-class problem to four binary problems. This step is essential because it optimizes the choice of features and classifiers for each problem.

For the next step, a stacked ensemble was employed to develop a multi-class model to label the recordings based on disease severity: Normal, Slight, Mild, and Moderate. The DDK task, feature, and machine learning model for each binary problem identified in step 1 were selected to build the ensemble. The MDS-UPDRS-Speech scale was used for the ground-truth concerning the severity level labels.

The contributions of this study are threefold:

- A systematic examination of the current effectiveness of DDK tasks for determining different PD severity levels.
- To employ an ensemble approach to integrate diverse DDK tasks into a multi-class model, based on the MDS-UPDRS-Speech criteria.
- To establish a computerized voice analysis method to determine the severity level of voice impairment in the Parkinson's group.

### 4.1.1 Methods

Six different diadochokinetic (DDK) tasks, four features (phonation, articulation, prosody, and their fusion), and three classifier architectures (GradientBoosting, Logistic Regression, and RandomForest) to identify the severity level classification of speech for people with PD were investigated. The system consists of two main stages: feature extraction and classification. In the feature extraction phase, four features are considered: phonation, articulation, prosody, and their combination. The need to identify features that capture phonation, articulation, and prosody is supported by prior research indicating that speech difficulties in PD (KODALI; KADIRI; ALKU, 2023; GARCIA *et al.*, 2017; LÓPEZ; OROZCO-ARROYAVE; GOSZTOLYA, 2019; CERNAK *et al.*, 2017).

#### 4.1.1.1 Dataset

This study was conducted using the PC-GITA dataset (OROZCO-ARROYAVE *et al.*, 2014), which comprises speech recordings from PwPD and healthy controls. My analysis focused on the participants from the Parkinson's group. The duration since the diagnosis of PD was in the range from 0 to 43 years, with an average of 11 ( $\pm 10$ ) years. All recordings were made when patients were in their "on" state of medication. The recordings include sustained vowels, isolated sentences, rapid word/syllable repetitions, reading, and monologues.

From the speaking tasks of PC-GITA, the rapid repetition of words (/pa-ta-ka/, /pa-ka-ta/, /pe-ta-ka/) and syllables (/pa/, /ta/, /ka/), which is referred to as the DDK task, were selected and are reported in this investigation. The DDK task is suitable for extracting phonation, articulation, and prosody features. From this task, variations in fundamental frequency, vocal fold vibration regularity, formant characteristics, and prosodic elements become evident, making it versatile and more appropriate for assessing speech quality, articulation proficiency, and prosodic attributes in both clinical and research contexts (GALAZ *et al.*, 2016; RUSZ *et al.*, 2011b; VÁSQUEZ-CORREA *et al.*, 2018b; MORO-VELAZQUEZ *et al.*, 2021). Furthermore, DDK is a routine clinical test familiar to clinicians and has a complex speech structure without being language-dependent. Free speech has limitations because the outcomes are based on the language skills and vision of the patient, while only vowels have a low correlation (PAH *et al.*, 2021).

In this study, I categorized individuals with Parkinson’s disease from the PC-GITA dataset into four different severity levels based on their MDS-UPDRS-Speech scales: 0 for normal, 1 for slight, 2 for mild, and 3 for moderate. Figure 17 displays the distribution of speakers with Parkinson Disease in each severity class.

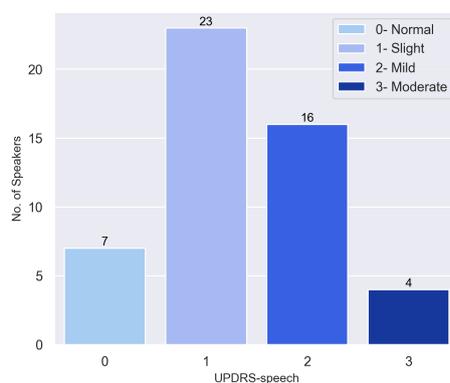


Figure 17 – The distribution of severity of speech symptoms of Parkinson’s disease among the participants in this study: UPDRS-speech average score of 1.34 ( $\pm$  0.82)

#### 4.1.1.2 Feature Extraction

Three distinct feature sets were employed to capture the different facets of speech. Phonation features describe the temporal and amplitude variations of vocal fold vibration extracted from voiced segments. This enables the computation of fundamental frequency, jitter, shimmer, and harmonics-to-noise ratio. Articulation features consider spectral attributes and energy profiles during speech onset and offset and are described by the formants, bandwidths, and spectral moments. Prosody features encapsulate the dynamic contours of fundamental frequency (F0) and energy. All features were extracted using the DisVoice framework<sup>1</sup>.

<sup>1</sup> <<https://github.com/jcvasquezc/DisVoice>>

1. **Phonation:** Analyzing phonation in continuous speech requires identifying voiced segments within the spoken expression (ARIAS-VERGARA; VÁSQUEZ-CORREA; OROZCO-ARROYAVE, 2017; VÁSQUEZ-CORREA *et al.*, 2018b). Seven phonation parameters were computed for the voiced segments: jitter and shimmer, the first and second derivatives of (F0), long-term perturbation features- amplitude perturbation quotient and pitch perturbation quotient, and energy. The mean, standard deviation, skewness, and kurtosis were calculated for all the seven parameters, resulting in a 28-dimensional feature vector for each utterance.
2. **Articulation:** Evaluating the articulatory capabilities of a person requires analyzing how they initiate and conclude their vocal movements. (VÁSQUEZ-CORREA *et al.*, 2018b; OROZCO-ARROYAVE *et al.*, 2018) For this analysis, a set of 12 Mel-Frequency Cepstral Coefficients (MFCCs) and their first and second derivatives from the start and end of vocal sounds were extracted. It also requires computing the log energy of the signal across 22 Bark Bands. I used the first and second formant frequencies and their derivatives to assess patients' speech articulation. In total, 87 descriptors were obtained to describe these characteristics. Additionally, four statistical functions (mean, standard deviation, skewness, and kurtosis) were calculated for each descriptor, resulting in a 488-dimensional feature vector.
3. **Prosody:** The prosodic comprises three aspects: the length of speech, the pitch variation (F0 contour), and the energy modulation (DEHAK; DUMOUCHEL; KENNY, 2007). I calculated 13 attributes for each spoken phrase, encompassing metrics such as the mean, variation, and peak pitch value (F0). I determined the pitch variability in semitones and metrics for the energy contour, mean, variation, and peak values. I also analyzed voice usage rate, average and variation in the duration of voiced segments, pause frequency, and the mean and deviation in pause duration. These measurements collectively produce a 103-dimensional prosodic feature vector for each spoken utterance.
4. **Fusion:** In the previous sub-sections, I explored three features: phonation, articulation, and prosody. The fusion feature is obtained by concatenating these three sets of features, creating a unified 619-dimensional feature vector for each utterance.

#### 4.1.1.3 Classification

To determine a suitable classifier, three ML-based methods were used: Random Forest (RF) (CUTLER; CUTLER; STEVENS, 2012), Logistic Regression (LR) (DREISEITL; OHNO-MACHADO, 2002), and GradientBoosting (GB) (NATEKIN; KNOLL, 2013). I considered the 6 tasks (subsection 2.1), 4 feature sets (subsection 2.2) and the 3 classifiers, a total of 72 binary classifications. These were repeated for the four disease stages for PwPD: Normal, Slight, Mild, and Moderate. Four binary classes were made for each DDK task recording: (i) Normal stage vs.

Not Normal Stage, (ii) Slight Stage vs. Not Slight Stage, (iii) Mild Stage vs. Not Mild Stage) and (iv) Moderate-Stage vs. Not Moderate Stage.

A pipeline was set for each machine learning method, each with specific hyperparameters. These were implemented using sci-kit-learn’s Pipeline class (PEDREGOSA *et al.*, 2011b), combining feature selection, data preprocessing, and a classifier to create a unified workflow. All pipelines employed recursive feature elimination (RFE) for feature selection to select two features and *StandardScaler* for data standardization. RFE is a method where an estimator recursively assesses and eliminates the least essential features from a dataset until a predefined number of critical features is retained. I used a *DecisionTreeClassifier*, with a maximum depth of 5, to manage non-linear relationships between features.

The GB pipeline incorporates a *GradientBoostingClassifier* with specific hyperparameters, including a learning rate of 0.01 and 300 estimators. Similarly, the LR pipeline used *LogisticRegression* with hyperparameters such as the ‘*liblinear*’ solver, L2 penalty, a regularization parameter *C* of 10, and balanced class weights. The RF pipeline was performed by a *RandomForestClassifier* with specific hyperparameters, including a maximum depth of 3, 50 estimators, and balanced class weights.

This study performed no optimization step. The number of selected features was kept constant at two, along with previously determined hyperparameters. The rationale for using simple features and regularized models was the need to work with a minimal and highly unbalanced dataset for each disease stage.



Figure 18 – Example of a Logistic Regression pipeline fitted to “ta-ta-ta”: phonation features are extracted to classify Normal stage vs. not Normal stage.

To assess the performance of a model, a leave-one-out cross-validation approach was used. Each data point is held out as a test sample while the pipeline (Fig.2) is trained on the remaining data. The selection and classification occur only within the training fold, ensuring a robust evaluation without data leakage. The Matthews Correlation Coefficient (MCC) was employed to identify the top four pipelines for evaluating the performance of the model. MCC is particularly effective for datasets with imbalanced classes, making it an ideal choice for this analysis.

Subsequently, I assessed the performance of these four pipelines by calculating three measures: accuracy, sensitivity, and specificity. To construct the confidence intervals for each metric, the standard error was multiplied by the critical z-value associated with the desired confidence level, typically 1.96 for a 95% confidence interval:

$$CI = p \pm z \cdot \sqrt{\frac{p \cdot (1 - p)}{n}},$$

where  $p$  denotes the estimated proportion (i.e., accuracy, sensitivity, or specificity) and  $n$  is the number of observations relevant to each measure (total sample size for accuracy, number of actual positives for sensitivity, and number of actual negatives for specificity).

I conducted the Mann-Whitney U Test to explore the feature selection step's effectiveness further. The rationale is to evaluate the differences between the features selected by RFE statistically. Cohen's  $d$  was also calculated to quantify the effect size of these differences.

#### 4.1.1.4 Ensemble Modeling

In this study, I have introduced a stack ensemble model to overcome the multi-class classification reported in literature (ARIAS-LONDOÑO; GÓMEZ-GARCÍA, 2020; ARIAS-VERGARA *et al.*, 2018; KODALI; KADIRI; ALKU, 2023). This approach also allows for more exact speech tasks, speech features, and classifiers based on the disease's severity level.

The best of each of the four binary problems described above were identified, and the DDK task, feature, and pipeline that gave the best-balanced accuracy were selected. After this selection process, a stack ensemble model was constructed to predict Normal vs. Slight vs. Mild vs. Moderate.

My ensemble model is novel because it harnesses the strengths of diverse base models, each having most suitable feature sets and trained for specific classification tasks related to dysarthria stages. This methodology allows for a more granular and accurate representation of the progression of speech difficulties in Parkinson's patients. The innovation lies in how the stacked ensemble combines these different models, improving overall accuracy and addressing the heterogeneity of speech impairments in Parkinson's disease.

This approach was designed to adapt to the subtleties of speech variations across different stages of Parkinson's by identifying suitable features for each condition. This is often overlooked in existing models. Consequently, my model demonstrates an improved performance, making it a tool in the nuanced detection and analysis of Parkinson's-related dysarthria.

### 4.1.2 Results

The results of the three machine learning methods, i.e., Gradient Boosting (GB), Logistic Regression (LR), and Random Forest (RF), in classifying the four stages of PD, i.e., Normal, Slight, Mild, and Moderate-Stage, are presented. Table 31 presents the Normal vs. Not-Normal Stage classification results. The table shows that the LR model achieved the highest MCC for the 'ka-ka-ka' prosody pattern with a score of 0.62, suggesting that LR is a suitable choice for Normal-stage classification.

The classification results for the Slight Stage vs. Not Slight Stage are shown in Table 32. The RF model stands out with the highest MCC in this case. Notably, 'Articulation' showed a

performance with an MCC of 0.56 with ‘petaka’, making RF a promising choice for classifying Slight-stage speech disorders.

Task	Feature	GB	LR	RF
ka-ka-ka	Phonation	-0.19	-0.33	-0.21
	Articulation	-0.13	0.06	-0.15
	<b>Prosody</b>	0.44	<b>0.62</b>	0.56
	Fusion	-0.15	0.42	0.23
pa-pa-pa	Phonation	0.56	-0.05	0.31
	Articulation	0.00	-0.08	-0.02
	Prosody	-0.12	-0.14	-0.02
	Fusion	0.00	-0.06	0.00
pakata	Phonation	0.11	0.11	0.11
	Articulation	-0.12	-0.04	0.03
	Prosody	-0.12	0.01	-0.04
	Fusion	-0.12	-0.04	0.03
pataka	Phonation	-0.18	-0.14	-0.15
	Articulation	0.06	0.34	0.06
	Prosody	0.09	0.11	0.09
	Fusion	0.17	0.31	0.03
petaka	Phonation	-0.16	-0.18	-0.18
	Articulation	-0.15	-0.29	-0.18
	Prosody	-0.13	-0.01	0.04
	Fusion	-0.16	-0.34	-0.23
ta-ta-ta	Phonation	-0.15	0.18	-0.15
	Articulation	-0.09	0.04	-0.03
	Prosody	-0.10	-0.19	-0.15
	Fusion	-0.18	-0.25	-0.18

Table 31 – Normal vs. Not Normal

Task	Feature	GB	LR	RF
ka-ka-ka	Phonation	-0.10	0.04	0.02
	Articulation	-0.36	-0.03	-0.44
	Prosody	0.10	0.04	-0.15
	Fusion	-0.45	0.01	-0.68
pa-pa-pa	Phonation	-0.08	-0.04	-0.05
	Articulation	0.23	0.23	0.19
	Prosody	-0.14	0.00	-0.05
	Fusion	0.35	0.36	0.35
pakata	Phonation	-0.08	-0.19	-0.19
	Articulation	0.05	-0.08	-0.10
	Prosody	-0.11	-0.10	-0.15
	Fusion	-0.11	-0.08	-0.07
pataka	Phonation	-0.08	-0.03	-0.03
	Articulation	-0.32	-0.19	-0.52
	Prosody	0.27	0.25	0.23
	Fusion	0.07	0.27	0.15
<b>petaka</b>	Phonation	-0.21	-0.36	-0.15
	<b>Articulation</b>	0.48	0.36	<b>0.56</b>
	Prosody	-0.07	-0.08	-0.10
	Fusion	0.48	0.36	0.56
ta-ta-ta	Phonation	-0.02	-0.13	0.05
	Articulation	-0.01	0.13	-0.16
	Prosody	-0.00	0.07	-0.01
	Fusion	-0.10	0.05	-0.17

Table 32 – Slight vs. Not Slight

Task	Feature	GB	LR	RF
ka-ka-ka	Phonation	0.02	-0.23	0.06
	Articulation	0.40	0.03	0.40
	Prosody	-0.08	-0.21	-0.05
	<b>Fusion</b>	0.40	0.06	<b>0.43</b>
pa-pa-pa	Phonation	0.32	0.11	0.14
	Articulation	-0.17	-0.02	-0.17
	Prosody	0.09	0.17	0.39
	Fusion	-0.13	-0.05	-0.28
pakata	Phonation	-0.26	-0.15	-0.25
	Articulation	-0.07	0.14	-0.01
	Prosody	0.05	0.05	-0.04
	Fusion	-0.04	0.08	-0.04
pataka	Phonation	0.05	0.26	0.23
	Articulation	-0.11	-0.27	-0.19
	Prosody	0.15	0.08	0.20
	Fusion	-0.14	0.12	0.08
petaka	Phonation	-0.18	-0.18	-0.15
	Articulation	0.02	-0.20	-0.18
	Prosody	-0.24	-0.21	-0.29
	Fusion	-0.26	-0.14	-0.24
ta-ta-ta	Phonation	-0.24	0.02	-0.16
	Articulation	0.12	0.20	0.08
	Prosody	0.17	0.20	0.17
	Fusion	-0.29	-0.10	-0.26

Table 33 – Mild vs. Not Mild

Task	Feature	GB	LR	RF
ka-ka-ka	Phonation	-0.07	0.41	-0.06
	Articulation	-0.07	-0.09	-0.07
	Prosody	0.12	0.20	0.12
	Fusion	-0.07	-0.10	-0.07
pa-pa-pa	Phonation	-0.06	-0.23	-0.06
	Articulation	-0.09	0.29	-0.09
	Prosody	-0.11	0.14	-0.10
	Fusion	-0.09	0.13	-0.09
pakata	Phonation	-0.09	0.38	0.18
	Articulation	-0.04	0.09	-0.04
	Prosody	0.18	0.01	-0.07
	Fusion	0.48	0.05	0.32
pataka	Phonation	-0.09	0.11	-0.07
	Articulation	-0.06	0.64	0.24
	Prosody	0.15	0.22	0.18
	Fusion	0.18	0.27	0.24
petaka	Phonation	-0.07	-0.16	0.00
	Articulation	0.48	0.31	-0.06
	Prosody	0.32	0.18	-0.06
	Fusion	0.48	0.31	0.32
<b>ta-ta-ta</b>	Phonation	-0.06	-0.15	-0.04
	Articulation	0.15	-0.01	0.12
	<b>Prosody</b>	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>
	<b>Fusion</b>	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>

Table 34 – Moderate vs. Not Moderate

Table 33 displays the classification results for the Mild Stage vs. Not Mild Stage. The highest MCC was 0.43 in the ‘ka-ka-ka’ task for the RF pipeline using fusion features. The highest MCC for the ‘Prosody’ feature in ‘pa-pa-pa’ was 0.39 using the RF pipeline.

The results for the Moderate-Stage vs. Not Moderate-Stage classification are presented in Table 34. While the performance appears to be perfect, with a MCC of 1.00, the dataset of this class was very small; hence, this may not be an accurate indicator. The ‘Prosody’ and ‘Fusion’ features stand, with the GB, LR and RF models achieving a perfectly MCC of 1.00.

Table 5 presents the accuracy, sensitivity and specificity of the best MCC found before. For the Normal vs. Not Normal classification, using the prosody and a LR model on the “ka-ka-ka” speech task, the results showed an accuracy of  $0.88\pm(0.090)$ . For the Slight vs. Not Slight classification on the “petaka” task with articulation features and a RF model, the accuracy was  $0.78\pm(0.115)$ . The Mild vs. Not Mild classification using a fusion of features on the “ka-ka-ka” task with a RF model reported an accuracy of  $0.72\pm(0.124)$ . Notably, the Moderate vs. Not Moderate classification on the “ta-ta-ta” task, which prosody/fusion features in a GB model, achieved a perfect accuracy of  $1.0\pm(0.00)$ . For all classifications, both sensitivity and specificity were above 70%.

Classification	Task	Feature	Model	Accuracy	Sensitivity	Specificity	MCC
Normal vs Not Normal	ka-ka-ka	prosody	LR	$0.88\pm(0.090)$	$0.86\pm(0.257)$	$0.88\pm(0.097)$	0.62
Slight vs Not Slight	petaka	articulation	RF	$0.78\pm(0.115)$	$0.74\pm(0.179)$	$0.81\pm(0.148)$	0.56
Mild vs Not Mild	ka-ka-ka	fusion	RF	$0.72\pm(0.124)$	$0.75\pm(0.212)$	$0.71\pm(0.153)$	0.43
Moderate vs Not Moderate	ta-ta-ta	prosody/fusion	GB	$1.0\pm(0.00)$	$1.0\pm(0.00)$	$1.0\pm(0.00)$	1.0

Table 35 – Classification Report.

The statistical analysis, presented in Table 36, shows the significance of the different descriptors in classifying different stages of a medical condition. For the Normal vs Not Normal classification using the “ka-ka-ka” task, prosody descriptors such as ‘kurtosis dur unvoiced’, representing the peakedness of duration in unvoiced segments, and ‘avg tilt E unvoiced’, representing the average slope of the energy contour across unvoiced segments, were identified as key. These descriptors achieved U-scores of 33 and 56, and p-values of 0.0011 and 0.0086 respectively, with negative Cohen-d values (-1.1900 and -1.1783), indicating significant effect sizes favoring the Normal stage.

In the Slight vs Not Slight classification, during the “petaka” task, articulation features, i.e. ‘skewness MFCCoff\_4’, indicating the asymmetry in the 4th Mel-Frequency Cepstral Coefficient during offset transitions, and ‘avg DMFCC on\_8’, representing the average first derivative of the 8th Mel-Frequency Cepstral Coefficient in onset transitions, proved crucial. These features displayed a U-score of 475 and a p-value of 0.0014 for skewness, and a U-score of 175 with a p-value of 0.0085 for average delta, with respective Cohen-d values of 1.2186 and -0.6060, demonstrating the impact on classification.

For the Mild vs Not Mild stage differentiation, using the “ka-ka-ka” task under the fusion category, the suitable features were ‘std DMFCC on\_10’, representing the standard deviation of the 10th first derivative of the Mel-Frequency Cepstral Coefficients during onset transitions, and ‘skewness DMFCCoff\_2’, indicating the skewness of the 2nd first derivative

Classification	Task	Feature	Descriptor	U-score	p-value	Cohen-d
Normal vs Not Normal	ka-ka-ka	prosody	kurtosis dur unvoiced	33	0.0011	-1.1900
			avg tilt E unvoiced	56	0.0086	-1.1783
Slight vs Not Slight	petaka	articulation	skewness MFCCoff_4	475	0.0014	1.2186
			avg DMFCCon_8	175	0.0085	-0.6060
Mild vs Not Mild	ka-ka-ka	fusion	std DMFCCon_10	182	0.0626	-0.3288
			skewness DMFCCoff_2	238	0.4837	-0.2700
Moderate vs Not Moderate	ta-ta-ta	prosody/fusion	max dur unvoiced	184	0.0010	3.0518
			UP	104	0.4663	0.4877

Table 36 – Descriptors selected using RFE.

of the Mel-Frequency Cepstral Coefficients during offset transitions. This class had relatively higher p-values and smaller effect sizes.

Lastly, in the Moderate vs Not Moderate classification using the “ta-ta-ta” task, features combining prosody and fusion like ‘max dur unvoiced’, indicating the maximum duration of unvoiced segments, and ‘UP’ represents the duration ratio of unvoiced segments to pauses, were significant. These features highlight how class based selection of voice descriptors can be used effectively to classify stages of medical conditions. In the future these may also offer insights into the progression of symptoms and aiding in tailored therapeutic approaches.

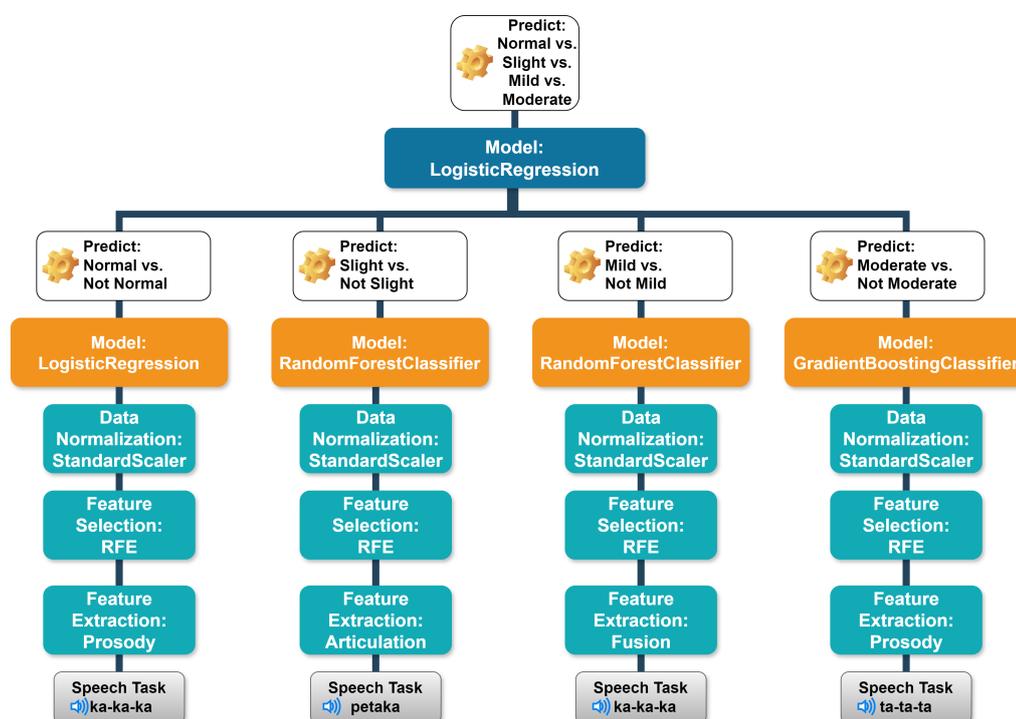


Figure 19 – Inference of the ensemble approach.

The next step was to stack the cross-validated predictions of the best four pipelines and use them as input to train the final step classifier. A logistic Regression model was used for the multi-class problem. Figure 19 illustrates the prediction of my approach. The results are shown in the confusion matrix in 20, with overall accuracy approximately  $0.72 \pm (0.124)$ .

True label	Normal	4 57.14%	1 14.29%	2 28.57%	0 0.00%
	Slight	1 4.35%	18 78.26%	4 17.39%	0 0.00%
	Mild	1 6.25%	5 31.25%	10 62.50%	0 0.00%
	Moderate	0 0.00%	0 0.00%	0 0.00%	4 100.00%
		Normal	Slight	Mild	Moderate
		Predicted label			

Figure 20 – Confusion Matrix of Ensemble approach

### 4.1.3 Discussion

The focus of automatic classification of speech changes associated with PD reported in the literature has primarily focused on the binary differentiation between PwPD and HC. While that is important for assisting in the diagnosis of the disease, it is not suitable for monitoring the progress of the disease in PwPD based on MDS-UPDRS-speech score. However, research for detecting the severity level of PD from speech appears in the early stages (ARIAS-LONDOÑO; GÓMEZ-GARCÍA, 2020; ARIAS-VERGARA *et al.*, 2018; VÁSQUEZ-CORREA *et al.*, 2018a; KARAN; SAHU; OROZCO-ARROYAVE, 2022).

Previous researchers, such as Arias (ARIAS-LONDOÑO; GÓMEZ-GARCÍA, 2020) and Kodali (KODALI; KADIRI; ALKU, 2023), treated the classification of Parkinson's disease severity as a multi-class problem, distinguishing between healthy individuals, mild PD, and severe PD. However, this method may overlook the subtle variations in speech that manifest as the disease progresses. To address this shortcoming, I have adopted a different approach by excluding healthy participants and segmenting the multi-class problem into several binary classification tasks.

The works of Arias (ARIAS-LONDOÑO; GÓMEZ-GARCÍA, 2020) and Kodali (KODALI; KADIRI; ALKU, 2023) have differentiated between healthy, mild PD, and severe PD, and they used speech recordings of 50 healthy and 50 PwPD. Based on the clinical feedback and my previous work where I differentiated between healthy and PwPD (PAH *et al.*, 2021; PAH; MOTIN; KUMAR, 2022b), this study focused on the classification of only PwPD in four classes: Normal, slight, mild, and moderate voice symptoms. It is important to note that "Normal" describes the PwPD who, while diagnosed with PD, does not have speech symptoms.

In the literature, the challenges of small-sized and imbalanced databases have been recognized as limitations for speech-based multi-class clinical diagnosis (DIBAZAR; BERGER; NARAYANAN, 2006; CHUI; LYTRAS; VASANT, 2020; KODRASI *et al.*, 2021). Additionally, using a single standard set of features for all classes may not be optimal. To address these issues, my approach treats each class as a distinct binary problem. This ensures that the speech task, speech features, and classifier are individually optimized for each specific class. Once that is achieved, the outcomes of the four binary classifiers are ensembled to obtain the outcome. This two-stage approach is advantageous as it allows me to select the most suitable tasks and features for each class, contributing to improved performance.

I investigated six DDK tasks and four features, i.e., phonation, articulation, prosody, and their fusion, and three classifier architectures, i.e., GB, LR, and RF, similar to those reported in the literature. However, by performing binary classification for each severity level, I identified the best-performing set of tasks and features for the class, optimizing each system. Classifications were made for each DDK task recording: (i) Normal stage vs. Not Normal Stage, (ii) Slight Stage vs. Not Slight Stage, (iii) Mild Stage vs. Not Mild Stage) and (iv) Moderate-Stage vs. Not Moderate Stage. In the next step, a stacked ensemble was used to develop a multiclass model that could distinguish between Normal, Slight, Mild, and Moderate. This two-stage method resulted in improved performance of the classification.

The findings presented in Table 36 can be investigated to identify the clinical relevance of specific acoustic descriptors in effectively classifying different stages of the medical condition under investigation. These results, obtained through RFE and evaluated using the Mann-Whitney U Test, reveal the importance of the acoustic features in discriminating between various stages of the condition, offering valuable insights for developing diagnostic and monitoring tools. Significance of descriptors such as ‘kurtosis dur unvoiced’, ‘avg tilt E unvoiced’, ‘skewness MFCCoff\_4’, ‘avg DMFCC on\_8’, ‘std DMFCC on\_10’, ‘skewness DMFCCoff\_2’, ‘max dur unvoiced’, and ‘UP’ show the potential utility of these features in clinical practice. These findings can enhance our understanding of the medical condition and facilitate the development of more accurate diagnostic and prognostic tools for improved patient care and management.

In my work, only two descriptors were selected from the feature set based on the recursive feature elimination. This approach was for the simplicity and interpretability of my study, making it easier to examine the relationship between two particular variables. However, reducing the features can lead to the loss of subtle differences. Besides that, the recursive feature elimination method relies on the choice of an estimator to assess feature importance. Thus the results of the feature elimination process can vary based on the choice of the estimator.

The use of DDK has the advantage because it is not language dependent while also has the complexity of speech and includes vowels and consonants. In literature, Khan *et al.* (KHAN *et al.*, 2020) have done similar classification but using speech, which has the limitation because the assessment is dependent on the language skills and vision of the person. The other researchers

such as Sakar et al. (SAKAR; SERBES; SAKAR, 2017) have used vowels for this purpose, which however has been reported to have low correlation (PAH *et al.*, 2021).

There are two novelties of this work. First, I did systematic binary classification to identify the most significant DDK tasks and speech features that are optimum for detecting each severity level of Parkinson's disease. The second novelty is that I have used a new ensemble approach that combines various machine learning models for a complete system to accurately determine the disease's severity level: Normal, Slight, Mild, and Moderate severity levels. This has resulted in a potential computerised system for assisting the clinicians. My approach is promising for addressing challenges posed by imbalanced and small datasets.

#### 4.1.3.1 Novelty of this study

This study introduces a novel, systematic binary classification approach that identifies the most significant DDK tasks, speech/voice features, and classifier architectures for each PD severity level, thereby mitigating the challenges of imbalanced and small datasets. By segmenting the multiclass problem into distinct binary tasks, my method optimizes the system for individually detecting Normal, Slight, Mild, and Moderate stages. It provides scientific insights into the relevance of specific acoustic features—such as alterations in prosody and spectral dynamics—in discerning subtle speech changes. Building on this framework, I further advance the field by integrating different machine learning models into an innovative ensemble approach that unifies these optimized binary classifiers. This integrated methodology enhances classification accuracy and reliability. It lays the groundwork for more precise diagnostic and monitoring tools in clinical settings, thereby contributing to the automated voice analysis for PD.

#### 4.1.3.2 Limitation

While this study has shown the potential of using speech to monitor PD progression, it is not yet ready for deployment. It has four major limitations. Firstly, the dataset used was relatively small and unbalanced; there were only a small number of people with Normal-stage and Moderate-stage Parkinson's disease which limits the generalisability of the model. Secondly, only one dataset was used, and all the recordings in this were collected from one region in Columbia; thus, it lacks demographic and ethnic diversity. Thirdly, the study has not been validated using independently collected data. Fourth, this study is cross-sectional and has not investigated the repeatability nor the ability for this to identify the progression of the disease. Thus, a longitudinal, multi-center study is essential to confirm the efficacy of this model to be suitable for clinical applications.

Unfortunately, I could not validate the model on an external dataset because I couldn't find other public datasets with such details.

#### 4.1.4 Conclusion

This work has demonstrated the potential of using computerized speech analysis to estimate the severity of Parkinson's disease symptoms using four sets of binary classifications. The findings suggest that this approach could potentially support clinicians in remotely monitoring their patients based on speech recordings. However, to enhance the robustness and generalisability of the model, further research is needed to have participants from a diverse range of demographics and ethnic groups. Additionally, longitudinal studies tracking patients throughout the progression of the disease are essential to confirm the model's practical utility.

#### 4.1.5 CRediT authorship contribution statement

I am grateful for the support and collaboration that made this work possible. As the lead investigator, I was responsible for the research, design, methodology development, and writing this study. I thank Quoc C. Ngo for project administration, validation, and manuscript review; Leandro A. Passos for software validation and manuscript review; Nicolás and Arissa for experimental validation; João P. Papa for project oversight; and Dinesh Kumar for his conceptual guidance, supervision and manuscript review. Their contributions were essential to the success of this work.

## 4.2 The Apparent Vocal Tract length

Diagnosis of Parkinson's disease involves a comprehensive evaluation of diverse symptoms, including motor impairments like tremors, rigidity, bradykinesia, and postural instability, as well as non-motor issues such as dysarthria, functional impairment, and cognitive challenges (SIMONET *et al.*, 2019). These assessments are used to calculate a Unified Parkinson's Disease Rating Score (UPDRS). Notably, Parkinsonian hypokinetic dysarthria—reported by 90% of people with PD (RUSZ *et al.*, 2011a)—is characterized by reduced voice intensity, increased nasality and acoustic noise, reduced speech prosody, imprecise articulation, a narrower pitch range, mono loudness, longer pauses, vocal tremor, harsh and breathy voice quality, and disfluency (VAICIUKYNAS *et al.*, 2017; YANG *et al.*, 2020).

Hypokinetic dysarthria in PD is primarily caused by poor activation and coordination of the speech-production muscles (YANG *et al.*, 2020; HUANG *et al.*, 2019). The stiffness and tremor of the laryngeal muscles affect the vibration pattern of the vocal cords, leading to changes in the fundamental frequency and irregular or asymmetrical closure phases during phonation (YANG *et al.*, 2020; SILBERGLEIT *et al.*, 2015). Additionally, reduced controllability of the diaphragm results in unstable phonatory airflow and pneumatic pressure to the larynx, while impaired control of the tongue and lips further complicates speech production (YANG *et al.*, 2020; Jiang J, O'Mara T, Chen HJ, Stern JI, Vlagos D, 1999; HAMMER, 2013).

Many acoustic features from sustained phonemes—such as pitch frequency variation, number of pulses, jitter, shimmer, autocorrelation, MFCCs, and harmonics-to-noise ratio—have been investigated for computerized detection of hypokinetic dysarthria in people with PD (GOYAL; KHANDNOR; ASERI, 2020; SAKAR *et al.*, 2013; SAKAR *et al.*, 2019; BRAGA *et al.*, 2019; MORO-VELÁZQUEZ *et al.*, 2018). The vocal tract length—the distance measured from the glottis to the lips—plays a significant role in the variability of speech production (LAMMERT; NARAYANAN, 2015) and is strongly correlated with formant frequencies (PETERSON; BARNEY, 1952; TURNER *et al.*, 2009). Although various models have been devised to estimate vocal tract length, one common approach is through the apparent vocal tract length (AVTL), calculated from the first four formants  $\{F_i\}_{i=1}^4$  as defined by Pisanski *et al.* in

$$AVTL(F_i) = (2i - 1) \frac{c}{4F_i}, \quad (4.1)$$

where the constant  $c = 33,500$  cm/s represents the speed of sound in a uniform tube with one end closed. Despite its established relevance in speech analysis, the potential relationship between modifications in vocal tract length and dysarthria symptoms in PD has not been thoroughly investigated. To address this gap, this study reports the results of a preliminary investigation that measured the average difference in AVTL between people with PD and age-matched control participants during sustained phonation of the vowel /a/, with implications for improving voice-based diagnostic tools.

### 4.2.1 Methods

The existence and characteristics of change in the length of the vocal tract of people with PD were investigated in this work by observing the change in the AVTL statistical distribution of people with PD in comparison to the age-matched healthy control (HC) participants. The AVTL of the four formants ( $F_1$ ,  $F_2$ ,  $F_3$ , and  $F_4$ ) was extracted from the recording of the sustained phoneme /a/ in two publicly available datasets. The datasets were the UCI Parkinson's Disease Classification Dataset of Okan Sakar (SAKAR *et al.*, 2019) and the Italian Parkinson's Voice and Speech Dataset (DIMAURO *et al.*, 2017), abbreviated in this paper with UCI and ITA, respectively. The datasets were approved by the clinical research ethics committee of Bahcesehir University and Universita' degli Studi di Bari Aldo Moro, respectively. The demographic information of the two datasets is presented in Table 37.

The sustained phoneme /a/ in the UCI dataset was recorded from Turkish participants using a common microphone with a 44.1 kHz sampling rate and 16-bit resolution in an uncontrolled environment. The ITA dataset was recorded from participants of predominantly Italian using a common microphone located about 15 – 25 cm from the subject in a quiet, echo-free room. The recordings were sampled at 16 kHz and 16 bits of resolution.

The AVTLs were calculated by using equation (4.1) from the formant's frequencies. The formants of the UCI dataset were provided in the downloaded tabulated matrix of the extracted

Table 37 – Demographics of the datasets

Features	PD		HC	
	Male	Female	Male	Female
<b>UCI Parkinson's Disease Classification Dataset (UCI)</b>				
# Subject	107	81	23	41
Age (years)	65.1 ± 10.9		61.1 ± 8.9	
<b>Italian Parkinson's Voice and Speech Dataset (ITA)</b>				
# Subject	19	9	10	12
Age (years)	67.2 ± 8.7		67.1 ± 5.2	

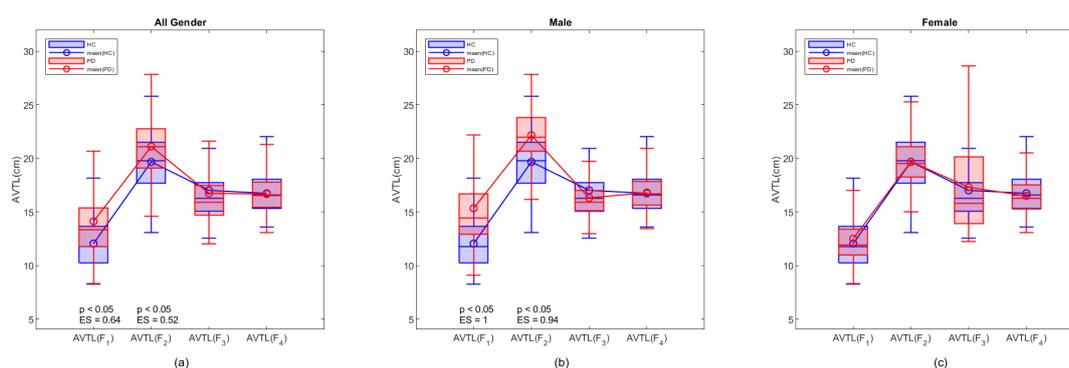


Figure 21 – The statistical distribution of apparent AVTL( $F_1$ ), AVTL( $F_2$ ), AVTL( $F_3$ ), and AVTL( $F_4$ ) or the sustained phoneme /a/ extracted from UCI dataset. (a) The distribution of all participants (b) The distribution of male participants, (c) The distribution of female participants. The effect size (ES) was displayed for the cases with a p-value of less than 0.05 and ES

features in the form of a CSV file. The formants of the WAV recordings in the ITA dataset were calculated using a publicly available speech analysis software, Praat (BOERSMA; HEUVEN, 2001).

Before the statistical analysis, the AVTL data were examined for their normality using the Anderson-Darling test (JÄNTSCHI; BOLBOACĂ, 2018). Due to the non-normality of the AVTL distribution, the Mann-Whitney U-test (MCDONALD, 2014) was used to compare the AVTL of PD and HC groups with a 95% confidence level. A p-value < 0.05 indicated a significant difference between the groups. The effect size, ES (SULLIVAN; FEINN, 2012)(COHEN, 1988), of the cases with p-value < 0.05 were calculated to indicate the strength of the difference. The AVTL calculation and the statistical distribution were calculated using Matlab 2022 of MathWorks.

## 4.2.2 Results

Figures 21 and 22 show the statistical distribution of the four AVTLs, i.e. AVTL( $F_1$ ), AVTL( $F_2$ ), AVTL( $F_3$ ), and AVTL( $F_4$ ) extracted from the phoneme /a/ of the two datasets. The figures show that the AVTL calculated using the first formant  $F_1$  was between 10 – 15 cm while the AVTL( $F_2$ ) was much longer (15 – 25 cm). The AVTL calculated using  $F_3$  and  $F_4$  were having a similar range of around 15 cm.

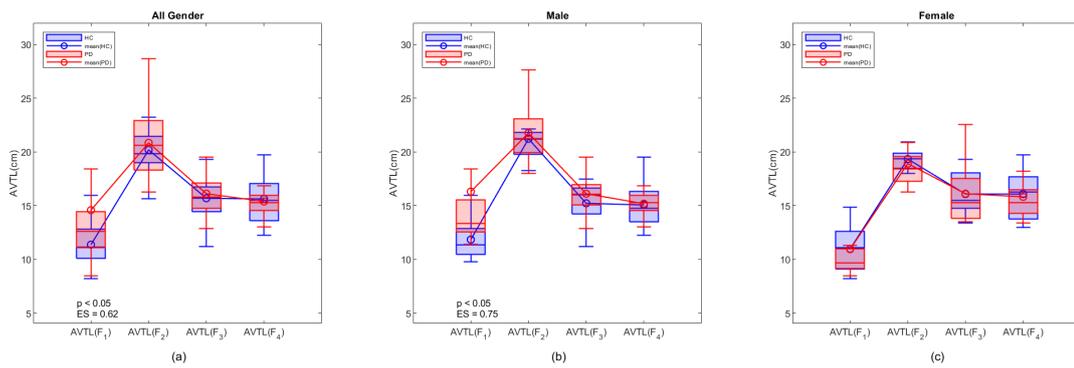


Figure 22 – The statistical distribution of apparent AVTL(F<sub>1</sub>), AVTL(F<sub>2</sub>), AVTL(F<sub>3</sub>), and AVTL(F<sub>4</sub>) or the sustained phoneme /a/ extracted from ITA dataset. (a) The distribution of all participants (b) The distribution of male participants, (c) The distribution of female participants. The effect size (ES) was displayed for the cases with a p-value of less than 0.05 and ES

The figures show the existence of a change in vocal tract length due to PD. The figures show that the AVTL calculated using F<sub>1</sub> was the most effective feature in capturing the difference in vocal tract length between PD and HC participants. People with Parkinson's disease tend to have longer AVTL compared to healthy people when pronouncing the sustained phoneme /a/. The AVTL(F<sub>1</sub>) distribution of UCI and ITA datasets were having a p-value of less than 0.05 with ES of 0.64 and 0.62, respectively. The statistical p-value was calculated using the non-parametric Mann-Whitney U-test (MCDONALD, 2014).

The phenomenon of having a longer vocal tract length in people with Parkinson's was also evidenced in the AVTL(F<sub>2</sub>) distribution but with less consistency and statistical significance. The AVTL(F<sub>2</sub>) of the phoneme /a/ in UCI and ITA datasets of people with PD were longer than that of HC, however, the difference was only significant in UCI dataset. The AVTL of F<sub>3</sub> and F<sub>4</sub> could not identify any change due to PD. The other interesting finding in this work is that the change in the vocal tract length captured by AVTL(F<sub>1</sub>) and AVTL(F<sub>2</sub>) has only appeared in the male participants as shown in the figures. The result indicates that the lengthening of the vocal tract during the pronunciation of /a/ was only happening in male PD patients. The phenomenon was observed in both datasets with different demographics.

### 4.2.3 Discussion

The reduction in the ability of people with PD to control and initiate speech-production muscle activities has been known to be the cause of their speech impairments, known as hypokinetic dysarthria (YANG *et al.*, 2020)(HUANG *et al.*, 2019). This has been investigated based on voice features of vocal cord vibrations such as the pitch frequency variation, number of pulses, jitter, shimmer, the change in the vocal tract harmonics such as formants, autocorrelation, MFCCs, and harmonics-to-noise ratio, and the change in voice intensity (PAH; MOTIN; KUMAR, 2022a). One of the changes in the vocal tract harmonics is related to the change in the length of the vocal tract.

PD has been found to alter the resonance properties of the vocal tract (PAH; INDRAWATI; KUMAR, 2022)(GILLIVAN-MURPHY; CARDING; MILLER, 2016)(PAH; MOTIN; KUMAR, 2022a). This preliminary study identified the change in the vocal tract length of patients with PD through the statistical distribution of the apparent vocal tract length (AVTL) features.

The result of this study indicates that people with PD tend to extend the length of their vocal tract by 2 – 3 cm when performing the pronunciation of the sustained phoneme /a/. The consistent results were observed in two datasets of different demographic settings. The result indicates that due to the reduced ability of the PD patients in controlling or activating their vocal tract muscle, a modification was made to the position of voice production organs in the vocal tract that resulted in the lengthening of the vocal tract. At this point, the authors have not been able to specifically identify the source or the vocal tract change.

Among the four AVTLs investigated in this work, the AVTL derived from the frequency of the first formant, the AVTL( $F_1$ ), was the most suitable feature to represent the difference between the voice of PD and healthy controls. The range of AVTL( $F_1$ ) was between 10 – 15 cm which agrees with the actual anatomical length of the vocal tract as reported in (PISANSKI *et al.*, 2016b). The AVTL( $F_1$ ) could capture the change in the vocal tract length with an effect size (ES) of 0.62 to 1. The AVTL calculated on the second formant,  $F_2$ , was also sensitive to the vocal tract change but with less consistency and accuracy. The third and fourth formants,  $F_3$  and  $F_4$ , were not suitable to represent the length of the vocal tract.

In this study, it was discovered that the alteration in vocal tract length, as measured by AVTL( $F_1$ ) and AVTL( $F_2$ ), only occurred in male participants as depicted in the figures. This suggests that lengthening of the vocal tract during the pronunciation of /a/ is specific to male PD patients. This observation was found in both datasets consisting of different groups of people. The possible explanation of the phenomenon is that, anatomically, a man has a relatively longer vocal tract than a woman (PISANSKI *et al.*, 2016b). However, further works need to be done to investigate the cause of the finding.

#### 4.2.4 Conclusion

This study has investigated the existence and the characteristics of differences in the vocal tract length of people with PD and healthy, age-matched people. It has been found that the AVTL of males with PD is extended while pronouncing the sustained phoneme /a/. The alteration, however, only occurred in male PD patients. This study also found that the most suitable feature to represent the length of the vocal tract is the AVTL( $F_1$ ).

Additional research is required to better understand this phenomenon. This would require a larger sample size, a wider range of sustained sounds, and people separated based on the severity of PD. To investigate the underlying causes of the changes in the vocal tract caused by PD, the sub-glottic pressure and dynamic imaging of the larynx will be required.

### 4.2.5 CRediT authorship contribution statement

I am grateful to Nemuel D. Pah for inviting me to be a part of this work. My role included assisting in project design and manuscript review.

## 4.3 Voice Screening for Chatbot

Neurological disorders, including Parkinson's disease, Alzheimer's, and multiple sclerosis, pose a significant healthcare challenge globally (FEIGIN *et al.*, 2019). According to the World Health Organization, these disorders account for a substantial proportion of the global disease burden. For instance, Parkinson's disease alone affects an estimated 10 million people worldwide (AKBAR *et al.*, 2021). Early and accurate screening for such diseases can lead to more effective management and improved patient outcomes.

However, access to specialized neurological healthcare is often limited, particularly in remote and underserved regions. Moreover, there is a lack of easy-to-use, widely accessible tools that utilize everyday technology like smartphones and messaging apps for preliminary screening. A chatbot that assists in screening neurological diseases represents a significant leap in making health assessments more accessible, cost-effective, and prompt, addressing a critical need in global health.

The use of artificial intelligence in healthcare has been extensively documented (JIANG *et al.*, 2017; KELLY *et al.*, 2019). Chatbots, powered by advanced algorithms and natural language processing capabilities, have been successfully employed in different healthcare applications, from mental health support to chronic disease management (OMOREGBE *et al.*, 2020; OMAROV; NARYNOV; ZHUMANOV, 2023; AGGARWAL *et al.*, 2023). Additionally, the use of chatbots in cardiovascular medicine presents significant potential, especially for cardiovascular prevention (MONTAGNA; MARIANI; PENGU, 2023). However, their application in neurological disease screening still needs to be explored.

Incorporating Large Language Models (LLMs) as intelligent agents in healthcare chatbots secures a significant advancement in digital health technology (JAVAID; HALEEM; SINGH, 2023). LLMs, known for their exceptional natural language processing abilities, are adept at facilitating more natural and engaging user interactions. LLMs enhance conversation flow, act as agents, and select appropriate screening methods based on user responses (XI *et al.*, 2023; ABBASIAN *et al.*, 2023). However, their potential in neurological disease screening is relatively untapped. Existing research highlights the feasibility of using vocal biomarkers as indicators of neurological disorders, a concept central to developing my chatbot for this study.

Studies have shown that vocal biomarkers can effectively indicate neurological disorders, including Parkinson's disease, where vocal cord changes are prominent symptoms (NGO *et al.*, 2022). Pah et al. (PAH; MOTIN; KUMAR, 2022b) investigated the effectiveness of apparent

vocal tract length (VTL) of phoneme parameters for differentiating people with PD from healthy subjects. However, they limited their approach to using a single machine learning model, specifically Support Vector Machine (SVM) (CORTES; VAPNIK, 1995). While this is a crucial step, the performance has room for improvement. I explored a range of machine-learning techniques to identify the best performance option.

This study aimed to develop a chatbot suitable for multi-platforms for screening people with Parkinson’s disease (PwPD). Telegram, an easily accessible platform, was chosen for this work due to its user-friendly nature, and its API is free of charge. The proposed approach uses a large language model as an agent designed to check users’ messengers and trigger the vocal test. This methodology would deliver “Software As a Medical Device” (SAMD) that leverages the widespread availability and ease of messaging apps with the analytical power of AI for the screening of neurodegenerative diseases.

### 4.3.1 Methods

#### 4.3.1.1 Dataset

This work employs the Viswanathan’s dataset (VISWANATHAN *et al.*, 2019), which is a public dataset that contains audio recordings from 24 People diagnosed with Parkinson’s Disease and 22 age-matched control (AMC) individuals without any motor or neurodegenerative disease. The PwPD participants were enlisted from the Movement Disorders Clinic at Monash Medical Centre in Australia. Each participant gave recordings of three phonemes captured in a clinical room using a Samson-SE50 microphone. These recordings were stored as single-channel WAV files with a 48 kHz sampling rate and 16-bit resolution. The database includes recordings of PwPD in both medicated (on-state) and unmedicated (off-state) periods. However, only recordings taken during the On-state were considered for this study. For more details, refer to (PAH *et al.*, 2021).

#### 4.3.1.2 Feature Extraction

I extracted the apparent vocal tract length (VTL) information from the sound signals, similar to the approach by Pah *et al.* (PAH; MOTIN; KUMAR, 2022b; PAH *et al.*, 2023b). VTL refers to the estimated length of a person’s vocal tract when they utter a particular sound and is determined based on the frequency of the formants. This technique has applications in different vocal analyses, including speaker verification and the identification of bodily measurements (TAN *et al.*, 2021; PISANSKI *et al.*, 2014).

The VTL for each recording was estimated in centimeters using the average of the four formants,  $F_i$ , following the formula by Pisanski (PISANSKI *et al.*, 2014):

$$VTL(F_i) = (2i - 1) \frac{c}{4F_i} \quad (4.2)$$

The fixed value,  $c = 33,500\text{cm/s}$ , represents the velocity of the sound in a uniformly shaped tube that is closed at one end. All the features were normalized using StandardScaler (PEDREGOSA *et al.*, 2011b), which adjusts each feature to have a mean of zero and a standard deviation of one.

#### 4.3.1.3 Classification and Evaluation

The dataset was split into training (80%) and test sets (20%), and the the proposed model was evaluated using sensitivity, specificity, and accuracy measures. Six machine learning models were evaluated: Logistic Regression, Decision Tree, Random Forest, Gradient Boosting, AdaBoost, and  $K$ -Nearest Neighbors.

#### 4.3.1.4 Chat-bot

I have developed a chatbot-based system to make the tool available to the global community. This approach offers a cost-effective and user-friendly alternative to traditional diagnostic methods, particularly benefiting those without easy access to specialized healthcare services. Accessible via a Telegram link, my chatbot seamlessly guides users through conversations and audio tasks. Users record and submit an audio sample, receiving a screening result from the bot server.

This study utilized ChatGPT-3.5, an LLM from OpenAI fine-tuned on instructions, along with the LangChain framework to create the agent chain and implement the conversation buffer memory. Having an agent enables my chatbot to run different assessments of neurological diseases using video, audio, or images. The tool was designed to be used as a plug-in.

Different LLMs are currently available, including Claude, Gemini, ChatGPT, Mistral, Llama3, and GPT-3.5. While all these models could potentially meet the needs of my experiment, I selected GPT-3 due to its suitability for tasks requiring limited reasoning complexity and its cost-effectiveness.

Figure 23 illustrates the entire process. The proposed architecture allows for streamlined user interaction, primarily facilitated by ChatGPT acting as an intelligent agent. The decision-making capability of ChatGPT determines whether to initiate the "PD Assessment" based on user commands. The "PD Assessment" is a conversational flow, prompting users to record specific phonemes and submit the corresponding audio files. Upon receiving these files, my machine-learning model analyzes them. The results from this analysis are then communicated back to the user through ChatGPT, which conveys the findings in a user-friendly and accessible manner. This pilot is to test the potential of the platform for this application.

This integration enables the evaluation of subtle audio cues, empowering early screening. By combining chatbot technology with audio analysis, I am reshaping Parkinson's screening to be efficient and accessible for all.

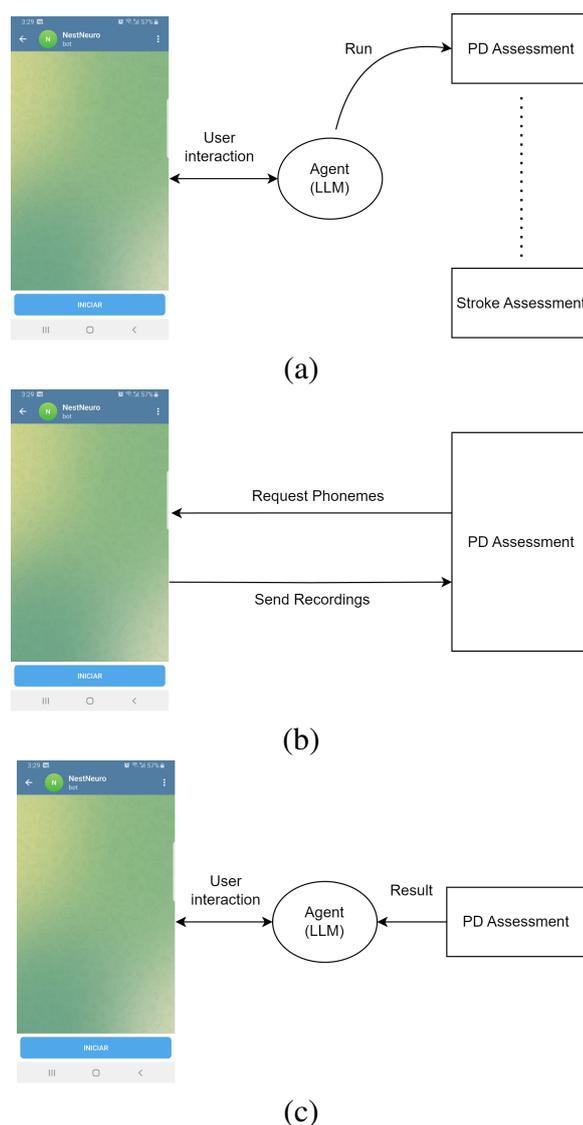


Figure 23 – Bot architecture: (a) The agent is responsible for starting the “PD Assessment” conversational flow; (b) The “PD Assessment” flow involves recording speech and running the machine learning model; (c) The result is returned to the agent, which then transmits it to the user in a user-friendly manner.

## 4.3.2 Experimental Results

### 4.3.2.1 Classification

The classification consists of identifying patients as positive or negative for PD. Table 38 presents the accuracy, sensitivity, and specificity concerning different machine learning models. Random Forest reached an accuracy of 0.98, sensitivity of 1.00, and specificity of 0.97. *K*-Nearest Neighbors and Gradient Boosting followed with accuracies of 0.92 and 0.90, respectively. AdaBoost and Logistic Regression showed moderate performances with accuracies of 0.85 and 0.77, respectively. These results highlight the varying effectiveness of each model in predictive analysis.

Table 38 – Comparison of model performances.

<b>Model</b>	<b>Accuracy</b>	<b>Sensitivity</b>	<b>Specificity</b>
Random Forest	0.98	1.00	0.97
K-Nearest Neighbors	0.92	0.97	0.88
Gradient Boosting	0.90	0.94	0.86
AdaBoost	0.85	0.86	0.84
Logistic Regression	0.77	0.86	0.70

#### 4.3.2.2 Chat-bot

In the chatbot-based screening system accessed via a Telegram link, users are greeted with a warm welcome. You can start the assessment asking for a voice health check for PD. The chatbot provides clear instructions on how to record and submit audio samples, emphasizing the need to articulate specific phonemes - /a/, /o/, and /m/ - for accurate analysis. Users are guided on recording. After recording, the audio is sent to the chatbot, which acknowledges receipt and executes the machine learning model. The model then analyzes the sounds. The result, along with any relevant recommendations, is promptly communicated back to the user, ensuring a seamless and informative experience. Finally, the chatbot offers guidance on further steps, maintaining a friendly and supportive tone throughout the interaction.

Figure 24 shows the screenshot of the user interacting with the chatbot. The results indicate that ChatGPT has learned to conduct a voice test to identify PD.

### 4.3.3 Discussion

Changes in the voice of people suffering from neurodegenerative disease are well recognized and routinely used by clinicians to evaluate their patients. However, there are large differences between the voices of different people, and hence, determining pathological or natural differences in the voice is a challenge. This can result in a high level of misdiagnosis that can result in poor patient outcomes.

I have developed a method that can overcome the above limitations. The Telegram chatbot-based screening system, aka NestNeuro, is a significant advancement, offering an accurate assessment of patients' voices.

NestNeuro has several advantages over other options. One key difference lies in Telegram's accessibility and familiarity as a platform. Unlike web-based systems that may require users to navigate multiple pages or unfamiliar interfaces, Telegram's widespread usage and intuitive design make the chatbot more user-friendly, especially for those less comfortable with technology. It operates on various devices, with apps available for smartphones (iOS, Android), tablets, and desktops (Windows, macOS, Linux). The service's cloud-based nature means users can access messages from multiple devices, maintaining synchronization across platforms.

The integration of chatbot technology in this context is crucial. Chatbots facilitate a more



Figure 24 – Screenshot of NestNeuro.

interactive and engaging experience, resembling a natural conversation, which can be particularly comforting for users who might feel overwhelmed by the clinical nature of traditional screening tools. Moreover, chatbots can provide instantaneous responses and guidance, reducing anxiety and uncertainty associated with waiting times often experienced in web-based systems.

A significant advantage of using Telegram for this screening tool is the ease of operating its built-in audio recording and transmission capabilities. Users can effortlessly record and submit their vocal samples directly through the app, streamlining the process. This integration eliminates the need for additional software or tools, making the system more accessible, especially for users with limited technical resources or knowledge.

LLM is important in my approach as it enables seamless interaction with the user and serves as an agent to initiate the vocal test. The architecture allows for the inclusion of more models that assess other symptoms, such as hypomimia and neurological conditions like stroke and amyotrophic lateral sclerosis. Works that use Action Units and Facial Graph Points can be coupled (OLIVEIRA *et al.*, 2023; OLIVEIRA *et al.*, 2024a; GOMES *et al.*, 2023; GOMES *et*

*al.*, 2024b), for NestNeuro is designed to be a plug-in.

However, this approach has limitations, for it depends on Telegram, a single platform, and excludes other similar options. Additionally, the outcomes may depend on the quality of the audio recordings, the user's device, the environment, and adherence to recording instructions. Finally, this study was conducted only using one small dataset. Therefore, the power of generalizability has yet to be discovered.

#### 4.3.3.1 *Novelty of this study*

This study introduces a novel approach by integrating chatbot technology with vocal biomarker analysis for Parkinson's disease screening, marking a significant advancement in digital health diagnostics. By harnessing the capabilities of large language models to serve as intelligent agents, my system enables a natural, conversational user experience while guiding individuals through precise vocal tests. This method overcomes traditional limitations of voice-based diagnostics by addressing the inherent variability in human speech, thereby enhancing both the accuracy and accessibility of early PD screening. Moreover, using the Telegram platform as a delivery mechanism not only leverages a widely adopted communication tool to facilitate cost-effective, user-friendly health assessments but also opens avenues for the future inclusion of multi-modal assessments—such as video and image analysis—for broader neurological screening.

#### 4.3.4 **Conclusion**

The Telegram chatbot-based screening system offers an innovative and user-friendly alternative to traditional web-based screening tools. Its reliance on a popular messaging platform for recording and sending audio provides ease of use, although it faces limitations in accessibility and potential variability in audio quality. Incorporating text-to-speech tool can further enhance its capabilities in talking, making it a promising mechanism for remote healthcare screening.

#### 4.3.5 ***CRediT authorship contribution statement***

I am grateful for the support and collaboration that made this work possible. As the lead investigator, I was responsible for the research, design, methodology development, and writing of this study. I acknowledge Nemuel D. Pah for his role in software validation; Quoc C. Ngo for his contributions in project administration, validation, and manuscript review and editing; João P. Papa for his efforts in supervision, project administration, and review and editing; and Dinesh Kumar for his guidance in conceptualization, supervision, and manuscript review and editing. Their collective contributions were essential to the success of this study.

---

## AI-POWERED SYNTHETIC IMAGING

---

The integration of advanced image analysis techniques in medical diagnostics has revolutionized the field, offering enhanced accuracy and efficiency in detecting, monitoring, and treating various conditions. This chapter examines the role of synthetic image generation in advancing non-invasive diagnostics by addressing key challenges, such as limited data availability and image translation issues. It investigates the application of state-of-the-art generative models, focusing on Generative Adversarial Networks (GANs) and Diffusion models, in medical image analysis. Specifically, it explores the use of StyleGAN-2 for data augmentation to improve the detection of age-related macular degeneration (AMD) and the application of Stable Diffusion for estimating thermal images from RGB images for leg ulcer assessment.

Section 5.1 was published as “**A Stable Diffusion Approach for RGB to Thermal Image Conversion for Leg Ulcer Assessment**” in the *IEEE 37th International Symposium on Computer Based Medical Systems (CBMS)*. The work in Section 5.2 was published in the *Biomedical Signal Processing and Control* journal, titled “**Robust deep learning for eye fundus images: Bridging real and synthetic data for enhancing generalization**”.

### 5.1 Stable Diffusion for Thermal Image Estimation

Chronic leg ulcers, a significant global health burden, affect millions worldwide each year. These complex wounds, characterized by their prolonged healing process and high recurrence rates, present a substantial challenge in terms of both patient care and healthcare system strain (CAI *et al.*, 2023; MULDOON, 2023; PEREIRA *et al.*, 2022). The effective management and treatment of leg ulcers are crucial for patient well-being and reducing the economic impact on health systems globally. Current methods for monitoring these ulcers typically involve RGB (Red, Green, Blue) photographic documentation, providing essential visual information. However, this lacks the thermal data that has been found useful in the assessment of the wound’s condition, particularly in terms of microcirculation and inflammation (GONZÁLEZ *et al.*, 2024).

Figure 25 illustrates a thermal image from a human leg that highlights the ulcer.

Using thermal imaging in evaluating leg ulcers has shown a more accurate and early assessment of chronic wounds (MONSHIPOURI *et al.*, 2021). It also provides a more comprehensive understanding of the wound's physiological state (DINI *et al.*, 2015). However, while RGB images can be recorded using standard smartphones, thermal images require expensive cameras and are not routinely available. On the other hand, some works have shown the potential of estimating the thermal images from the RGB images using deep-learning techniques (NGO; OGRIN; KUMAR, 2022; OGRIN *et al.*, 2023), bridging the gap between the simplicity and ubiquity of RGB imaging and the information provided by thermal imaging, that could be helpful for the clinicians. While the spectrum of the RGB images is lower than the thermal images that are typically in the near infrared range, it is hypothesized that because of the border of the two spectrums, such estimation may be possible. However, estimating thermal images from RGB data concerning leg ulcers still needs to be reported in the literature.

Advances in machine learning, especially in diffusion techniques, have shown promising results (YANG *et al.*, 2023). Such technologies have the potential to develop a model to estimate one modality from the other as has been shown (MENG *et al.*, 2021). This could assist clinicians in estimating the thermal images of the ulcers while using a smartphone. However, the specific application of these techniques in leg ulcer management remains unexplored.

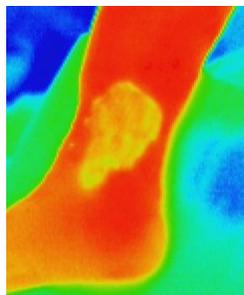


Figure 25 – Example of a thermal image: the ulcer is clearly highlighted.

I have explored a fine-tuned Stable Diffusion (ROMBACH *et al.*, 2022) model for transforming RGB photographs of leg ulcers into thermal images. By employing DreamBooth (RUIZ *et al.*, 2023) and LoRA (HU *et al.*, 2021) for model fine-tuning and incorporating ControlNet (ZHANG; RAO; AGRAWALA, 2023) for enhancing the image transformation process, I aim to create a pipeline that estimates thermal images from RGB photos of the leg ulcers. The estimated thermal images were compared with the thermal images obtained using the thermal camera and the performance was recorded. This study could improve the screening and monitoring of wounds and facilitate the accessibility of advanced assessment tools by using only a smartphone camera, potentially leading to better patient outcomes without requiring expensive equipment.

This section main contribution is to propose a methodology and testing its performance

for converting RGB images to estimate thermal images for leg ulcer assessment. To my knowledge, this is the first attempt to reconstruct thermal images from RGB images of the leg ulcers.

### 5.1.1 Methods

This section describes the datasets that were used in the study, the techniques employed to generate the synthetic images, and the methodology to evaluate the different neural architectures considered in the experimental section.

#### 5.1.1.1 Dataset

The study requires a dataset that has the thermal and RGB images of the ulcers recorded simultaneously. For this purpose, the dataset of Venous Leg Ulcers (VLU) collected by the Biosignals Laboratory at the Royal Melbourne Institute of Technology (RMIT), Australia, was employed (MONSHIPOURI *et al.*, 2021; NGO; OGRIN; KUMAR, 2022; OGRIN *et al.*, 2023). To ensure consistency in data collection, all images were taken by a single trained individual. The ulcers were photographed weekly using a Nikon D90 DSLR camera, and thermal images were captured with ULRIvision TI160 (Zhejiang Ulirvision Technology Co., Ltd). Ethical approval for conducting data collection was obtained from the Human Research and Ethics Committee of Bolton Clarke (Project number 194) and RMIT University (BSEHAPP 21-15), Melbourne, Australia. I used 12 picture pairs, each consisting of an RGB and a thermal image. The first set of six pairs was employed to fine-tune my Stable Diffusion method through Lora Dreambooth. Subsequently, I used an extra set of six RGB and thermal image pairs to test the proposed approach.

#### 5.1.1.2 Preprocessing

In the preprocessing phase, I adapted the dataset to match the requirements of Stable Diffusion v1.5, which is optimized for  $512 \times 512$  pixel images. This required manually cropping the original images to the specified dimensions above, particularly on the wound area. Subsequently, I employed the ‘Segment Anything’ (KIRILLOV *et al.*, 2023) approach to effectively isolate and remove the background, ensuring that my focus remained solely on the leg with the wound. This step was critical to minimize potential biases and noise in my training. This approach was conducted on all images, i.e., RGB and thermal. Figure 26 depicts the original RGB photo and its appearance after preprocessing.

#### 5.1.1.3 Generation

I employed LoRA Dreambooth to fine-tune Stable Diffusion 1.5. The software used for this implementation was Kohya<sup>1</sup>. ControlNet was not fine-tuned. I chose the ‘Standard’ LoRA

<sup>1</sup> <[https://github.com/bmaltais/kohya\\_ss](https://github.com/bmaltais/kohya_ss)>

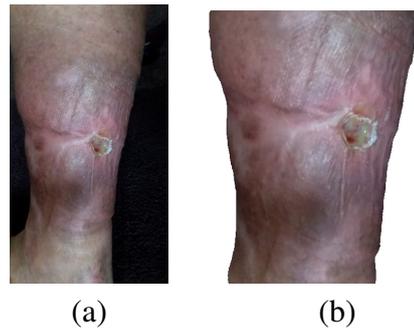


Figure 26 – Preprocessing step: (a) original RGB photo and (b) its postprocessed version.

type, and key features such as an ‘adaptive\_noise\_scale’ of 0 and ‘enable\_bucket’ set to true with ‘bucket\_reso\_steps’ at 64 were included for stability and efficiency.

The training was set to run during four epochs with a learning rate of 0.0003, managed by a cosine learning rate scheduler. I employed mixed precision (‘fp16’) to balance computational demands with performance. The ‘Lion8bit’ optimizer was chosen for its efficiency, complemented by a training batch size of 1, which allowed for effective resource utilization without compromising the training quality.

My approach extended to specific details like the ‘caption\_extension’ set to ‘.txt’, facilitating ease in handling text-based inputs for transforming images. Each .txt file contains a caption formatted as ‘x, thermal photo of the leg with ulcer’ where ‘x’ represents the image ID. The model’s ‘network\_dim’ was set to 8, which played a crucial role in determining the network’s capacity and performance. I also set the ‘save\_every\_n\_epochs’ to 1, ensuring regular checkpoints of the model’s state for potential recovery or analysis.

The RGB photos were placed in the regularization data folder. Each corresponding .txt file has a caption, ‘y, RGB photo of the leg with ulcer,’ where ‘y’ denotes the image ID.

The settings were thought for efficient resource management: a maximum bucket resolution of 2048 and a maximum training resolution of  $512 \times 512$  pixels. My approach to handling model outputs and logging was methodical, directing these to specific directories within my user data path, and I opted to use ‘runwayml/stable-diffusion-v1-5’ from my model list.

After the training phase, I used Automatic1111<sup>2</sup>, a web interface for Stable Diffusion, implemented using the Gradio library, to load the LoRA weights from Kohya and generate images using ControlNet. The images were generated using an image-to-image pipeline. To enhance and maintain the integrity of the shapes, I conditioned the process on ControlNet, utilizing both a soft model and a recolor approach.

<sup>2</sup> <<https://github.com/AUTOMATIC1111/stable-diffusion-webui>>

#### 5.1.1.4 Prompt

In my study, I employed two different prompts, each tailored for specific condition tasks. The first prompt was designed for ControlNet conditioned on soft edges. The details of this process included 30 steps, using the Euler A sampler with a CFG scale of 6 and a seed value of 64. The image size was set to  $512 \times 512$ , using the v1-5-pruned model. The denoising strength was calibrated at 0.6. ControlNet 0 was configured with "Text Prompt: 'thermal photo of the leg with ulcer', Module: softedge\_pidinet, Model: control\_v11p\_sd15\_softedge, Weight: 1, Resize Mode: Crop and Resize, Low Vram: False, Processor Res: 512, Guidance Start: 0, Guidance End: 1, Pixel Perfect: True, Control Mode: Balanced, Hr Option: Both, Save Detected Map: True".

For ControlNet conditioned on Recolor, a similar approach was taken. This process involved 35 steps with the same Euler A sampler, CFG scale, seed, and image size as the first. However, it differed in using the v1-5-pruned model combined with the VAE: vae-ft-mse-840000-ema-pruned.safetensors. The denoising strength remained at 0.6. ControlNet 0 specifications for this process were "Text Prompt: 'thermal photo of the leg with ulcer', Module: recolor\_luminance, Model: ioclab\_sd15\_recolor, Weight: 1, Resize Mode: Crop and Resize, Low Vram: False, Threshold A: 1, Guidance Start: 0, Guidance End: 1, Pixel Perfect: True, Control Mode: Balanced, Hr Option: Both, Save Detected Map: True".

#### 5.1.1.5 Evaluation Measures

To evaluate my approach, I employed the well-known Structural Image Similarity Index (SSIM) (WANG *et al.*, 2004) on the test set. The SSIM index between two images  $x$  and  $y$  is calculated as follows:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)},$$

where  $\mu_x, \mu_y$  denote the images' mean values,  $\sigma_x^2, \sigma_y^2$  are their pixels' brightness variances,  $\sigma_{xy}$  is the covariance among those values, and  $C_1, C_2$  are constants to stabilize the division. This index assesses similarity between the thermal image generated by Stable Diffusion and the original thermal one, with a range from -1 to 1, where 1 indicates perfect similarity.

## 5.1.2 Results

Figure 27 is an example of the thermal image estimation of the venous leg ulcer. Derived from RGB image 27(a), the images 27(b) and 27(c) were thermal photos generated without and with using LoRA Dreambooth fine-tuned weights employing ControlNet with Soft Edge. Figure 27(b) fails to accurately represent a thermal image of a leg with an ulcer, whereas Figure 27(c) demonstrates that a generated photo has learned from the distribution of my dataset.

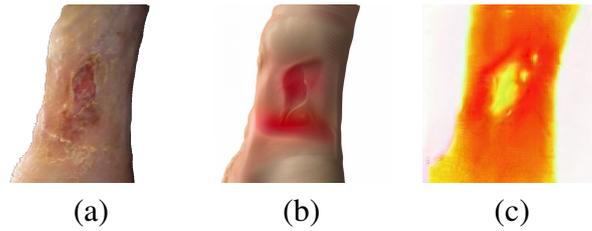


Figure 27 – Comparison of a venous leg ulcer image: (a) shows the original RGB photo, while (b) and (c) display the thermal photo before and after LoRA Dreambooth fine-tuning, respectively.

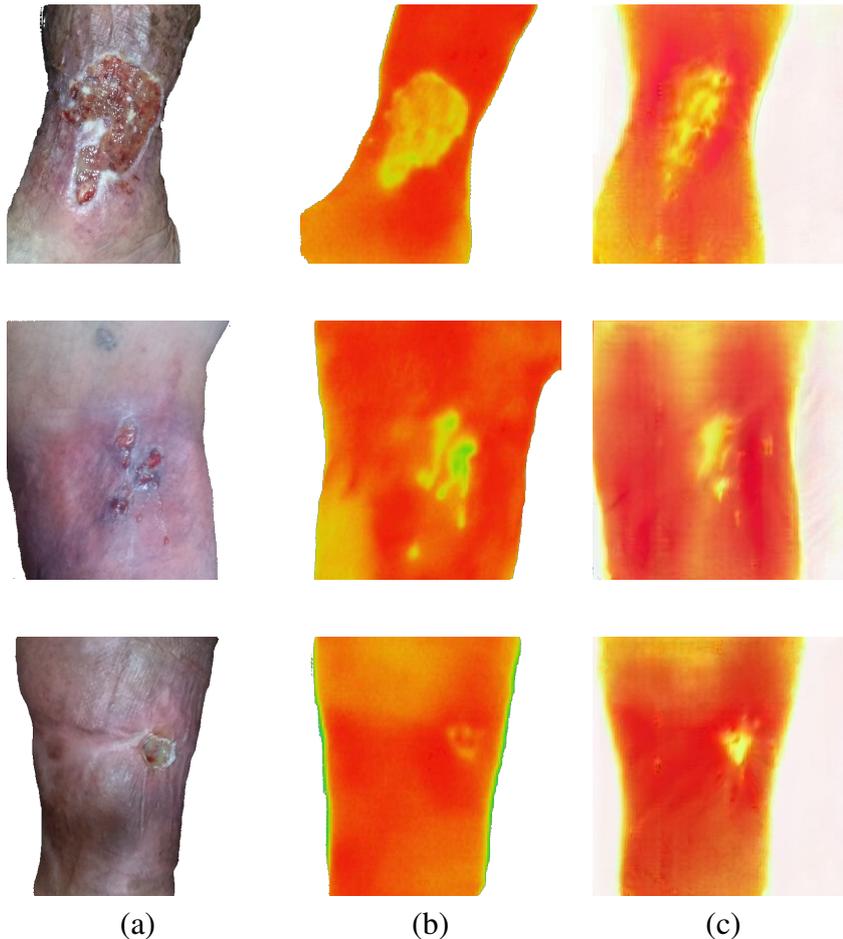


Figure 28 – Results of ControlNet conditioned on Soft edges, featuring (a) RGB image, (b) captured thermal image, and (c) Stable Diffusion generated thermal image.

My approach achieved an average SSIM score of 0.84 across the test set, comparing the original thermal images with the thermal ones generated by Stable Diffusion using Control Net with Soft Edge. Although my thermal image generation method achieves a high average SSIM score throughout the test, indicating that my thermal image generated approach preserves the essential structural and perceptual features, it fails to replicate the original contour in the wound region as shown in Figure 28.

Figure 28 presents some results using the soft preprocessing step in which the proposed approach did not perform as expected: Figures 28a-c show the RGB version, the original thermal

image, and the generated thermal image by Stable Diffusion, respectively. The shape of the leg was retained, with a similar color palette. However, the wound's shape was not properly reconstructed.

The comparison between the original and the generated images reveals notable similarities in the heat distribution patterns. The heat signature of the wound (Figure 28b) is evident but lacks the detailed contours found in the RGB image. In contrast, Figure 28(c) demonstrates Stable Diffusion's capacity to mimic the overall thermal profile while subtly altering the wound's shape, suggesting that while the algorithm is adept at reproducing general thermal patterns, it struggles with precise anatomical details.

Figure 29 presents the results where the proposed approach did perform well: Figures 29a-c show the RGB image, the reconstructed images generated by Stable Diffusion, and the wound segmented using 'Segment Anything' approach, respectively. One can observe the wound area is highlighted in a solid yellow, making it easier to identify.

In contrast to Figure 28, Figure 29 illustrates a more distinct approach in wound visualization and segmentation. The recolor preprocessing employed in Figure 29b significantly enhances the wound's visibility against the background, using a vivid yellow to delineate the affected area. This enhancement dramatically aids in the subsequent segmentation process, as shown in Figure 29c, where 'Segment Anything' can isolate the wound accurately. This suggests that recolor preprocessing could be helpful in medical imaging and analysis.

### 5.1.3 Discussion

This work has shown the possibility of estimating the thermal images of venous leg ulcers from the RGB versions using stable diffusion. Visual comparison of the estimated and the original thermal images shows the potential of using this approach to assist clinicians in assessing the VLU without requiring specialized equipment. This approach differs fundamentally from Generative Adversarial Networks (GANs) (PAN *et al.*, 2019), another image-to-image transformation option. While GANs involve two networks competing against each other – a generator and a discriminator – to produce realistic images, stable diffusion operates through a series of diffusion processes that gradually adjust the image toward a desired output.

This study has not employed GAN approaches, such as Pix2Pix (ISOLA *et al.*, 2017) and CycleGAN (ZHU *et al.*, 2017), due to the amount of data required. These approaches can lead to overlap and overfitting in a small dataset. In contrast, to fine-tune Stable Diffusion using LoRA DreamBooth, I only need a few high-quality images to obtain promising results. Besides that, this technique offers the advantage of having more controlled transformations using ControlNet.

The thermal images are very useful in leg ulcer management. While the thermal and RGB images correspond to different wavelengths, I hypothesis that due to the proximity of the two spectrums, it is possible to estimate thermal images from RGB images. A reasonable estimate



Figure 29 – Outcomes of ControlNet conditioned on recolor: (a) the RGB image, (b) the Stable Diffusion-generated image with the wound highlighted in strong yellow for easier segmentation, and (c) the wound segmented using "Segment Anything".

of the thermal image of the VLU can assist the clinician in better visualizing the ulcers. While thermal imaging provides essential information about the wound's physiological state, such as inflammation and microcirculation, which are not seen in RGB images, I am unable to check if this information is available in the estimated thermal images obtained by my method. However, the visual appearance of the ulcers in the estimated and actual thermal images is similar, and the SSIM score is 0.84. While this is a pilot, it shows the potential of this approach, but further work is essential.

Works by [Monshipouri \*et al.\* \(2021\)](#), [Ngo, Ogrin and Kumar \(2022\)](#) have shown the effectiveness of using computerized assessment of thermal images to differentiate between chronic and regular VLU. The next step in this work is an objective measure of the relevant information in the estimated thermal images. This information is crucial for diagnosis, monitoring progress, and tailoring treatment strategies, ultimately contributing to improved patient care and outcomes.

In my approach, ControlNet and LoRA DreamBooth played significant roles. LoRA DreamBooth enhanced the stable diffusion process by fine-tuning the model to generate more

specific and relevant thermal images from RGB photos of leg ulcers. Conversely, ControlNet addressed a significant challenge encountered in the initial application of stable diffusion. While stable diffusion is conditioned by text and demonstrates remarkable capabilities in image transformation, its application in my context revealed a limitation in precisely maintaining the wound's position in the new thermal image. ControlNet tries to mitigate this problem by incorporating edge detection, enabling more accurate conditioning of the stable diffusion process and thereby maintaining the integrity of the wound's location in the transformed image.

This study has some limitations. One significant constraint is the reliance on a relatively small dataset, which might affect the generalizability of the findings. The diversity and size of the dataset are crucial in machine learning, especially in medical applications, where variations in wound characteristics are vast. The limited dataset size may restrict the model's ability to learn and adapt to the wide range of possible presentations of leg ulcers, potentially impacting the applicability of the thermal images produced.

While my approach using stable diffusion, enhanced by DreamBooth and ControlNet, marks a substantial advancement in leg ulcer imaging, further work is needed. The first is to obtain a larger dataset. The second is to test the estimated thermal images for computerized analysis to detect chronic wounds.

#### 5.1.3.1 Novelty of this study

This study pioneers RGB-to-thermal image conversion for venous leg ulcer assessment using Stable Diffusion, overcoming the cost barrier of thermal imaging. Unlike previous GAN-based methods, my approach, fine-tuned with DreamBooth, LoRA, and ControlNet, requires minimal data while achieving a high SSIM score (0.84). This enables clinicians to visualize thermal patterns using standard RGB images, expanding accessibility to wound assessment tools. Beyond its technical innovation, this study provides new insights into estimating thermal characteristics from RGB images in a medical context. I improve anatomical consistency by integrating ControlNet for edge retention, addressing a key limitation in AI-driven medical imaging. This approach opens new possibilities for AI-assisted wound monitoring in telemedicine and low-resource settings.

#### 5.1.4 Conclusion

My study presents a method to estimate the thermal images of venous leg ulcers from the RGB images. The estimated images can assist the clinician in better observing the ulcers. This method is based on stable diffusion techniques, and it enhances them with DreamBooth and ControlNet to convert RGB photographs into thermal images. This innovative approach could help a clinician estimate the thermal image of the ulcer using an RGB image acquired with a smartphone and without requiring the thermal camera.

Future work should test the use of these estimated images to detect chronic wounds and enhancing the contour definition within the wound region in the estimated thermal images. A bigger dataset is also needed. Another future work is testing this technique for other medical imaging applications.

### 5.1.5 CRediT authorship contribution statement

I am grateful for the support and collaboration that made this work possible. As the lead investigator, I was responsible for the research, design, methodology development, and writing of this study. I acknowledge Quoc C. Ngo for his contributions in project administration, validation, and manuscript review and editing; João P. Papa for his efforts in supervision, project administration, and review and editing; and Dinesh Kumar for his guidance in conceptualization, supervision, and manuscript review and editing. Their collective contributions were essential to the success of this study.

## 5.2 StyleGAN-2 for Synthetic Eye Fundus Image Generation

Age-related macular degeneration (AMD) is a major cause of vision impairment and has affected approximately 200 million people worldwide in 2020 (JONAS, 2014). With an aging population, such numbers are expected to rise to 288 million by 2040 (WONG *et al.*, 2014). AMD is a progressive disorder of the macular region that causes central vision loss and is one of the most common causes of irreversible vision impairment in people over 50 years-old (HARVEY, 2003). Figure 30 depicts an example of a retina fundus affected by AMD.

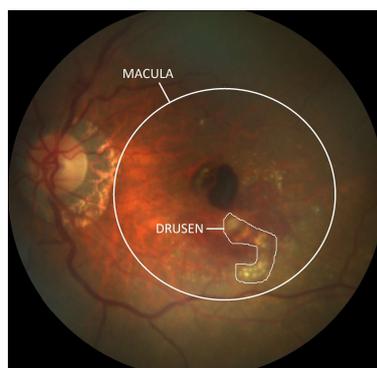


Figure 30 – Retina fundus image positive to age-related macular degeneration identified by the presence of drusen. The image was extracted from the iChallenge-AMD dataset (FU *et al.*, 2020).

StyleGAN2 has been utilized to generate detailed images of the eye fundus. Mayya *et al.* (MAYYA *et al.*, 2023) performed data augmentation with StyleGAN2. They developed a method to diagnose a range of conditions such as myopia, diabetic retinopathy, age-related macular degeneration, glaucoma, and cataract. Meanwhile, Wang *et al.* (WANG *et al.*, 2023)

employed a distinct approach using StyleGAN2 and augmentation, categorizing conditions into four classes: no AMD, early AMD, intermediate AMD, and advanced AMD. However, Mayya et al. (MAYYA *et al.*, 2023) and Wang et al. (WANG *et al.*, 2023) didn't specify a method for removing images with significant quality issues, such as those that are blurry, have low contrast, or are inadequately illuminated. Veturi et al. (VETURI *et al.*, 2023) conducted a diagnostic study focusing on gene-labeled fundus autofluorescence (FAF) images for inherited retinal diseases (IRDs). However, They found that synthetic data augmentation failed to enhance disease classification when applied to IRD datasets.

This work has introduced an alternative approach for generating synthetic images for training deep networks and tested it for AMD identification, which consists in using a retinal image quality assessment model (FU *et al.*, 2019) and the StyleGAN2-ADA (KARRAS *et al.*, 2020). Retina images, positive and negative to AMD, from multiple databases having a range of image qualities and lesions were used. Ten different GAN architectures were compared to generate synthetic eye-fundus images and the quality was assessed using the Fréchet Inception Distance (FID), two independent clinical experts who were label blinded and deep-learning classification. Different percentages of synthetic data were employed in the augmentation.

The primary contributions of this work are fourfold:

- To introduce StyleGAN2-ADA (KARRAS *et al.*, 2020) for eye fundus image generation;
- Test generalization across different datasets;
- Free access to software for generating the synthetic images;
- Accessible web-based tool for diagnosing AMD that combines computer vision and deep learning techniques.

### 5.2.1 Methods

This subsection describes the datasets that were used in the study, the techniques employed to generate the synthetic images, and the methodology to evaluate the different neural architectures considered in the experimental subsection.

#### 5.2.1.1 Dataset

This study used eye-fundus images from four public datasets from four countries which have been reported in the literature. These capture the typical differences due to equipment and demographics, which is necessary for verifying the generalisability of the data. Below, I summarized the datasets' preliminary information:

- **iChallenge-AMD:** Comprises of 1,200 retinal fundus images that have been annotated for drusen and hemorrhage. The training set was made of 400 images (89 images of eyes

with AMD and 311 from eyes without AMD), while the test sets contained the remaining images (FU *et al.*, 2020).

- **ODIR-2019:** Contains colored fundus images from both left and right eyes of 5,000 patients obtained from multiple hospitals/medical centers in China, with varying image resolutions and observations from several specialists. The dataset has been designed to address normal and six diseases: diabetes, glaucoma, cataract, AMD, hypertension, myopia, and other diseases/abnormalities<sup>3</sup>. The training set is made up of a structured dataset with 7,000 images, of which 280 images are labelled as having AMD. The testing set consists of 500 colored fundus images, eliminating age and gender.
- **RIADD:** contains 3,200 fundus images recorded using 3 different cameras and multiple conditions. The images have been annotated through the consensus of two retina experts. The dataset has been sub-divided based on six diseases/abnormalities; diabetic retinopathy, AMD, media haze, drusen, myopia, and branch retinal vein occlusion (PACHADE *et al.*, 2021). The dataset was subdivided into three subsets: 60% for training (1,920 images), 20% for validation (640 images), and the remaining 20% for testing purposes (640 images).
- **STARE:** STRUCTURED Analysis of the RETina (STARE), contains 400 high-quality fundus images with different diseases.<sup>4</sup> I randomly sampled 40 AMD and 40 normal images from STARE for testing my model. This database was not part of the mix that was used to train the model and hence testing the model from images from this database was a test for its generalisability. All the ground truths are annotated by the expert.

I have used publicly available and online datasets, each of which have described their patient recruitment, clinical outcomes and human experiments ethics approval in their associated publications. Registration was necessary for access to the data from the three sources.

i-Challenge dataset authors confirm in their publications they received ethics approval guidelines from Sun-Yat Sen University, China and Sixth People's Hospital Affiliated to Shanghai Jiao Tong University, Shanghai, China. Details on <https://amd.grand-challenge.org/Home/>

ODIR dataset confirm that all data is from routine clinical examinations, have been de-identified and published after receiving clearance from Peiking University board. Details on <https://odir2019.grand-challenge.org/dataset/>

RIADD dataset confirms that data collection was conducted after ethics approval from Review Board of Shri Guru Gobind Singhji Institute of Engineering and Technology, Nanded, India. Details on <https://riadd.grand-challenge.org/>

<sup>3</sup> <https://odir2019.grand-challenge.org/dataset/>

<sup>4</sup> <https://cecas.clemson.edu/~ahoover/stare/>

The datasets mentioned above were designed to address a challenge and hence the labels of the test subset were not available. Therefore, I used only the training subset.

These datasets have an imbalanced class distribution, which is an inherent problem for most medical image datasets and using such a set generally leads to a bias towards the predominant class. Different research groups have developed these datasets with differences in the quality of the images and the demographics. Thus, these datasets offered both, data imbalance and wide range of image quality.

The ODIR-2019 and RIADD datasets were organized into two subsets, AMD and non-AMD images. Preprocessing methodology and quality classification was same as proposed by Fu et al. (FU *et al.*, 2019), which comprises of a step that detects the retinal mask using the Hough Circle Transform and then crops it to remove the impact of the black background. The cropped region was then scaled to a  $224 \times 224$  resolution. The resultant images were submitted to a DenseNet trained by Fu et al. (FU *et al.*, 2019) which classified these in 3 classes: “Good”, “Usable” and “Reject”. The rejected images were of poor quality based on significant blurring, low contrast, inadequate illumination, which has been described by Fu et al. (FU *et al.*, 2019). As a result, the number of images positive to AMD decreased from 89 to 74, 280 to 227, and 100 to 79 images in the iChallenge-AMD, ODIR-2019, and RIADD datasets, respectively while the number of non-AMD images decreased from 311 to 290, 6,720 to 4,993, and 1,820 to 1,143 images concerning the iChallenge-AMD, ODIR-2019, and RIADD datasets, respectively. Figure 31 illustrates the steps mentioned above for a sample image from the ODIR-2019 dataset.

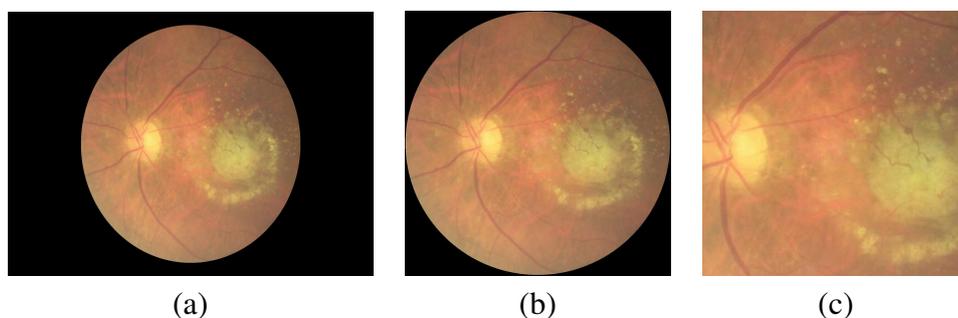


Figure 31 – Sample image extracted from ODIR-2019 dataset and its corresponding transformations: (a) original image, (b) background removal using Hough Circle Transform and resizing, and (c) central cropping.

These images were then resampled to  $390 \times 390$  pixels, followed by a cropping procedure keeping the center of the image to  $256 \times 256$  pixels. Such a procedure is required to drive StyleGAN2-ADA generating images focused on the macula area (Figure 31-c). Ultimately, images were resized  $224 \times 224$  pixels and normalized within the range  $[-1, 1]$  to be used as proper inputs to the deeper architectures considered in the manuscript.

After quality assessment and selecting the images for the final dataset using the criterion described earlier, the resulting single dataset comprised total of 7,106 images. In this, 6,896 images were used to train the models (275 with AMD and 6,621 without AMD) and the

remaining 210 images (105 with AMD) were used as a holdout test set. Figure 32 displays the number of images used per dataset to compose the final test set.

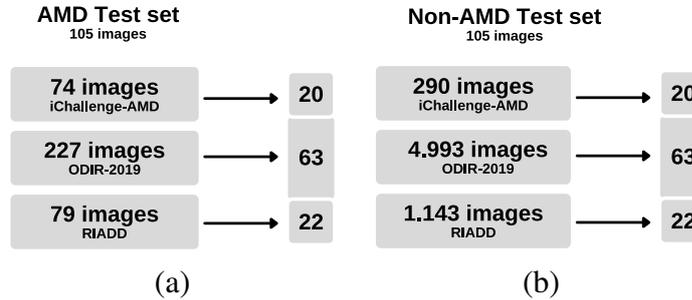


Figure 32 – Number of images per dataset to compose the test set: (a) images positive to AMD and (b) non-AMD images.

### 5.2.1.2 Evaluation Measures

Evaluating the quality of synthetic images is important for establishing their usability in practical applications, such as training deep learning models. It can significantly influence the training of these models. If the data does not accurately represent reality or lacks diversity, the synthetic data may introduce noise into the training, decreasing the model performance.

I employed three evaluation measures, i.e., the Fréchet Inception Distance (FID), a well-known GAN evaluation score (HEUSEL *et al.*, 2017), the ability of human experts to identify the synthetic images and the classification accuracy. FID is often used to assess the quality and variety of the generated images and, even though it has been proposed to improve the standard Inception Score, it still uses the Inception architecture to extract features from both, synthetic and real images.

Structural Similarity Index (SSIM), and Peak Signal-to-Noise Ratio (PSNR) focus on pixel-wise comparisons and are limited in their capacity to assess higher-level features and perceptual quality. FID leverages the features extracted from a pre-trained Inception Network to measure the similarity of feature representations between real and generated images.

### 5.2.1.3 Experimental Setup

There were four stages of the experiments: (i) comparison of ten different GAN architectures, (ii) evaluation of synthetic images based on human experts' ability to distinguish between real and synthetic images, (iii) evaluation of data augmentation with synthetic images in three different deep learning networks, and (iv) measuring the accuracy in identifying AMD images between human experts and deep learning networks trained with mixed data, with both real and synthetic images.

The first experiment employed FID to compare StyleGAN2-ADA and nine distinct GAN models. The following models were considered in the experiment: Deep Convolutional GAN

(DCGAN) (RADFORD; METZ; CHINTALA, 2015), Least Squares Generative Adversarial Networks (LSGAN) (MAO *et al.*, 2017), Wasserstein GAN (WGAN) (ARJOVSKY; CHINTALA; BOTTOU, 2017), Wasserstein GAN with Gradient Penalty (WGAN-GP) (GULRAJANI *et al.*, 2017), Deep Regret Analytic Generative Adversarial Networks (DRAGAN) (KODALI *et al.*, 2017), Energy-based Generative Adversarial Network (EBGAN) (ZHAO; MATHIEU; LECUN, 2016), Boundary Equilibrium Generative Adversarial Networks (BEGAN) (BERTHELOT; SCHUMM; METZ, 2017), Conditional GAN (CGAN) (MIRZA; OSINDERO, 2014), and Auxiliary Classifier GAN (ACGAN) (ODENA; OLAH; SHLENS, 2017). All models were trained with 50 epochs, considering samples of size  $100 \times 100$  pixels and a batch size of 32. The training step used the ADAM (KINGMA; BA, 2017) optimizer with a learning rate of 0.0002 and decay rates of 0.5 and 0.999 regarding the generator and the discriminator, respectively. The experiments were conducted using the training set with an Nvidia RTX 2060 GPU. Therefore, after training each GAN model, new images were generated and FID was computed.

The second experiment determines whether clinical experts, who are very experienced with the analysis of eye fundus images, can distinguish between synthetic and real images. Such a step is essential to evaluate the effectiveness of StyleGAN2-ADA for generating synthetic eye fundus images. The experts were provided with randomly generated image sets, one for AMD-diagnosed images and the other for non-AMD images, each consisting of ten synthetic images and ten real images. They were asked to identify the synthetic images in the mix.

In the next experiment, I considered three deep architectures pre-trained with ImageNet dataset (DENG *et al.*, 2009b) for performance comparison when the model is augmented with synthetic images, i.e., SqueezeNet (IANDOLA *et al.*, 2016), AlexNet (KRIZHEVSKY; SUTSKEVER; HINTON, 2012), and ResNet18 (HE *et al.*, 2016). During training, synthetic and real images were mixed within each batch according to a pre-determined hyperparameter  $p \in [0, 1]$ . For each image, a uniform distributed number  $x$  was sampled and compared with  $p$ : if the latter was greater than  $x$ , the image was replaced by a synthetic one. I considered images generated by StyleGAN2-ADA, for it had obtained the best FID values in the first experiment (such outcomes are later described in subsection 5.2.2). The deep networks were trained using a learning rate of 0.0001, a decay rate of 0.9, batch size of 32, number of epochs equal to 5, and samples of size  $224 \times 224$  pixels.

To address the issue of unbalanced dataset, Weighted Random Sampler (EFRAIMIDIS; SPIRAKIS, 2008) was used. This approach can be implemented in PyTorch (PASZKE *et al.*, 2019). It assigns weights to each data point in the dataset. These weights are inversely proportional to the class frequency to which they belong. When batches of data are drawn for training the model during the training process, this sampler uses these weights to influence the selection process, making it more likely for points from minority classes to be included in each batch. Further augmentation was performed using classical image transformations, such as resizing and color jittering, which changes the brightness, contrast, saturation, and horizontal

flipping. The test set was kept intact.

The final round was aimed at comparing the best deep model obtained in the previous phase, ResNet-18, against human experts concerning the task of AMD identification. In this step, twenty real images (ten AMD images and ten non-AMD images) were randomly selected from the test set and provided to human experts for classification purposes. Following that, the same images were submitted to a ResNet-18 for comparison purposes. To allow a fair comparison between humans and deep models, I considered the following measures: standard accuracy (ACC), sensitivity, and specificity.

In this work, I used the StyleGAN2-ADA official source code, and the hyperparameters suggested by Karras et al. (KARRAS *et al.*, 2020). Concerning StyleGAN2-ADA hyperparameters, I used a batch size of 12, ADA target equal to 0.8 (i.e., the probability of using ADA mechanism), the Adam algorithm with a learning rate of 0.0025, decay values of 0 and 0.99, and a convergence error of  $1e^{-8}$  for the generator and discriminator. StyleGAN2-ADA framework enables different augmentations (rotation, geometric transformations, and color transformations) and class-conditional training. The output image resolution was set to  $256 \times 256$  pixels.

## 5.2.2 Results

The experimental results are presented in four sub-sections: (i) FID for synthetic image assessment, (ii) human experts detecting synthetic images, (iii) data augmentation assessment, (iv) comparison between human experts and deep models for detecting AMD images, and (v) web-based tool to access and validate the deep learning method model.

### 5.2.2.1 Synthetic Image Assessment

Table 39 presents the FID values for each GAN-based architecture. StyleGAN2-ADA achieved the lowest FID value of 166, while EBGAN was placed in last and its FID value of 380 was the highest. The smaller the FID value, the better is the quality of the generated image. Therefore, all further experiments only considered the StyleGAN2-ADA architecture for synthetic image generation.

Figure 33 shows a comparison between (a) real images from the training dataset and (b) synthetic images produced by StyleGAN2-ADA. The trained model yields realistic-looking images for both, with and without AMD, conditioned by sampling from latent representations. Visual inspection shows that the generated images are similar to the real images. In the AMD images, macula degeneration is evident.

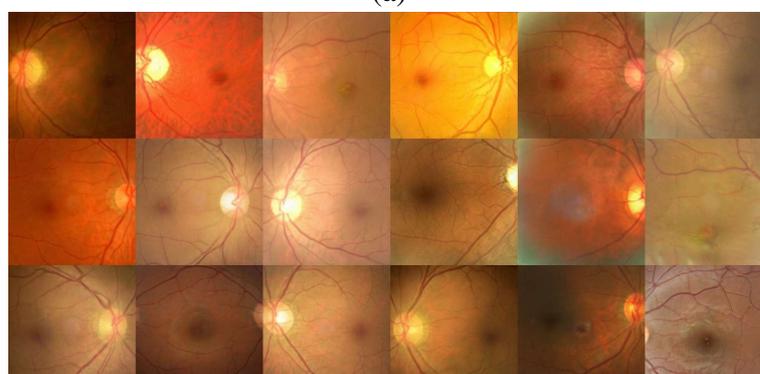
Figure 34 provides examples of real and synthetic images that are from eyes, positive and negative to AMD. One can observe the high-quality images that were generated for both, AMD and non-AMD images.

Table 39 – Mean FID values for image quality assessment (the best result is highlighted in bold).

Type	Architecture	FID
Unconditional	EBGAN	380.18
	DCGAN	326.85
	DRAGAN	317.82
	WGAN	307.00
	LSGAN	305.59
	WGAN-GP	295.23
	BEGAN	225.89
Conditional	CGAN	342.59
	ACGAN	315.36
	<b>StyleGAN2-ADA</b>	<b>166.17</b>



(a)



(b)

Figure 33 – Examples of (a) real retina images extracted from the training dataset, and (b) synthetic images generated by StyleGAN2-ADA.

### 5.2.2.2 Distinguishing between Synthetic and Real Images

Table 40 presents the outcomes of each clinical expert. For AMD images, the accuracy was 50% (standard deviation of 21.91%) for clinician #1 and 55% (standard deviation of 21.80%) for clinician #2. For Non-AMD images, clinician #1 achieved an accuracy of 60% (standard deviation of 21.47%), and clinician #2 obtained an accuracy of 50% (standard deviation of 21.47%). These results highlight that both clinicians could not differentiate between real and synthetic images.

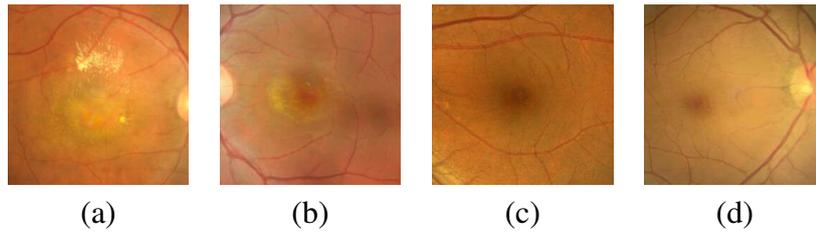


Figure 34 – Examples of synthetic and real images for AMD and Non\_AMD. (a) real, positive AMD, (b) synthetic, positive AMD, (c) real, Non-AMD and (d) synthetic, non-AMD.

Table 40 – Synthetic versus real images by humans experts.

		ACC	Sensitivity	Specificity
AMD	Clinician #1	0.50	0.50	0.50
	Clinician #2	0.55	0.40	0.70
Non-AMD	Clinician #1	0.60	0.60	0.60
	Clinician #2	0.50	0.40	0.60

### 5.2.2.3 Data Augmentation Assessment

In this study I performed on-the-fly data augmentation during training. Figure 35 shows the accuracy over the test set concerning different percentages ( $p$  value) of real images that were replaced by synthetic images. Overall, the accuracy improved when combining synthetic and real images. While the accuracy lies between 50% and 55% when using only synthetic images, and between 78% and 81% when using only real images for training, the combination of both types of images gave the best results. However, this was network dependent: while SqueezeNet accuracy peaked at 81% when using 70% of synthetic images, AlexNet obtained its highest accuracy (82%) when using only 20% of synthetic images. ResNet18 achieved its best of 83% with 60% of synthetic images. In general, the networks performed poorly when the percentage of synthetic images exceeded 70%.

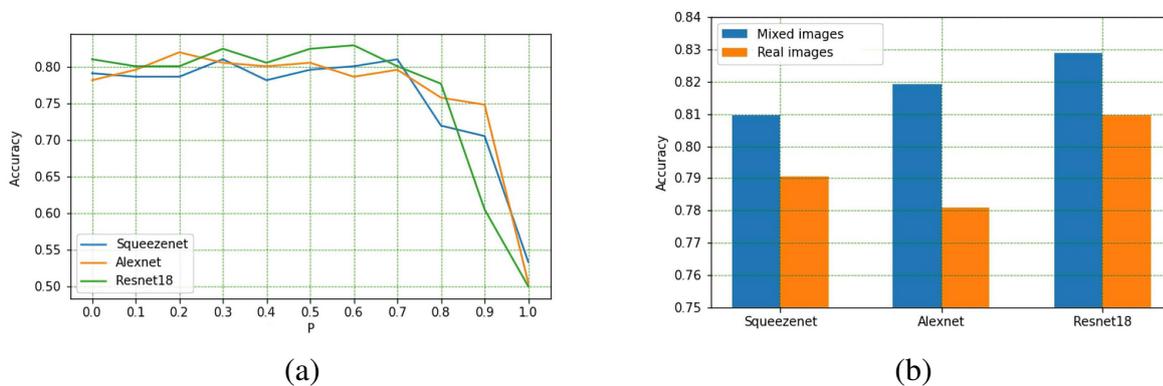


Figure 35 – Accuracy over the test set for different percentages of synthetic image for augmentation purposes (a). Accuracy over the test set concerning ResNet18, AlexNet, and SqueezeNet architectures (b).

Figure 35(b) shows the accuracy over the test set using only real ( $p = 0$ ) and also mixed data concerning ResNet18, AlexNet, and SqueezeNet architectures. The most significant

improvement was by AlexNet, its accuracy increased by approximately 8%, while ResNet18 had highest accuracy (about 83%) using a mix of synthetic and real images.

#### 5.2.2.4 Comparison between Human Experts and Deep Models

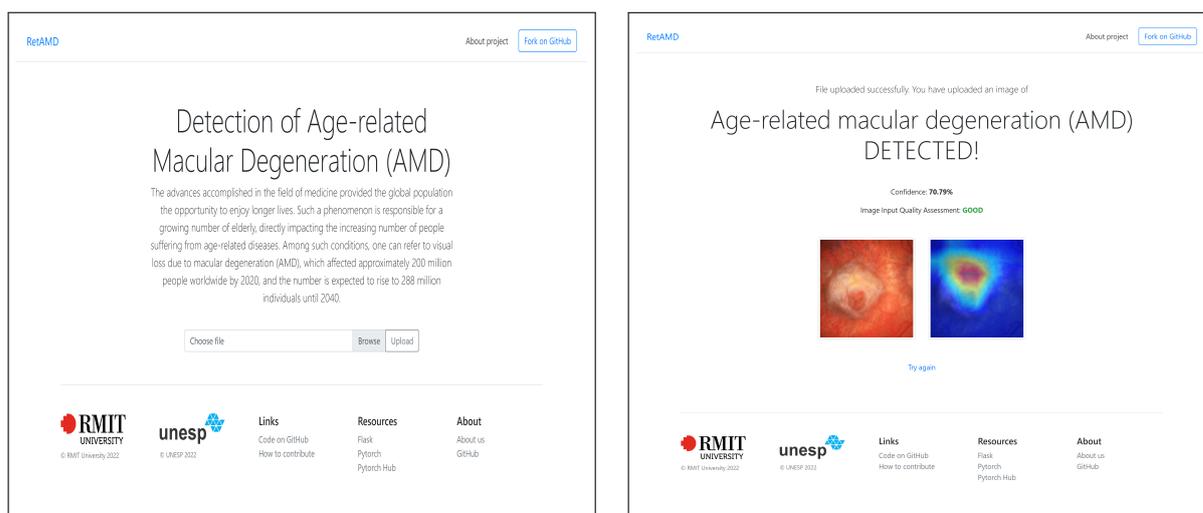
Table 41 presents the comparison of AMD detection by human experts, SqueezeNet, AlexNet and Resnet-18. Overall, the results of both clinicians and deep-learning were similar, with the best performance by deep-learning while the lowest specificity was by clinician #2. This shows that deep models are at least as good, and may outperform clinicians for diagnosing AMD in eye fundus images.

Table 41 – Comparison between human experts and deep models to classify AMD and real Non-AMD images.

	ACC	Sensitivity	Specificity
<b>Clinician #1</b>	0.80	0.80	0.80
<b>Clinician #2</b>	0.75	1.00	0.50
<b>SqueezeNet</b>	0.80	0.80	0.80
<b>AlexNet</b>	0.80	0.70	0.90
<b>Resnet-18</b>	0.85	0.90	0.80

#### 5.2.2.5 Web Application

The web application was designed to be easy to use and accessed by any user. The application's home page is minimalist and features a title along with a brief description of its functionality and a button with which the users can upload an image using the "browse" button (Figure 36(a)).



(a) RetAMD home page.

(b) Interface showing positive to AMD.

Figure 36 – Screenshots.

Once an image file is selected, the system automatically uploads the image to the cloud, performs the analysis in approximately ten to third seconds (largely based on the internet

conditions) displays the output in the form of a figure. The output also informs the user of the success of uploading the file and gives it the label, i.e., AMD detected or not. Additionally, the results also state the level of confidence in the diagnosis and an assessment of the quality of the input image into three quality grades: “Good”, “Usable” and “Reject”.

One of the main issues regarding deep learning in medicine is the difficulty to interpret the decision mechanisms. On the result screen, users can view two images (Figure 36(b)). The uploaded image is displayed on the left and a heatmap created by GradCAM (SELVARAJU *et al.*, 2017) is displayed on the right. This heatmap highlights the regions that were used in the diagnosis and allows the clinical users to understand the basis of the system’s decision. Overall, the results demonstrate the effectiveness and usability of RetAMD in diagnosing AMD.

### 5.2.2.6 Model Generalization

The validation of the ResNet18 architecture trained with mixed data ( $p = 0.6$ ) are shown in Table 42 provides the outcomes. The specificity, sensitivity, and accuracy were 80%, 85.7%, and 82.8%, respectively. To test this for generalisability, this deep learning trained model was validated using STARE, a database that was not used during the training phase. 80 images were randomly sampled from STARE; 40 AMD and 40 non-AMD images.

	AMD	Non-AMD
AMD	90	15
Non-AMD	21	84
Specificity	80%	
Sensitivity	85.71%	
Accuracy	82.86%	

Table 42 – Confusion matrix in my test set.

	AMD	Non-AMD
AMD	30	10
Non-AMD	5	35
Specificity	87.5%	
Sensitivity	75.0%	
Accuracy	81.25%	

Table 43 – Confusion matrix in STARE.

### 5.2.3 Discussion

Burlina et al have successfully developed a method for (BURLINA *et al.*, 2019; BURLINA *et al.*, 2018) generating the synthetic images. Their method is based on GAN, and they tested their method by showing that experts were unable to identify the synthetic images. However, their method has been patented and hence not available for being used by others. I have introduced an approach that integrates a deep learning quality assessment model and StyleGAN2-ADA, an extension of the progressive GAN. This approach filters out poor-quality retinal images and generates synthetic medical images that human experts are unable to distinguish from real ones. I have also shown that these images were suitable for increase the performance of deep-learning networks. The results were similar to those reported in the literature using patented technology. While Burlina et al. trained a Progressive GAN over a large number of images positive to AMD, my technique has the potential to generate similar high-quality synthetic images with only a small number of images.

My method has the potential to enhance the approach outlined by Burlina (BURLINA *et al.*, 2019; BURLINA *et al.*, 2018). Assessing image quality could lead to the exclusion of poor-quality images. Additionally, implementing on-the-fly augmentation and selecting the appropriate percentage of synthetic images have shown to be fundamental aspects.

ResNet18 architecture trained over real and synthetic images provided the best results, marginally outperforming the human experts' performance. While works of Anh *et al.* (AHN; SONG; SHIN, 2023) report significantly better results, these are all from the same database, with similar equipment used for all the images, images having similar quality and people from the similar ethnic group, which however does not represent the real-world situation. My work has shown the potential of using deep learning over multiple datasets where there are differences in people, equipment and image quality.

One limitation of my study is that it has only handled the binary problem, i.e., AMD versus non-AMD images. However, medical images are often non-binary with more classes. While this study and Burlina *et al.* (BURLINA *et al.*, 2019) utilized the expertise of two clinical professionals for image evaluation, it's important to acknowledge that having more experts such as three experts can make the study stronger and tests for the generalisability of the model. Further, three experts permits the disagreements being resolved by voting which tests the strength of the study.

Another limitation of this study is that it used a single-source dataset with 80 images for validation purposes. However, this may be not sufficient to test the model's generalisability.

Different studies have presented results reaching more than 90% accuracy (PEAD *et al.*, 2019; LENG *et al.*, 2023). However, the aim of this study was not to get the best results but for the model to be not limited to any one database. Thus, the focus was on training on a mixed dataset to increase generalization and test this using a different dataset. Another objective was also to develop a web-based tool that provides a model capable of identifying AMD in a wide variety of images from different sources. My model works on standard CPU computers, which makes it suitable for inexpensive deployment. One limitation in my work is that I have used traditional architectures and not used the number of layers reported nor some recent advances in the state-of-the-art techniques such as used by Anh *et al.* (AHN; SONG; SHIN, 2023), Grassmann *et al.* (GRASSMANN *et al.*, 2018), Govindaiah *et al.* (GOVINDAIAH; SMITH; BHUIYAN, 2018), Keel *et al.* (KEEL *et al.*, 2019), Bhuiyan *et al.* (BHUIYAN *et al.*, 2020). This decision was influenced by a common drawback of state-of-the-art convolutional neural networks, which is their complexity. They were typically developed to be trained in extensive datasets, resulting in overfitting when applied to small datasets. However, this study focused on developing a system suitable for being used for web-based applications.

I have made available an online system with this trained network so that anyone can use it and test it, simply by uploading images. The software automatically labels the images as positive or negative to AMD. I have also provided the source code of the entire software and it

is available publicly to facilitate researchers to use this as it is, or improve it. I am focused on fostering partnerships to facilitate and conduct research towards the usage of deep-learning to generate and recognize medical images.

Moreover, diffusion models (CROITORU *et al.*, 2023) have recently emerged as a promising alternative for generating high-quality and realistic synthetic images due to their ability to effectively capture complex patterns and textures. Unlike GAN-based approaches, diffusion models iteratively refine images by progressively removing noise, resulting in detailed and diverse outputs. Future work should include evaluating the performance and generalizability of diffusion models, specifically in generating synthetic eye-fundus images for AMD diagnosis, and comparing their effectiveness against the StyleGAN2-ADA method presented in this study. Such comparisons will help determine whether diffusion models can further enhance synthetic image quality and improve the accuracy of deep learning models in clinical ophthalmology.

#### 5.2.3.1 Novelty of this study

This study uniquely integrates a retinal image quality assessment model with StyleGAN2-ADA, enhancing the generation of high-quality synthetic fundus images for AMD detection. Unlike prior approaches, my method systematically filters low-quality images before synthetic augmentation, ensuring more reliable deep learning training. A comparative analysis of ten GAN architectures confirmed StyleGAN2-ADA as the most effective, producing synthetic images indistinguishable from real ones by clinical experts. This refinement addresses key limitations in existing work, where synthetic image quality and practical applicability have often been overlooked.

Furthermore, by training on multiple datasets with diverse demographics and imaging conditions, I demonstrate the superior generalizability of my approach. My findings highlight the optimal balance of synthetic and real data for deep learning, improving model performance beyond human expert accuracy. This study advances synthetic data applications in medical imaging and establishes a scalable framework for broader clinical AI integration.

#### 5.2.4 Conclusion

I have employed a retinal image quality assessment model in preprocessing step. I have compared a number of synthetic medical image generation techniques and found StyleGAN2-ADA to be the most suitable using which I have developed a method to generate synthetic images. I have investigated the use of the synthetic images obtained using examples from publicly available databases to train the model for distinguishing between AMD and healthy eyes. I found that experienced clinical experts were unable to differentiate between synthetic and real images. I have tested the model for generalisability by training the model using images from three databases and validated it using a fourth database. I also have demonstrated that the classification

accuracy of deep learning networks marginally outperformed clinical experts in separating the AMD and Non-AMD retinal images.

I have made the source code for generating the synthetic images publicly available to facilitate joint research in the field. I have also provided free access through this chapter for the online use of the AMD detection model. This will facilitate future work to broaden the scope for detecting the severity of AMD, and for differentiating from other diseases. For generating synthetic medical images, there is the need to consider a broader range of deep architectures and the effectiveness of heatmaps helping the clinicians.

### 5.2.5 Data availability

The iChallenge-AMD dataset can be found in <https://ai.baidu.com/broad/introduction?dataset=amd>, while ODIR-2019 dataset is available on <https://odir2019.grand-challenge.org/dataset/> and RIADD is available on <https://riadd.grand-challenge.org/Home/>.

### 5.2.6 Code availability

The official code for StyleGAN2-ADA is available at <https://github.com/NVlabs/stylegan2-ada-pytorch>. My implementation of Deep Convolutional GAN (DCGAN), Least Squares Generative Adversarial Networks (LSGAN), Wasserstein GAN (WGAN), Wasserstein GAN with Gradient Penalty (WGAN-GP), Deep Regret Analytic Generative Adversarial Networks (DRAGAN), Energy-based Generative Adversarial Network (EBGAN), Boundary Equilibrium Generative Adversarial Networks (BEGAN), Conditional GAN (CGAN), and Auxiliary classifier GAN (ACGAN) are available for download at <https://github.com/GuiCamargoX/gans-pytorch>. All source code concerning image processing, Style-GAN2-ADA data generation, and the pre-trained networks are available at <https://github.com/GuiCamargoX/synthetic-retina-aml>. The web application can be found in <https://amdfundus.space/>. The official Flask code of is available for download in <https://github.com/GuiCamargoX/RetAMD>.

### 5.2.7 CRediT authorship contribution statement

I am grateful for the support and collaboration that made this work possible. As the lead investigator, I was responsible for the research, design, methodology development, software development and writing of this study. I acknowledge Gustavo H. Rosa for his contributions in supervision, project administration, investigation, conceptualization, and original draft writing; Daniel C.G. Pedronette for his role in project administration, investigation, conceptualization, validation, and original draft writing; João P. Papa for his supervision, project administration, investigation, conceptualization, validation, and contributions to both the original draft and review; Himeesh Kumar for his work in data curation, formal analysis, validation, visualization,

supervision, and investigation; Leandro A. Passos for his input on methodology, investigation, validation, visualization, and original draft writing; and Dinesh Kumar for his contributions to conceptualization, methodology, investigation, supervision, validation, visualization, and writing of the original draft and review. Their collective efforts were vital to the success of this study.

---

## CONCLUSION

---

---

This thesis explored the potential of AI-assisted tools to enhance medical diagnostics across three domains: video-based facial expression analysis, voice-based assessment, and AI-powered synthetic imaging. These investigations were guided by the following research questions:

1. **(Video)** How can AI-assisted facial expression analysis enhance the detection and understanding of neurological conditions such as Parkinson's disease, stroke, and amyotrophic lateral sclerosis?
2. **(Voice)** In what ways can AI-based voice analysis tools improve the remote assessment of Parkinson's disease severity and support ongoing monitoring?
3. **(Image)** How do AI-powered synthetic imaging techniques contribute to the detection and diagnosis of medical conditions like age-related macular degeneration and venous leg ulcers?

The following sections summarize how each research question was addressed through the findings of this thesis.

### 6.1 Video Analysis for Neurological Conditions

In response to RQ1, this thesis demonstrates that AI-assisted facial expression analysis can identify hypomimia in Parkinson's disease (Section 3.1) and detect and enhance the understanding of facial weakness in post-stroke and ALS patients (Sections 3.2 and 3.3), for example, using action units. Collectively, these findings show that, while not yet diagnostic on their own, such tools can serve as non-invasive first-line screening aids and support clinical assessments, especially in resource-limited or emergency settings.

- **Section 2.1.3: Systematic Literature Review of Hypomimia in Parkinson’s Disease.** A systematic review from 2019 to 2024 revealed that AI models, particularly machine learning techniques, have made strides in identifying PD-specific facial changes. Facial expressions, as potential biomarkers for Parkinson’s disease, have shown promise, but clinical implementation is hindered by limited dataset diversity and a lack of generalization across different demographics.
- **Section 3.1: Facial Action Unit Analysis for PD Hypomimia Detection.** The study showed that AI-assisted facial analysis could effectively detect hypomimia, a symptom of Parkinson’s disease. However, the occurrence of false positives indicates that while the technology is useful for population screening and clinical support, it is not yet suitable for definitive diagnosis. Additionally, research emphasizes the need to test the tool across different ethnic groups to enhance accuracy.
- **Section 3.2: Facial Action Unit Analysis for Post-Stroke and Amyotrophic Lateral Sclerosis.** Pilot studies indicated that AI-based facial expression analysis has the potential to detect facial weakness in post-stroke and amyotrophic lateral sclerosis patients. This method shows promise for initial assessments, particularly by first responders using portable devices, although it requires further validation in real-world conditions to ensure reliability.
- **Section 3.3: Graph Neural Networks for Post-Stroke and Amyotrophic Lateral Sclerosis.** A novel deep learning framework using Graph Neural Networks was introduced to detect orofacial impairments associated with post-stroke and ALS conditions. Unlike conventional approaches that depend on handcrafted features, this method automatically learns geometric and motion features from facial landmarks represented as graphs. While promising, further research is required to incorporate temporal dynamics and expand the dataset for broader clinical applicability. A key drawback is the difficulty in interpreting the outcomes; in comparison, action units (Section 3.2) are easier to interpret because they provide a direct mapping to specific facial muscle movements, and have demonstrated better or similar performance.

## 6.2 Voice Analysis for Remote Monitoring

In answer to RQ 2, the thesis demonstrated that AI-based voice analysis tools can enhance the assessment of Parkinson’s disease severity and facilitate remote monitoring. Across DDK task analysis (Section 4.1) and formant-based vocal tract measurements (Section 4.2), AI models extracted vocal biomarkers that correlate with PD. Moreover, integrating large language models into conversational agents (Section 4.3) enables scalable, user-guided voice assessments.

- **Section 4.1: Diadochokinetic (DDK) Task Analysis.** This study established that AI could accurately categorize Parkinson's disease symptoms through voice analysis, using DDK tasks to extract relevant speech features. Although the study was based on a limited dataset, the results suggest that vocal biomarkers are a viable tool for assessing disease severity, encouraging further exploration with larger datasets.
- **Section 4.2: Apparent Vocal Tract.** This study measured the apparent vocal tract length using phoneme-based formant analysis to compare Parkinson's patients with healthy controls. Results indicate an increased vocal tract length in patients, particularly among males, underscoring its potential as a remote monitoring biomarker.
- **Section 4.3: Integration of Large Language Models and Chatbots.** The use of AI-driven chatbots in conducting voice assessments illustrated a significant advancement in remote monitoring. By guiding users through specific phoneme-based vocal tests, these tools can make early Parkinson's screening more accessible, particularly in underserved regions. The technology offers a scalable, cost-effective solution that improves early intervention and patient care.

## 6.3 AI-Powered Synthetic Imaging

In answer to RQ 3, the thesis confirmed that AI-powered deep learning techniques could improve the detection and diagnosis of medical conditions by generating high-quality synthetic images that enhance medical datasets. StyleGAN-2 augmentation (Section 5.1) improved diagnostic model performance for age-related macular degeneration, while Stable Diffusion-based thermal reconstruction (Section 5.2) achieved an SSIM of 0.84 for leg-ulcer imagery.

- **Section 5.1: Data Augmentation with StyleGAN-2 for Ophthalmology.** The validation of synthetic image generation using StyleGAN-2 showed that this approach could robustly augment datasets for ophthalmology, enhancing diagnostic accuracy for conditions like age-related macular degeneration. This AI-driven technique allows for the development of more precise diagnostic models and supports further research through the open-source availability of the code.
- **Section 5.2: Thermal Imaging Using Stable Diffusion Techniques.** This study demonstrated that AI-based diffusion techniques could generate thermal images from RGB data, offering an affordable and accessible alternative to conventional thermal imaging. Preliminary results, with an SSIM score of 0.84, suggest the potential utility of these images for assessing chronic wounds like venous leg ulcers. However, additional validation is needed to confirm their effectiveness in clinical assessments.

## 6.4 Summary of Innovations and Work Implications

This thesis successfully addressed the research questions by demonstrating the potential of AI to revolutionize medical diagnostics:

- **AI-assisted facial expression analysis** has proven effective for detection of neurological conditions like Parkinson's disease, stroke, and amyotrophic lateral sclerosis. While these tools are not yet ready for standalone diagnosis, they are highly valuable for screening, clinical assistance and monitoring the loss of facial expression. Future work must focus on improving model generalization and reducing demographic biases. This approach offers a groundbreaking non-invasive way to identify subtle symptoms that might otherwise go unnoticed. The stroke app can assist in screening cases with just a smile in emergency departments.
- **AI-based voice analysis** has shown promise for non-invasive, remote monitoring of Parkinson's disease. Vocal biomarkers provide useful indicators of disease severity, and AI-powered chatbots make these tools accessible in resource-limited settings, facilitating early intervention and better disease management. This novelty allows for continuous monitoring without the need for in-person visits.
- **AI-powered synthetic imaging** techniques like StyleGAN-2 and stable diffusion models enhance the quality and accessibility of medical datasets. These methods contribute to more inclusive diagnostic models, addressing traditional data limitations and allowing for more accurate and accessible diagnostics. To my knowledge, this is the first attempt to reconstruct thermal images from RGB images of the leg ulcers.

### 6.4.1 Ethical Implications, Challenges, and Risks

While AI-assisted diagnostic systems present transformative potential for healthcare, their deployment also introduces significant ethical implications and practical challenges. The reliance on extensive patient data raises concerns about privacy and security breaches, emphasizing the need for safeguards. Additionally, biases in AI models, often arising from insufficiently diverse training datasets, risk leading to unequal diagnostic outcomes across different demographics. Addressing these issues requires continuous efforts to improve dataset diversity and implement fairness evaluations to ensure ethical and equitable AI deployment.

Transparency and accountability remain critical concerns, as many AI models function as "black boxes", making their decision-making processes challenging to interpret. Establishing explainable AI methodologies and clear accountability frameworks is essential to fostering trust among healthcare providers and patients. Furthermore, regulatory compliance and ethical oversight must evolve alongside AI advancements to define responsibility in cases of misdiagnosis or harm. AI should serve as an assistive tool rather than replace clinical judgment. Proactively

addressing these challenges will enable AI's responsible and effective integration into medical practice.

## 6.5 Future Work

While this thesis has made significant strides, further research is required in several key areas to fully realize the potential of AI in healthcare. These areas can be broadly categorized as follows: (1) Data, (2) Software Development, (3) Real-world Validation, and (4) Data Security.

1. **Data:** Expanding and diversifying datasets is crucial for enhancing the generalizability of AI models. This is particularly important in video and voice analysis, where diverse datasets help mitigate biases and ensure that AI tools are more inclusive and effective. In addition, conducting clinical trials with larger and more varied sample sizes is essential for validating AI-based tools, particularly in fields like facial and voice analysis. Larger datasets will not only improve model accuracy but also increase the reliability and robustness of diagnostic assessments.
2. **Software Development:** Integrating video, voice, and imaging data presents an opportunity to create a more holistic view of complex diseases, such as Parkinson's and Amyotrophic Lateral Sclerosis. Combining these data sources could lead to more accurate diagnoses, earlier detection of symptoms, and personalized treatment plans. Enhancing AI algorithms to handle multimodal data will be a key step in improving diagnostic performance.
3. **Real-world Validation:** Bridging the gap between research settings and clinical practice requires thorough testing of AI models in real-world environments. This step is critical to ensure that models perform consistently across diverse clinical scenarios, patient populations, and healthcare systems. Real-world validation will also highlight practical challenges and areas for improvement, making AI tools more reliable and user-friendly for clinicians.
4. **Data Security:** Addressing ethical concerns and ensuring robust data privacy are fundamental to building trust in AI-based healthcare solutions. It is crucial to develop secure frameworks that protect patient data while maintaining transparency in AI decision-making processes. Minimizing biases, adhering to ethical guidelines, and creating clear audit trails for AI decisions will be essential for fostering confidence among clinicians and patients.

# BIBLIOGRAPHY

---

---

ABBAS, A.; YADAV, V.; SMITH, E.; RAMJAS, E.; RUTTER, S. B.; BENAVIDEZ, C.; KOES-MAHARGYO, V.; ZHANG, L.; GUAN, L.; ROSENFELD, P. *et al.* Computer vision-based assessment of motor functioning in schizophrenia: use of smartphones for remote measurement of schizophrenia symptomatology. **Digital biomarkers**, S. Karger AG, v. 5, n. 1, p. 29–36, 2021. Citations on pages [82](#) and [84](#).

ABBASIAN, M.; AZIMI, I.; RAHMANI, A. M.; JAIN, R. Conversational health agents: A personalized llm-powered agent framework. **arXiv preprint arXiv:2310.02374**, 2023. Citation on page [114](#).

ABRAHAM, M. S.; AREVALO, G. G.; GARCIA, S. S.; MIZRAJI, G.; CHADE, A.; GER-SHANIK, O. Decreased palpebral fissure in patients with parkinson’s disease. **Movement Disorders Clinical Practice**, Wiley Online Library, v. 4, n. 1, p. 58–61, 2017. Citation on page [36](#).

ABRAMI, A.; GUNZLER, S.; KILBANE, C.; OSTRAND, R.; HO, B.; CECCHI, G. Automated computer vision assessment of hypomimia in parkinson disease: proof-of-principle pilot study. **Journal of Medical Internet Research**, Gunther Eysenbach MD MPH, Associate Professor, v. 23, n. 2, p. e21037, 2021. Citations on pages [39](#), [44](#), [46](#), and [49](#).

ADADI, A.; BERRADA, M. Peeking inside the black-box: a survey on explainable artificial intelligence (xai). **IEEE access**, IEEE, v. 6, p. 52138–52160, 2018. Citation on page [52](#).

AGGARWAL, A.; TAM, C. C.; WU, D.; LI, X.; QIAO, S. Artificial intelligence–based chatbots for promoting health behavioral changes: Systematic review. **Journal of Medical Internet Research**, JMIR Publications Toronto, Canada, v. 25, p. e40789, 2023. Citation on page [114](#).

AHN, S.; SONG, S. J.; SHIN, J. Fundusgan: Fundus image synthesis based on semi-supervised learning. **Biomedical Signal Processing and Control**, Elsevier, v. 86, p. 105289, 2023. Citations on pages [59](#) and [141](#).

AHO-ÖZHAN, H. E.; KELLER, J.; HEIMRATH, J.; UTTNER, I.; KASSUBEK, J.; BIRBAUMER, N.; LUDOLPH, A. C.; LULÉ, D. Perception of emotional facial expressions in amyotrophic lateral sclerosis (als) at behavioural and brain metabolic level. **PLoS One**, Public Library of Science San Francisco, CA USA, v. 11, n. 10, p. e0164655, 2016. Citations on pages [30](#) and [74](#).

AHONEN, T.; HADID, A.; PIETIKÄINEN, M. Face recognition with local binary patterns. In: SPRINGER. **European conference on computer vision**. [S.l.], 2004. p. 469–481. Citation on page [39](#).

AHONEN, T.; HADID, A.; PIETIKAINEN, M. Face description with local binary patterns: Application to face recognition. **IEEE transactions on pattern analysis and machine intelligence**, IEEE, v. 28, n. 12, p. 2037–2041, 2006. Citation on page [43](#).

AKBAR, U.; MCQUEEN, R. B.; BEMSKI, J.; CARTER, J.; GOY, E. R.; KUTNER, J.; JOHNSON, M. J.; MIYASAKI, J. M.; KLUGER, B. Prognostic predictors relevant to end-of-life palliative care in parkinson's disease and related disorders: A systematic review. **Journal of Neurology, Neurosurgery & Psychiatry**, BMJ Publishing Group Ltd, 2021. Citation on page 114.

AKBARI, H.; YUAN, L.; QIAN, R.; CHUANG, W.-H.; CHANG, S.-F.; CUI, Y.; GONG, B. Vatt: Transformers for multimodal self-supervised learning from raw video, audio and text. **Advances in Neural Information Processing Systems**, v. 34, 2021. Citation on page 52.

ALI, M. R.; MYERS, T.; WAGNER, E.; RATNU, H.; DORSEY, E.; HOQUE, E. Facial expressions can detect parkinson's disease: preliminary evidence from videos collected online. **NPJ Digital Medicine**, Nature Publishing Group, v. 4, n. 1, p. 1–4, 2021. Citations on pages 40, 44, 46, 61, 62, 65, 70, 71, 72, and 82.

ALI, S. M.; ARJUNAN, S. P.; PETERS, J.; PERJU-DUMBRAVA, L.; DING, C.; ELLER, M.; RAGHAV, S.; KEMPSTER, P.; MOTIN, M. A.; RADCLIFFE, P. *et al.* Wearable sensors during drawing tasks to measure the severity of essential tremor. **Scientific Reports**, Nature Publishing Group UK London, v. 12, n. 1, p. 5242, 2022. Citation on page 97.

ALMUTIRY, R.; COUTH, S.; POLIAKOFF, E.; KOTZ, S.; SILVERDALE, M.; COOTES, T. Facial behaviour analysis in parkinson's disease. In: SPRINGER. **Medical Imaging and Augmented Reality: 7th International Conference, MIAR 2016, Bern, Switzerland, August 24-26, 2016, Proceedings 7**. [S.l.], 2016. p. 329–339. Citation on page 35.

ALSHARIF, I.; RUSCELLI, A. L.; CECCHETTI, G.; SAHU, S. K.; GHARBAOUI, M.; ORLANDI, G.; CASTOLDI, P. Face recognition to support pre-hospital stroke diagnosis. In: IEEE. **2024 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops)**. [S.l.], 2024. p. 227–232. Citation on page 32.

ALTMAN, N. S. An introduction to kernel and nearest-neighbor nonparametric regression. **The American Statistician**, Taylor & Francis, v. 46, n. 3, p. 175–185, 1992. Citations on pages 63 and 64.

ANAND, S. Speed of pitch change in people with parkinson's disease: A pilot study. **Revista de investigación e innovación en ciencias de la salud**, Fundación Universitaria María Cano, v. 5, n. 2, p. 149–163, 2023. Citation on page 56.

ANTONIOU, A.; STORKEY, A.; EDWARDS, H. Data augmentation generative adversarial networks. **arXiv preprint arXiv:1711.04340**, 2017. Citation on page 61.

ARCH, A. E.; WEISMAN, D. C.; COCA, S.; NYSTROM, K. V.; III, C. R. W.; SCHINDLER, J. L. Missed ischemic stroke diagnosis in the emergency department by emergency medicine and neurology services. **Stroke**, Am Heart Assoc, v. 47, n. 3, p. 668–673, 2016. Citation on page 74.

ARIAS-LONDOÑO, J. D.; GÓMEZ-GARCÍA, J. A. Predicting updrs scores in parkinson's disease using voice signals: A deep learning/transfer-learning-based approach. In: SPRINGER. **Automatic Assessment of Parkinsonian Speech: First Workshop, AAPS 2019, Cambridge, Massachusetts, USA, September 20–21, 2019, Revised Selected Papers 1**. [S.l.], 2020. p. 100–123. Citations on pages 97, 102, and 106.

ARIAS, T. R.; CORTÉS, I. R.; OLMO, A. Pérez del. Biomechanical parameters of voice in parkinson's disease patients. **Folia Phoniatica et Logopaedica**, S. Karger AG, v. 76, n. 1, p. 91–101, 2024. Citation on page [57](#).

ARIAS-VERGARA, T.; VÁSQUEZ-CORREA, J. C.; OROZCO-ARROYAVE, J. R. Parkinson's disease and aging: analysis of their effect in phonation and articulation of speech. **Cognitive Computation**, Springer, v. 9, p. 731–748, 2017. Citation on page [100](#).

ARIAS-VERGARA, T.; VÁSQUEZ-CORREA, J. C.; OROZCO-ARROYAVE, J. R.; KLUMPP, P.; NÖTH, E. Unobtrusive monitoring of speech impairments of parkinson's disease patients through mobile devices. In: IEEE. **2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)**. [S.l.], 2018. p. 6004–6008. Citations on pages [97](#), [102](#), and [106](#).

ARJOVSKY, M.; CHINTALA, S.; BOTTOU, L. Wasserstein generative adversarial networks. In: PMLR. **International conference on machine learning**. [S.l.], 2017. p. 214–223. Citation on page [135](#).

ARKIN, R. C.; SCHEUTZ, M.; TICKLE-DEGNEN, L. Preserving dignity in patient caregiver relationships using moral emotions and robots. In: IEEE. **2014 IEEE International Symposium on Ethics in Science, Technology and Engineering**. [S.l.], 2014. p. 1–5. Citation on page [35](#).

ARORA, S.; VENKATARAMAN, V.; ZHAN, A.; DONOHUE, S.; BIGLAN, K. M.; DORSEY, E. R.; LITTLE, M. A. Detecting and monitoring the symptoms of parkinson's disease using smartphones: A pilot study. **Parkinsonism & related disorders**, Elsevier, v. 21, n. 6, p. 650–653, 2015. Citation on page [20](#).

ARRIETA, A. B.; DÍAZ-RODRÍGUEZ, N.; SER, J. D.; BENNETOT, A.; TABIK, S.; BARBADO, A.; GARCÍA, S.; GIL-LÓPEZ, S.; MOLINA, D.; BENJAMINS, R. *et al.* Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai. **Information fusion**, Elsevier, v. 58, p. 82–115, 2020. Citation on page [52](#).

ARTHUR, K. C.; CALVO, A.; PRICE, T. R.; GEIGER, J. T.; CHIO, A.; TRAYNOR, B. J. Projected increase in amyotrophic lateral sclerosis from 2015 to 2040. **Nature communications**, Nature Publishing Group UK London, v. 7, n. 1, p. 12408, 2016. Citation on page [88](#).

ATALAR, M. S.; OGUZ, O.; GENÇ, G. Hypokinetic dysarthria in parkinson's disease: A narrative review. **Medical Bulletin of Sisli Etfal Hospital**, v. 57, n. 2, 2023. Citation on page [56](#).

BALTRUŠAITIS, T.; ROBINSON, P.; MORENCY, L.-P. Openface: an open source facial behavior analysis toolkit. In: IEEE. **2016 IEEE winter conference on applications of computer vision (WACV)**. [S.l.], 2016. p. 1–10. Citation on page [90](#).

BANDINI, A.; GREEN, J.; RICHBURG, B.; YUNUSOVA, Y. Automatic Detection of Orofacial Impairment in Stroke. In: **Proc. Interspeech 2018**. [S.l.: s.n.], 2018. p. 1711–1715. ISSN 2958-1796. Citation on page [32](#).

BANDINI, A.; GREEN, J. R.; RICHBURG, B.; YUNUSOVA, Y. Automatic detection of orofacial impairment in stroke. In: **Interspeech**. [S.l.: s.n.], 2018. p. 1711–1715. Citations on pages [82](#) and [85](#).

BANDINI, A.; GREEN, J. R.; TAATI, B.; ORLANDI, S.; ZINMAN, L.; YUNUSOVA, Y. Automatic detection of amyotrophic lateral sclerosis (als) from video-based analysis of facial movements: speech and non-speech tasks. In: IEEE. **2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)**. [S.l.], 2018. p. 150–157. Citations on pages [30](#), [82](#), [83](#), and [84](#).

BANDINI, A.; ORLANDI, S.; ESCALANTE, H. J.; GIOVANNELLI, F.; CINCOTTA, M.; REYES-GARCIA, C. A.; VANNI, P.; ZACCARA, G.; MANFREDI, C. Analysis of facial expressions in parkinson's disease through video-based automatic methods. **Journal of neuroscience methods**, Elsevier, v. 281, p. 7–20, 2017. Citations on pages [35](#), [36](#), and [61](#).

BANDINI, A.; REZAEI, S.; GUARÍN, D. L.; KULKARNI, M.; LIM, D.; BOULOS, M. I.; ZINMAN, L.; YUNUSOVA, Y.; TAATI, B. A new dataset for facial motion analysis in individuals with neurological disorders. **IEEE Journal of Biomedical and Health Informatics**, IEEE, v. 25, n. 4, p. 1111–1119, 2020. Citations on pages [75](#) and [88](#).

BANDINI, A.; REZAEI, S.; GUARÍN, D. L.; KULKARNI, M.; LIM, D.; BOULOS, M. I.; ZINMAN, L.; YUNUSOVA, Y.; TAATI, B. A new dataset for facial motion analysis in individuals with neurological disorders. **IEEE Journal of Biomedical and Health Informatics**, v. 25, n. 4, p. 1111–1119, 2021. Citations on pages [85](#) and [87](#).

BAT-ERDENE, B.-O.; SAVER, J. L. Automatic acute stroke symptom detection and emergency medical systems alerting by mobile health technologies: A review. **Journal of Stroke and Cerebrovascular Diseases**, Elsevier, v. 30, n. 7, p. 105826, 2021. Citation on page [74](#).

BAUGH, R. F.; BASURA, G. J.; ISHII, L. E.; SCHWARTZ, S. R.; DRUMHELLER, C. M.; BURKHOLDER, R.; DECKARD, N. A.; DAWSON, C.; DRISCOLL, C.; GILLESPIE, M. B. *et al.* Clinical practice guideline: Bell's palsy. **Otolaryngology–Head and Neck Surgery**, SAGE Publications Sage CA: Los Angeles, CA, v. 149, n. 3\_suppl, p. S1–S27, 2013. Citation on page [54](#).

BELLEMO, V.; BURLINA, P.; YONG, L.; WONG, T. Y.; TING, D. S. W. Generative adversarial networks (gans) for retinal fundus image synthesis. In: SPRINGER. **Computer Vision–ACCV 2018 Workshops: 14th Asian Conference on Computer Vision, Perth, Australia, December 2–6, 2018, Revised Selected Papers 14**. [S.l.], 2019. p. 289–302. Citation on page [59](#).

BERARDELLI, A.; WENNING, G.; ANTONINI, A.; BERG, D.; BLOEM, B.; BONIFATI, V.; BROOKS, D.; BURN, D.; COLOSIMO, C.; FANCIULLI, A. *et al.* Efn/nds-es recommendations for the diagnosis of parkinson's disease. **European Journal of Neurology**, Wiley Online Library, v. 20, n. 1, p. 16–34, 2013. Citation on page [32](#).

BERTHELOT, D.; SCHUMM, T.; METZ, L. Began: Boundary equilibrium generative adversarial networks. **arXiv preprint arXiv:1703.10717**, 2017. Citation on page [135](#).

BHUIYAN, A.; WONG, T. Y.; TING, D. S. W.; GOVINDAIAH, A.; SOUIED, E. H.; SMITH, R. T. Artificial intelligence to stratify severity of age-related macular degeneration (amd) and predict risk of progression to late amd. **Translational vision science & technology**, The Association for Research in Vision and Ophthalmology, v. 9, n. 2, p. 25–25, 2020. Citation on page [141](#).

BIASE, L. D.; SANTO, A. D.; CAMINITI, M. L.; LISO, A. D.; SHAH, S. A.; RICCI, L.; LAZZARO, V. D. Gait analysis in parkinson's disease: An overview of the most accurate

markers for diagnosis and symptoms monitoring. **Sensors**, MDPI, v. 20, n. 12, p. 3529, 2020. Citation on page 19.

BOCKLET, T.; STEIDL, S.; NÖTH, E.; SKODDA, S. Automatic evaluation of parkinson's speech-acoustic, prosodic and voice related cues. In: **Interspeech**. [S.l.: s.n.], 2013. p. 1149–1153. Citation on page 97.

BOERSMA, B. P.; HEUVEN, V. V. Speak and unSpeak with P RAAT. **Glott International**, v. 5, n. 9-10, p. 341–347, 2001. Citation on page 111.

BOLOGNA, M.; BERARDELLI, I.; PAPARELLA, G.; MARSILI, L.; RICCIARDI, L.; FABBRINI, G.; BERARDELLI, A. Altered kinematics of facial emotion expression and emotion recognition deficits are unrelated in parkinson's disease. **Frontiers in neurology**, Frontiers, v. 7, p. 230, 2016. Citation on page 35.

BORIS, N. Delaunay. sur la sphere vide. **Izvestia Akademia Nauk SSSR, VII Seria, Otdelenie Matematicheskii i Estestvennyka Nauk**, v. 7, p. 793–800, 1934. Citation on page 91.

BOWERS, D.; MILLER, K.; BOSCH, W.; GOKCAY, D.; PEDRAZA, O.; SPRINGER, U.; OKUN, M. Faces of emotion in parkinsons disease: micro-expressivity and bradykinesia during voluntary facial expressions. **Journal of the International Neuropsychological Society**, Cambridge University Press, v. 12, n. 6, p. 765–773, 2006. Citation on page 34.

BRAGA, D.; MADUREIRA, A. M.; COELHO, L.; AJITH, R. Automatic detection of Parkinson's disease based on acoustic analysis of speech. **Engineering Applications of Artificial Intelligence**, Elsevier Ltd, v. 77, p. 148–158, 2019. ISSN 0952-1976. Available: <<https://doi.org/10.1016/j.engappai.2018.09.018>>. Citation on page 110.

BRAKEDAL, B.; TOKER, L.; HAUGARVOLL, K.; TZOULIS, C. A nationwide study of the incidence, prevalence and mortality of parkinson's disease in the norwegian population. **npj Parkinson's Disease**, Nature Publishing Group UK London, v. 8, n. 1, p. 19, 2022. Citation on page 70.

BRANCO, P.; TORGO, L.; RIBEIRO, R. P. A survey of predictive modeling on imbalanced domains. **ACM Computing Surveys (CSUR)**, ACM New York, NY, USA, v. 49, n. 2, p. 1–50, 2016. Citation on page 63.

BRIGATO, L.; IOCCHI, L. A close look at deep learning with small data. In: IEEE. **2020 25th international conference on pattern recognition (ICPR)**. [S.l.], 2021. p. 2490–2497. Citation on page 59.

BRIGGS, P.; SCHEUTZ, M.; TICKLE-DEGNEN, L. Are robots ready for administering health status surveys' first results from an hri study with subjects with parkinson's disease. In: **Proceedings of the tenth annual acm/ieee international conference on human-robot interaction**. [S.l.: s.n.], 2015. p. 327–334. Citation on page 35.

BULAT, A.; TZIMIROPOULOS, G. How far are we from solving the 2d & 3d face alignment problem?(and a dataset of 230,000 3d facial landmarks). In: **Proceedings of the IEEE international conference on computer vision**. [S.l.: s.n.], 2017. p. 1021–1030. Citation on page 90.

BURLINA, P.; JOSHI, N.; PACHECO, K. D.; FREUND, D. E.; KONG, J.; BRESSLER, N. M. Utility of deep learning methods for referability classification of age-related macular degeneration. **JAMA ophthalmology**, American Medical Association, v. 136, n. 11, p. 1305–1307, 2018. Citations on pages 140 and 141.

BURLINA, P. M.; JOSHI, N.; PACHECO, K. D.; LIU, T. A.; BRESSLER, N. M. Assessment of deep generative models for high-resolution synthetic retinal image generation of age-related macular degeneration. **JAMA ophthalmology**, American Medical Association, v. 137, n. 3, p. 258–264, 2019. Citations on pages 59, 140, and 141.

CAI, P. L.; HITCHMAN, L. H.; MOHAMED, A. H.; SMITH, G. E.; CHETTER, I.; CAR-RADICE, D. Endovenous ablation for venous leg ulcers. **Cochrane Database of Systematic Reviews**, John Wiley & Sons, Ltd, n. 7, 2023. Citation on page 121.

CALVO-ARIZA, N. R.; GÓMEZ-GÓMEZ, L. F.; OROZCO-ARROYAVE, J. R. Classical fe analysis to classify parkinson's disease patients. **Electronics**, MDPI, v. 11, n. 21, p. 3533, 2022. Citations on pages 61 and 72.

CARAMIA, C.; TORRICELLI, D.; SCHMID, M.; MUÑOZ-GONZALEZ, A.; GONZALEZ-VARGAS, J.; GRANDAS, F.; PONS, J. L. Imu-based classification of parkinson's disease from gait: A sensitivity analysis on sensor location and feature selection. **IEEE Journal of Biomedical and Health Informatics**, v. 22, n. 6, p. 1765–1774, 2018. Citation on page 97.

CARELLI, L.; SOLCA, F.; TAGINI, S.; TORRE, S.; VERDE, F.; TICOZZI, N.; CONSONNI, M.; FERRUCCI, R.; PRAVETTONI, G.; POLETTI, B. *et al.* Emotional processing and experience in amyotrophic lateral sclerosis: A systematic and critical review. **Brain Sciences**, MDPI, v. 11, n. 10, p. 1356, 2021. Citation on page 74.

CERNAK, M.; OROZCO-ARROYAVE, J. R.; RUDZICZ, F.; CHRISTENSEN, H.; VÁSQUEZ-CORREA, J. C.; NÖTH, E. Characterisation of voice quality of parkinson's disease using differential phonological posterior features. **Computer Speech & Language**, Elsevier, v. 46, p. 196–208, 2017. Citations on pages 55 and 98.

CHAWLA, N. V.; BOWYER, K. W.; HALL, L. O.; KEGELMEYER, W. P. Smote: synthetic minority over-sampling technique. **Journal of artificial intelligence research**, v. 16, p. 321–357, 2002. Citations on pages 61 and 63.

CHEN, J.; CHEN, Z.; CHI, Z.; FU, H. Facial expression recognition in video with multiple feature fusion. **IEEE Transactions on Affective Computing**, IEEE, v. 9, n. 1, p. 38–50, 2016. Citation on page 41.

CHEN, T.; GUESTRIN, C. XGBoost: A scalable tree boosting system. In: **Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining**. New York, NY, USA: ACM, 2016. (KDD '16), p. 785–794. ISBN 978-1-4503-4232-2. Available: <<http://doi.acm.org/10.1145/2939672.2939785>>. Citation on page 76.

CHEONG, J. H.; JOLLY, E.; XIE, T.; BYRNE, S.; KENNEY, M.; CHANG, L. J. Py-feat: Python facial expression analysis toolbox. **Affective Science**, Springer, p. 1–16, 2023. Citation on page 75.

CHRISTODOULOU, E.; MA, J.; COLLINS, G. S.; STEYERBERG, E. W.; VERBAKEL, J. Y.; CALSTER, B. V. A systematic review shows no performance benefit of machine learning over

logistic regression for clinical prediction models. **Journal of clinical epidemiology**, Elsevier, v. 110, p. 12–22, 2019. Citation on page 77.

CHUI, K. T.; LYTRAS, M. D.; VASANT, P. Combined generative adversarial network and fuzzy c-means clustering for multi-class voice disorder detection with an imbalanced dataset. **Applied Sciences**, MDPI, v. 10, n. 13, p. 4571, 2020. Citation on page 107.

COHEN, J. **Statistical Power Analysis for the Behavioral Sciences**. New York, NY.: Routledge Academic, 1988. Citation on page 111.

CORTES, C.; VAPNIK, V. Support-vector networks. **Machine learning**, Springer, v. 20, p. 273–297, 1995. Citations on pages 83 and 115.

COX, D. R. The regression analysis of binary sequences. **Journal of the Royal Statistical Society: Series B (Methodological)**, Wiley Online Library, v. 20, n. 2, p. 215–232, 1958. Citations on pages 63, 64, 77, and 81.

CROITORU, F.-A.; HONDRU, V.; IONESCU, R. T.; SHAH, M. Diffusion models in vision: A survey. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, IEEE, 2023. Citation on page 142.

CUTLER, A.; CUTLER, D. R.; STEVENS, J. R. Random forests. **Ensemble machine learning: Methods and applications**, Springer, p. 157–175, 2012. Citation on page 100.

CWAJDA-BIAŁASIK, J.; MOŚCICKA, P.; JAWIEŃ, A.; SZEWCZYK, M. T. Infrared thermography to prognose the venous leg ulcer healing process—preliminary results of a 12-week, prospective observational study. **Wound Repair and Regeneration**, Wiley Online Library, v. 28, n. 2, p. 224–233, 2020. Citation on page 58.

DAHLMANN, S.; REICH-SCHUPKE, S.; SCHOLLEMAN, F.; STÜCKER, M.; LEONHARDT, S.; TEICHMANN, D. Classification of chronic venous diseases based on skin temperature patterns. **Physiological Measurement**, IOP Publishing, v. 42, n. 4, p. 045001, 2021. Citation on page 58.

DALAL, N.; TRIGGS, B. Histograms of oriented gradients for human detection. In: **IEEE. 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)**. [S.l.], 2005. v. 1, p. 886–893. Citation on page 39.

DEHAK, N.; DUMOUCHEL, P.; KENNY, P. Modeling prosodic features with joint factor analysis for speaker verification. **IEEE Transactions on Audio, Speech, and Language Processing**, IEEE, v. 15, n. 7, p. 2095–2103, 2007. Citation on page 100.

DEMROZI, F.; BACCHIN, R.; TAMBURIN, S.; CRISTANI, M.; PRAVADELLI, G. Toward a wearable system for predicting freezing of gait in people affected by parkinson's disease. **IEEE journal of biomedical and health informatics**, IEEE, v. 24, n. 9, p. 2444–2451, 2019. Citation on page 33.

DENG, J.; DONG, W.; SOCHER, R.; LI, L.-J.; LI, K.; FEI-FEI, L. Imagenet: A large-scale hierarchical image database. In: **2009 IEEE Conference on Computer Vision and Pattern Recognition**. [S.l.: s.n.], 2009. p. 248–255. Citation on page 61.

\_\_\_\_\_. Imagenet: A large-scale hierarchical image database. In: **IEEE. 2009 IEEE conference on computer vision and pattern recognition**. [S.l.], 2009. p. 248–255. Citation on page 135.

DÉNIZ, O.; BUENO, G.; SALIDO, J.; TORRE, F. De la. Face recognition using histograms of oriented gradients. **Pattern recognition letters**, Elsevier, v. 32, n. 12, p. 1598–1603, 2011. Citation on page 39.

DIBAZAR, A. A.; BERGER, T. W.; NARAYANAN, S. S. Pathological voice assessment. In: IEEE. **2006 international conference of the IEEE engineering in medicine and biology society**. [S.l.], 2006. p. 1669–1673. Citation on page 107.

DIMAURO, G.; Di Nicola, V.; BEVILACQUA, V.; CAIVANO, D.; GIRARDI, F. Assessment of speech intelligibility in Parkinson's disease using a speech-to-text system. **IEEE Access**, v. 5, p. 22199–22208, 2017. ISSN 21693536. Citation on page 110.

DINI, V.; SALVO, P.; JANOWSKA, A.; FRANCESCO, F. D.; BARBINI, A.; ROMANELLI, M. Correlation between wound temperature obtained with an infrared camera and clinical wound bed score in venous leg ulcers. **Wounds: a compendium of clinical research and practice**, v. 27, n. 10, p. 274–278, 2015. Citation on page 122.

DISEASE, M. D. S. T. F. on Rating Scales for P. The unified parkinson's disease rating scale (updrs): status and recommendations. **Movement Disorders**, Wiley Online Library, v. 18, n. 7, p. 738–750, 2003. Citation on page 97.

DOOLUB, G.; MAMALAKIS, M.; ALABED, S.; GEEST, R. J. Van der; SWIFT, A. J.; RODRIGUES, J. C.; GARG, P.; JOSHI, N. V.; DASTIDAR, A. Artificial intelligence as a diagnostic tool in non-invasive imaging in the assessment of coronary artery disease. **Medical Sciences**, MDPI, v. 11, n. 1, p. 20, 2023. Citation on page 19.

DREISEITL, S.; OHNO-MACHADO, L. Logistic regression and artificial neural network classification models: a methodology review. **Journal of biomedical informatics**, Elsevier, v. 35, n. 5-6, p. 352–359, 2002. Citation on page 100.

EFRAIMIDIS, P.; SPIRAKIS, P. Weighted random sampling. In: KAO, M.-Y. (Ed.). **Encyclopedia of Algorithms**. Boston, MA: Springer US, 2008. p. 1024–1027. ISBN 978-0-387-30162-4. Available: <[https://doi.org/10.1007/978-0-387-30162-4\\_478](https://doi.org/10.1007/978-0-387-30162-4_478)>. Citation on page 135.

EKMAN, P.; FRIESEN, W. V. Constants across cultures in the face and emotion. **Journal of personality and social psychology**, American Psychological Association, v. 17, n. 2, p. 124, 1971. Citation on page 40.

\_\_\_\_\_. Facial action coding system. **Environmental Psychology & Nonverbal Behavior**, 1978. Citation on page 62.

ELMAGARMID, A.; FEDOROWICZ, Z.; HAMMADY, H.; ILYAS, I.; KHABSA, M.; OUZZANI, M. Rayyan: a systematic reviews web app for exploring and filtering searches for eligible studies for cochrane reviews. In: JOHN WILEY & SONS HYDERABAD, INDIA, INDIA. **Evidence-Informed Public Health: Opportunities and Challenges. Abstracts of the 22nd Cochrane Colloquium**. [S.l.], 2014. p. 21–26. Citation on page 36.

ESMAEILZADEH, P. Challenges and strategies for wide-scale artificial intelligence (ai) deployment in healthcare practices: A perspective for healthcare organizations. **Artificial Intelligence in Medicine**, Elsevier, v. 151, p. 102861, 2024. Citation on page 20.

FÁDEL, T. M. D. M.; CARVALHO, S. D.; SANTOS, B. A. D. *et al.* Facial expression recognition in alzheimer's disease: a systematic review. **Journal of Clinical and Experimental Neuropsychology (Neuropsychology, Developm**, Routledge, part of the Taylor & Francis Group, v. 41, n. 2, p. 192–203, 2019. Citation on page [54](#).

FARHAD, M.; MASUD, M. M.; BEG, A.; AHMAD, A.; AHMED, L. A review of medical diagnostic video analysis using deep learning techniques. **Applied Sciences**, MDPI, v. 13, n. 11, p. 6582, 2023. Citations on pages [29](#) and [30](#).

FEENY, A. K.; CHUNG, M. K.; MADABHUSHI, A.; ATTIA, Z. I.; CIKES, M.; FIROUZANIA, M.; FRIEDMAN, P. A.; KALSCHEUR, M. M.; KAPA, S.; NARAYAN, S. M. *et al.* Artificial intelligence and machine learning in arrhythmias and cardiac electrophysiology. **Circulation: Arrhythmia and Electrophysiology**, Am Heart Assoc, v. 13, n. 8, p. e007952, 2020. Citation on page [19](#).

FEIGIN, V. L.; NICHOLS, E.; ALAM, T.; BANNICK, M. S.; BEGHI, E.; BLAKE, N.; CULPEPPER, W. J.; DORSEY, E. R.; ELBAZ, A.; ELLENBOGEN, R. G. *et al.* Global, regional, and national burden of neurological disorders, 1990–2016: a systematic analysis for the global burden of disease study 2016. **The Lancet Neurology**, Elsevier, v. 18, n. 5, p. 459–480, 2019. Citation on page [114](#).

FLECKENSTEIN, M.; SCHMITZ-VALCKENBERG, S.; CHAKRAVARTHY, U. Age-related macular degeneration: a review. **Jama**, American Medical Association, v. 331, n. 2, p. 147–157, 2024. Citation on page [59](#).

FRID-ADAR, M.; DIAMANT, I.; KLANG, E.; AMITAI, M.; GOLDBERGER, J.; GREENSPAN, H. Gan-based synthetic medical image augmentation for increased cnn performance in liver lesion classification. **Neurocomputing**, Elsevier, v. 321, p. 321–331, 2018. Citations on pages [59](#) and [61](#).

FRIESEN, E.; EKMAN, P. Facial action coding system: a technique for the measurement of facial movement. **Palo Alto**, v. 3, n. 2, p. 5, 1978. Citation on page [74](#).

FRÖHLICH, H.; BONTRIDDER, N.; PETROVSKA-DELACRÉTA, D.; GLAAB, E.; KLUGE, F.; YACOUBI, M. E.; VALERO, M. M.; CORVOL, J.-C.; ESKOFIER, B.; GYSEGHEM, J.-M. V. *et al.* Leveraging the potential of digital technology for better individualized treatment of parkinson's disease. **Frontiers in neurology**, Frontiers Media SA, v. 13, 2022. Citation on page [52](#).

FU, H.; LI, F.; ORLANDO, J.; BOGUNOVIC, H.; SUN, X.; LIAO, J.; XU, Y.; ZHANG, S.; ZHANG, X. Adam: Automatic detection challenge on age-related macular degeneration. **IEEE Dataport**, 2020. Citations on pages [8](#), [130](#), and [132](#).

FU, H.; WANG, B.; SHEN, J.; CUI, S.; XU, Y.; LIU, J.; SHAO, L. Evaluation of retinal image quality assessment networks in different color-spaces. In: SPRINGER. **International Conference on Medical Image Computing and Computer-Assisted Intervention**. [S.l.], 2019. p. 48–56. Citations on pages [131](#) and [133](#).

FULLARD, M. E.; MORLEY, J. F.; DUDA, J. E. Olfactory dysfunction as an early biomarker in parkinson's disease. **Neuroscience bulletin**, Springer, v. 33, n. 5, p. 515–525, 2017. Citations on pages [32](#) and [33](#).

GALAZ, Z.; MEKYSKA, J.; MZOUREK, Z.; SMEKAL, Z.; REKTOROVA, I.; ELIASOVA, I.; KOSTALOVA, M.; MRACKOVA, M.; BERANKOVA, D. Prosodic analysis of neutral, stress-modified and rhymed speech in patients with parkinson's disease. **Computer methods and programs in biomedicine**, Elsevier, v. 127, p. 301–317, 2016. Citations on pages [56](#) and [99](#).

GARCIA, N.; OROZCO-ARROYAVE, J. R.; FERNANDO, D. L.; DEHAK, N.; NÖTH, E. Evaluation of the neurological state of people with parkinson's disease using i-vectors. In: **Interspeech**. [S.l.: s.n.], 2017. p. 299–303. Citations on pages [55](#) and [98](#).

GHIMIRE, D.; LEE, J. Geometric feature-based facial expression recognition in image sequences using multi-class adaboost and support vector machines. **Sensors**, Multidisciplinary Digital Publishing Institute, v. 13, n. 6, p. 7714–7734, 2013. Citation on page [43](#).

GHORAANI, B.; HSSAYENI, M. D.; BRUACK, M. M.; JIMENEZ-SHAHED, J. Multilevel features for sensor-based assessment of motor fluctuation in parkinson's disease subjects. **IEEE Journal of Biomedical and Health Informatics**, IEEE, v. 24, n. 5, p. 1284–1295, 2019. Citation on page [33](#).

GILLIVAN-MURPHY, P.; CARDING, P.; MILLER, N. Vocal tract characteristics in Parkinson's disease. **Speech Therapy and Rehabilitation**, v. 24, n. 3, p. 175–182, 2016. Citation on page [113](#).

GOETZ, C. G. The history of parkinson's disease: early clinical descriptions and neurological therapies. **Cold Spring Harbor perspectives in medicine**, Cold Spring Harbor Laboratory Press, v. 1, n. 1, p. a008862, 2011. Citation on page [33](#).

GOETZ, C. G.; TILLEY, B. C.; SHAFTMAN, S. R.; STEBBINS, G. T.; FAHN, S.; MARTINEZ-MARTIN, P.; POEWE, W.; SAMPAIO, C.; STERN, M. B.; DODEL, R. *et al.* Movement disorder society-sponsored revision of the unified parkinson's disease rating scale (mds-updrs): scale presentation and clinimetric testing results. **Movement disorders: official journal of the Movement Disorder Society**, Wiley Online Library, v. 23, n. 15, p. 2129–2170, 2008. Citation on page [97](#).

GOMES, N. B.; YOSHIDA, A.; OLIVEIRA, G. C.; RODER, M.; PAPA, J. P. Facial point graphs for stroke identification. In: SPRINGER. **Iberoamerican Congress on Pattern Recognition**. [S.l.], 2023. p. 685–699. Citations on pages [119](#) and [120](#).

GOMES, N. B.; YOSHIDA, A.; RODER, M.; OLIVEIRA, G. C. de; PAPA, J. P. Facial point graphs for amyotrophic lateral sclerosis identification. In: **Proceedings of the 19th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 3: VISAPP**. [S.l.]: SciTePress, 2024. p. 207–214. Citations on pages [83](#) and [84](#).

GOMES, N. B.; YOSHIDA, A.; RODER, M.; OLIVEIRA, G. C.; PAPA, J. P. Facial point graphs for amyotrophic lateral sclerosis identification. In: INSTICC. **Proceedings of the 19th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 3: VISAPP**. [S.l.]: SciTePress, 2024. p. 207–214. ISBN 978-989-758-679-8. ISSN 2184-4321. Citations on pages [119](#) and [120](#).

GOMEZ-GOMEZ, L. F.; MORALES, A.; FIERREZ, J.; OROZCO-ARROYAVE, J. R. Exploring facial expressions and affective domains for parkinson detection. **arXiv preprint arXiv:2012.06563**, 2020. Citations on pages [38](#), [44](#), [46](#), [49](#), and [50](#).

GOMEZ, L. F.; MORALES, A.; FIERREZ, J.; OROZCO-ARROYAVE, J. R. Exploring facial expressions and action unit domains for parkinson detection. **Plos one**, Public Library of Science San Francisco, CA USA, v. 18, n. 2, p. e0281248, 2023. Citations on pages 41, 45, 47, 61, and 72.

GOMEZ, L. F.; MORALES, A.; OROZCO-ARROYAVE, J. R.; DAZA, R.; FIERREZ, J. Improving parkinson detection using dynamic features from evoked expressions in video. In: **Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition**. [S.l.: s.n.], 2021. p. 1562–1570. Citations on pages 40, 44, 46, 48, 49, 50, 61, and 72.

GONZÁLEZ, J. J. V.; BERGER, M.; SCHIELE, S.; RUBECK, A.; MÜLLER, G.; WELZEL, J.; SCHUH, S. Dynamic optical coherence tomography of chronic venous ulcers. **Journal of the European Academy of Dermatology and Venereology**, Wiley Online Library, v. 38, n. 1, p. 223–231, 2024. Citation on page 121.

GOVINDAIAH, A.; SMITH, R. T.; BHUIYAN, A. A new and improved method for automated screening of age-related macular degeneration using ensemble deep neural networks. In: IEEE. **2018 40th Annual international conference of the IEEE engineering in medicine and biology society (EMBC)**. [S.l.], 2018. p. 702–705. Citation on page 141.

GOYAL, J.; KHANDNOR, P.; ASERI, T. C. Engineering Applications of Artificial Intelligence Classification, Prediction, and Monitoring of Parkinson’s disease using Computer Assisted Technologies: A Comparative Analysis \*. **Engineering Applications of Artificial Intelligence**, Elsevier Ltd, v. 96, n. August, p. 103955, 2020. ISSN 0952-1976. Available: <<https://doi.org/10.1016/j.engappai.2020.103955>>. Citation on page 110.

GRAMMATIKOPOULOU, A.; GRAMMALIDIS, N.; BOSTANTJOPOULOU, S.; KATSAROU, Z. Detecting hypomimia symptoms by selfie photo analysis: for early parkinson disease detection. In: **Proceedings of the 12th ACM international conference on pervasive technologies related to assistive environments**. [S.l.: s.n.], 2019. p. 517–522. Citations on pages 38, 41, 44, 45, and 46.

GRASSMANN, F.; MENGELKAMP, J.; BRANDL, C.; HARSCH, S.; ZIMMERMANN, M. E.; LINKOHR, B.; PETERS, A.; HEID, I. M.; PALM, C.; WEBER, B. H. A deep learning algorithm for prediction of age-related eye disease study severity scale for age-related macular degeneration from color fundus photography. **Ophthalmology**, Elsevier, v. 125, n. 9, p. 1410–1420, 2018. Citation on page 141.

GUAN, Y. Application of logistic regression algorithm in the diagnosis of expression disorder in parkinson’s disease. In: IEEE. **2021 IEEE 2nd International Conference on Information Technology, Big Data and Artificial Intelligence (ICIBA)**. [S.l.], 2021. v. 2, p. 1117–1120. Citations on pages 40, 44, 46, and 49.

GULRAJANI, I.; AHMED, F.; ARJOVSKY, M.; DUMOULIN, V.; COURVILLE, A. Improved training of wasserstein gans. **arXiv preprint arXiv:1704.00028**, 2017. Citation on page 135.

GUNNERY, S. D.; HABERMANN, B.; SAINT-HILAIRE, M.; THOMAS, C. A.; TICKLE-DEGNEN, L. The relationship between the experience of hypomimia and social wellbeing in people with parkinson’s disease and their care partners. **Journal of Parkinson’s disease**, IOS Press, v. 6, n. 3, p. 625–630, 2016. Citation on page 33.

GUNNERY, S. D.; NAUMOVA, E. N.; SAINT-HILAIRE, M.; TICKLE-DEGNEN, L. Mapping spontaneous facial expression in people with parkinson's disease: A multiple case study design. **Cogent Psychology**, Taylor & Francis, v. 4, n. 1, p. 1376425, 2017. Citation on page 35.

HAAN, J. de; STOOP, M.; ZUIJLEN, P. P. van; PIJPE, A. Thermal imaging for burn wound depth assessment: A mixed-methods implementation study. **Journal of Clinical Medicine**, MDPI, v. 13, n. 7, p. 2061, 2024. Citation on page 58.

HAMMADI, Y.; GRONDIN, F.; FERLAND, F.; LEBEL, K. Evaluation of various state of the art head pose estimation algorithms for clinical scenarios. **Sensors**, MDPI, v. 22, n. 18, p. 6850, 2022. Citation on page 62.

HAMMER, M. J. Aerodynamic Assessment of Phonatory Onset in Parkinson's Disease: Evidence of Decreased Scaling of Laryngeal and Respiratory Control. v. 3, p. 173–179, 2013. Citation on page 109.

HARRIS, C. R.; MILLMAN, K. J.; WALT, S. J. van der; GOMMERS, R.; VIRTANEN, P.; COURNAPEAU, D.; WIESER, E.; TAYLOR, J.; BERG, S.; SMITH, N. J.; KERN, R.; PICUS, M.; HOYER, S.; KERKWIJK, M. H. van; BRETT, M.; HALDANE, A.; RÍO, J. F. del; WIEBE, M.; PETERSON, P.; GÉRARD-MARCHANT, P.; SHEPPARD, K.; REDDY, T.; WECKESSER, W.; ABBASI, H.; GOHLKE, C.; OLIPHANT, T. E. Array programming with NumPy. **Nature**, Springer Science and Business Media LLC, v. 585, n. 7825, p. 357–362, Sep. 2020. Available: <<https://doi.org/10.1038/s41586-020-2649-2>>. Citation on page 65.

HARVEY, P. T. Common eye diseases of elderly people: identifying and treating causes of vision loss. **Gerontology**, Karger Publishers, v. 49, n. 1, p. 1–11, 2003. Citation on page 130.

HE, H.; BAI, Y.; GARCIA, E. A.; LI, S. Adasyn: Adaptive synthetic sampling approach for imbalanced learning. In: IEEE. **2008 IEEE international joint conference on neural networks (IEEE world congress on computational intelligence)**. [S.l.], 2008. p. 1322–1328. Citation on page 63.

HE, K.; ZHANG, X.; REN, S.; SUN, J. Deep residual learning for image recognition. In: **Proceedings of the IEEE conference on computer vision and pattern recognition**. [S.l.: s.n.], 2016. p. 770–778. Citation on page 135.

HEARST, M. A.; DUMAIS, S. T.; OSUNA, E.; PLATT, J.; SCHOLKOPF, B. Support vector machines. **IEEE Intelligent Systems and their applications**, IEEE, v. 13, n. 4, p. 18–28, 1998. Citations on pages 63 and 64.

HEUSEL, M.; RAMSAUER, H.; UNTERTHINER, T.; NESSLER, B.; HOCHREITER, S. Gans trained by a two time-scale update rule converge to a local nash equilibrium. **Advances in neural information processing systems**, v. 30, 2017. Citation on page 134.

HIREŠ, M.; GAZDA, M.; DROTAR, P.; PAH, N. D.; MOTIN, M. A.; KUMAR, D. K. Convolutional neural network ensemble for parkinson's disease detection from voice recordings. **Computers in biology and medicine**, Elsevier, v. 141, p. 105021, 2022. Citation on page 97.

HO, M. W.-R.; CHIEN, S. H.-L.; LU, M.-K.; CHEN, J.-C.; AOH, Y.; CHEN, C.-M.; LANE, H.-Y.; TSAI, C.-H. Impairments in face discrimination and emotion recognition are related to aging and cognitive dysfunctions in parkinson's disease with dementia. **Scientific reports**, Nature Publishing Group, v. 10, n. 1, p. 1–8, 2020. Citation on page 33.

HORNYKIEWICZ, O. Biochemical aspects of parkinson's disease. **Neurology**, AAN Enterprises, v. 51, n. 2 Suppl 2, p. S2–S9, 1998. Citation on page [96](#).

HOU, X.; ZHANG, Y.; WANG, Y.; WANG, X.; ZHAO, J.; ZHU, X.; SU, J. *et al.* A markerless 2d video, facial feature recognition-based, artificial intelligence model to assist with screening for parkinson disease: Development and usability study. **Journal of medical Internet research**, JMIR Publications Inc., Toronto, Canada, v. 23, n. 11, p. e29554, 2021. Citations on pages [39](#), [44](#), [46](#), [48](#), [51](#), [61](#), [70](#), and [72](#).

HU, C.; ZHANG, P.; HUANG, W. A novel face-based approach for the early diagnosis of parkinson's disease. In: **IEEE. 2021 International Conference on Information Technology and Biomedical Engineering (ICITBE)**. [S.l.], 2021. p. 248–252. Citations on pages [40](#), [44](#), [46](#), [50](#), and [52](#).

HU, E. J.; SHEN, Y.; WALLIS, P.; ALLEN-ZHU, Z.; LI, Y.; WANG, S.; WANG, L.; CHEN, W. Lora: Low-rank adaptation of large language models. **arXiv preprint arXiv:2106.09685**, 2021. Citation on page [122](#).

HUANG, J.; LIN, L.; YU, F.; HE, X.; SONG, W.; LIN, J.; TANG, Z.; YUAN, K.; LI, Y.; HUANG, H. *et al.* Parkinson's severity diagnosis explainable model based on 3d multi-head attention residual network. **Computers in Biology and Medicine**, Elsevier, v. 170, p. 107959, 2024. Citations on pages [43](#), [45](#), and [48](#).

HUANG, M.; PU, T.; MIRT, G.; LEI, X.; JIANG, Y.; WEI, Y.; ZOU, X.; LIU, X.; LIU, F.; XU, F.; CHEN, W. Chapter 2: The Reasoning of Dysarthria in Parkinson ' s Disease. In: **Neurodegenerative Diseases Symptoms and Treatment**. Las Vegas, USA: [s.n.], 2019. Citations on pages [109](#) and [112](#).

HUANG, W.; XU, W.; WAN, R.; ZHANG, P.; ZHA, Y.; PANG, M. Auto diagnosis of parkinson's disease via a deep learning model based on mixed emotional facial expressions. **IEEE journal of biomedical and health informatics**, IEEE, 2023. Citations on pages [42](#), [45](#), and [47](#).

IANDOLA, F. N.; HAN, S.; MOSKEWICZ, M. W.; ASHRAF, K.; DALLY, W. J.; KEUTZER, K. Squeezenet: Alexnet-level accuracy with 50x fewer parameters and < 0.5 mb model size. **arXiv preprint arXiv:1602.07360**, 2016. Citation on page [135](#).

IENCA, M.; IGNATIADIS, K. Artificial intelligence in clinical neuroscience: methodological and ethical challenges. **AJOB neuroscience**, Taylor & Francis, v. 11, n. 2, p. 77–87, 2020. Citation on page [20](#).

IKECHUKWU, M. C.; WANG, J. Tackling visual illumination variations in fall detection for healthcare applications. 2024. Citation on page [30](#).

IPAPO, T. J.; ROSARIO, C. D.; ABU, P. A.; ALAMPAY, R. Clinical score estimation for determining oro-facial dysfunction severity. In: **Proceedings of the 2023 International Conference on Robotics, Control and Vision Engineering**. [S.l.: s.n.], 2023. p. 76–80. Citation on page [31](#).

ISOLA, P.; ZHU, J.-Y.; ZHOU, T.; EFROS, A. A. Image-to-image translation with conditional adversarial networks. In: **Proceedings of the IEEE conference on computer vision and pattern recognition**. [S.l.: s.n.], 2017. p. 1125–1134. Citation on page [127](#).

JAEGLER, A.; BORGEAUD, S.; ALAYRAC, J.-B.; DOERSCH, C.; IONESCU, C.; DING, D.; KOPPULA, S.; ZORAN, D.; BROCK, A.; SHELHAMER, E. *et al.* Perceiver io: A general architecture for structured inputs & outputs. **arXiv preprint arXiv:2107.14795**, 2021. Citation on page 52.

JAEGLER, A.; GIMENO, F.; BROCK, A.; VINYALS, O.; ZISSERMAN, A.; CARREIRA, J. Perceiver: General perception with iterative attention. In: PMLR. **International Conference on Machine Learning**. [S.l.], 2021. p. 4651–4664. Citation on page 52.

JAKUBOWSKI, J.; POTULSKA-CHROMIK, A.; BIAŁEK, K.; NOJSZEWSKA, M.; KOSTERA-PRUSZCZYK, A. A study on the possible diagnosis of parkinson's disease on the basis of facial image analysis. **Electronics**, Multidisciplinary Digital Publishing Institute, v. 10, n. 22, p. 2832, 2021. Citations on pages 39, 44, 46, 61, and 72.

JANKOVIC, J. Parkinson's disease: clinical features and diagnosis. **Journal of neurology, neurosurgery & psychiatry**, BMJ Publishing Group Ltd, v. 79, n. 4, p. 368–376, 2008. Citations on pages 32 and 33.

\_\_\_\_\_. Parkinson's disease: clinical features and diagnosis. **Journal of Neurology, Neurosurgery, and Psychiatry**, v. 79, p. 368 – 376, 2008. Available: <<https://api.semanticscholar.org/CorpusID:4824951>>. Citation on page 96.

JÄNTSCHI, L.; BOLBOACĂ, S. D. Computation of Probability Associated with Anderson–Darling Statistic. **Mathematics**, v. 6, n. 6, p. 1–16, 2018. Citation on page 111.

JAVAID, M.; HALEEM, A.; SINGH, R. P. Chatgpt for healthcare services: An emerging stage for an innovative perspective. **BenchCouncil Transactions on Benchmarks, Standards and Evaluations**, Elsevier, v. 3, n. 1, p. 100105, 2023. Citation on page 114.

JIANG, F.; JIANG, Y.; ZHI, H.; DONG, Y.; LI, H.; MA, S.; WANG, Y.; DONG, Q.; SHEN, H.; WANG, Y. Artificial intelligence in healthcare: past, present and future. **Stroke and vascular neurology**, BMJ Specialist Journals, v. 2, n. 4, 2017. Citation on page 114.

Jiang J, O'Mara T, Chen HJ, Stern JJ, Vlagos D, H. D. Aerodynamic measurements of patients with Parkinson's disease. **Journal of Voice**, v. 13, n. 4, 1999. Citation on page 109.

JIN, B.; QU, Y.; ZHANG, L.; GAO, Z. Diagnosing parkinson disease through facial expression recognition: video analysis. **Journal of medical Internet research**, JMIR Publications Toronto, Canada, v. 22, n. 7, p. e18697, 2020. Citations on pages 10, 38, 39, 44, 45, 46, 50, 51, 61, and 72.

JONAS, J. B. Global prevalence of age-related macular degeneration. **The Lancet Global Health**, Elsevier, v. 2, n. 2, p. e65–e66, 2014. Citation on page 130.

JOSHI, A.; TICKLE-DEGNEN, L.; GUNNERY, S.; ELLIS, T.; BETKE, M. Predicting active facial expressivity in people with parkinson's disease. In: **Proceedings of the 9th ACM International Conference on PErvasive Technologies Related to Assistive Environments**. [S.l.: s.n.], 2016. p. 1–4. Citations on pages 35 and 61.

KADU, A.; SINGH, M. Comparative analysis of e-health care telemedicine system based on internet of medical things and artificial intelligence. In: IEEE. **2021 2nd International conference on smart electronics and communication (ICOSEC)**. [S.l.], 2021. p. 1768–1775. Citation on page 20.

KAEWMAHANIN, W.; RASSAMEECHAROENCHAI, T.; JUTHAREE, W.; TONGSKUL-ROONGRUANG, T.; WIPHUNAWAT, P.; JENNAWASIN, T.; KAEWKAMNERDPONG, B. Automatic facial asymmetry analysis for elderly stroke detection by using cosine similarity. In: IEEE. **2022 19th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)**. [S.l.], 2022. p. 1–4. Citations on pages [31](#) and [85](#).

KAN, Y.; KAWAMURA, M.; HASEGAWA, Y.; MOCHIZUKI, S.; NAKAMURA, K. Recognition of emotion from facial, prosodic and written verbal stimuli in parkinson's disease. **Cortex**, Elsevier, v. 38, n. 4, p. 623–630, 2002. Citation on page [33](#).

KANG, J.; DERVA, D.; KWON, D.-Y.; WALLRAVEN, C. Voluntary and spontaneous facial mimicry toward other's emotional expression in patients with parkinson's disease. **PLoS one**, Public Library of Science San Francisco, CA USA, v. 14, n. 4, p. e0214957, 2019. Citation on page [36](#).

KARAN, B.; SAHU, S. S.; OROZCO-ARROYAVE, J. R. An investigation about the relationship between dysarthria level of speech and the neurological state of parkinson's patients. **Biocybernetics and Biomedical Engineering**, Elsevier, v. 42, n. 2, p. 710–726, 2022. Citation on page [106](#).

KARRAS, T.; AITTALA, M.; HELLSTEN, J.; LAINE, S.; LEHTINEN, J.; AILA, T. Training generative adversarial networks with limited data. In: **Proc. NeurIPS**. [S.l.: s.n.], 2020. Citations on pages [131](#) and [136](#).

KASULA, B. Y. Advancements in ai-driven healthcare: A comprehensive review of diagnostics, treatment, and patient care integration. **International Journal of Machine Learning for Sustainable Development**, v. 6, n. 1, p. 1–5, 2024. Citation on page [19](#).

KATSIKITIS, M.; PILOWSKY, I. A study of facial expression in parkinson's disease using a novel microcomputer-based method. **Journal of Neurology, Neurosurgery & Psychiatry**, BMJ Publishing Group Ltd, v. 51, n. 3, p. 362–366, 1988. Citation on page [33](#).

KAZANDJIAN, S.; BOROD, J. C.; BRICKMAN, A. M. Facial expression during emotional monologues in unilateral stroke: an analysis of monologue segments. **Applied neuropsychology**, Taylor & Francis, v. 14, n. 4, p. 235–246, 2007. Citation on page [54](#).

KEEL, S.; LI, Z.; SCHEETZ, J.; ROBMAN, L.; PHUNG, J.; MAKEYEVA, G.; AUNG, K.; LIU, C.; YAN, X.; MENG, W. *et al.* Development and validation of a deep-learning algorithm for the detection of neovascular age-related macular degeneration from colour fundus photographs. **Clinical & Experimental Ophthalmology**, Wiley Online Library, v. 47, n. 8, p. 1009–1018, 2019. Citation on page [141](#).

KELLY, C. J.; KARTHIKESALINGAM, A.; SULEYMAN, M.; CORRADO, G.; KING, D. Key challenges for delivering clinical impact with artificial intelligence. **BMC medicine**, Springer, v. 17, p. 1–9, 2019. Citation on page [114](#).

KHAN, B.; FATIMA, H.; QURESHI, A.; KUMAR, S.; HANAN, A.; HUSSAIN, J.; ABDULLAH, S. Drawbacks of artificial intelligence and their potential solutions in the healthcare sector. **Biomedical Materials & Devices**, Springer, v. 1, n. 2, p. 731–738, 2023. Citation on page [19](#).

KHAN, T.; LUNDGREN, L. E.; ANDERSON, D. G.; NOWAK, I.; DOUGHERTY, M.; VERIKAS, A.; PAVEL, M.; JIMISON, H.; NOWACZYK, S.; AHARONSON, V. Assessing parkinson's disease severity using speech analysis in non-native speakers. **Computer Speech & Language**, Elsevier, v. 61, p. 101047, 2020. Citation on page [107](#).

KIERNAN, M. C.; VUCIC, S.; CHEAH, B. C.; TURNER, M. R.; EISEN, A.; HARDIMAN, O.; BURRELL, J. R.; ZOING, M. C. Amyotrophic lateral sclerosis. **The lancet**, Elsevier, v. 377, n. 9769, p. 942–955, 2011. Citation on page [74](#).

KINGMA, D. P.; BA, J. **Adam: A Method for Stochastic Optimization**. 2017. Citations on pages [64](#) and [135](#).

KIRILLOV, A.; MINTUN, E.; RAVI, N.; MAO, H.; ROLLAND, C.; GUSTAFSON, L.; XIAO, T.; WHITEHEAD, S.; BERG, A. C.; LO, W.-Y. *et al.* Segment anything. In: **Proceedings of the IEEE/CVF International Conference on Computer Vision**. [S.l.: s.n.], 2023. p. 4015–4026. Citation on page [123](#).

KODALI, M.; KADIRI, S. R.; ALKU, P. Automatic classification of the severity level of parkinson's disease: A comparison of speaking tasks, features, and classifiers. **Computer Speech & Language**, Elsevier, v. 83, p. 101548, 2023. Citations on pages [55](#), [97](#), [98](#), [102](#), and [106](#).

KODALI, N.; ABERNETHY, J.; HAYS, J.; KIRA, Z. On convergence and stability of gans. **arXiv preprint arXiv:1705.07215**, 2017. Citation on page [135](#).

KODRASI, I.; PERNON, M.; LAGANARO, M.; BOURLARD, H. Automatic and perceptual discrimination between dysarthria, apraxia of speech, and neurotypical speech. In: IEEE. **ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)**. [S.l.], 2021. p. 7308–7312. Citation on page [107](#).

KOLARIK, M.; SARNOVSKY, M.; PARALIC, J.; BABIC, F. Explainability of deep learning models in medical video analysis: a survey. **PeerJ Computer Science**, PeerJ Inc., v. 9, p. e1253, 2023. Citation on page [29](#).

KOLLIAS, D.; ZAFEIRIOU, S. Aff-wild2: Extending the aff-wild database for affect recognition. **arXiv preprint arXiv:1811.07770**, 2018. Citation on page [76](#).

KOOB, J. L.; GORSKI, M.; KRICK, S.; MUSTIN, M.; FINK, G. R.; GREFKES, C.; REHME, A. K. Behavioral and neuroanatomical correlates of facial emotion processing in post-stroke depression. **NeuroImage: Clinical**, Elsevier, v. 41, p. 103586, 2024. Citation on page [32](#).

KOTSIA, I.; PITAS, I. Facial expression recognition in image sequences using geometric deformation features and support vector machines. **IEEE transactions on image processing**, IEEE, v. 16, n. 1, p. 172–187, 2006. Citation on page [43](#).

KOTSIANTIS, S.; KANELLOPOULOS, D.; PINTELAS, P. *et al.* Handling imbalanced datasets: A review. **GESTS international transactions on computer science and engineering**, v. 30, n. 1, p. 25–36, 2006. Citation on page [61](#).

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. **Advances in neural information processing systems**, v. 25, p. 1097–1105, 2012. Citation on page [135](#).

LAMMERT, A. C.; NARAYANAN, S. S. On short-time estimation of vocal tract length from formant frequencies. **PLoS ONE**, v. 10, n. 7, p. 1–23, 2015. ISSN 19326203. Citation on page [110](#).

LANGEVIN, R.; ALI, M. R.; SEN, T.; SNYDER, C.; MYERS, T.; DORSEY, E. R.; HOQUE, M. E. The park framework for automated analysis of parkinson's disease characteristics. **Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies**, ACM New York, NY, USA, v. 3, n. 2, p. 1–22, 2019. Citations on pages [36](#) and [62](#).

LAPCHAK, P. A.; ZHANG, J. H. The high cost of stroke and stroke cytoprotection research. **Translational stroke research**, Springer, v. 8, p. 307–317, 2017. Citations on pages [87](#) and [88](#).

LEE, T. K.; YANKEE, E. L. A review on parkinson's disease treatment. **Neuroimmunology and Neuroinflammation**, OAE Publishing Inc., v. 8, p. 222–44, 2021. Citation on page [32](#).

LENG, X.; SHI, R.; WU, Y.; ZHU, S.; CAI, X.; LU, X.; LIU, R. Deep learning for detection of age-related macular degeneration: A systematic review and meta-analysis of diagnostic test accuracy studies. **Plos one**, Public Library of Science San Francisco, CA USA, v. 18, n. 4, p. e0284060, 2023. Citation on page [141](#).

LI, F.; ZHANG, H.; XU, S.; MA, X.; LUO, N.; YU, Y.; HE, W.; JIN, H.; WANG, M.; WANG, T. *et al.* Exploration of machine learning models for surgical incision healing assessment based on thermal imaging: A feasibility study. **International Wound Journal**, Wiley Online Library, v. 21, n. 2, p. e14677, 2024. Citation on page [58](#).

LI, T.; BO, W.; HU, C.; KANG, H.; LIU, H.; WANG, K.; FU, H. Applications of deep learning in fundus images: A review. **Medical Image Analysis**, Elsevier, v. 69, p. 101971, 2021. Citation on page [59](#).

LIM, Y.-J.; KO, Y.-G.; SHIN, H.-C.; CHO, Y. Prevalence and correlates of complete mental health in the south korean adult population. In: **Mental well-being**. [S.l.]: Springer, 2013. p. 91–109. Citation on page [36](#).

LIOPYRIS, K.; GREGORIOU, S.; DIAS, J.; STRATIGOS, A. J. Artificial intelligence in dermatology: challenges and perspectives. **Dermatology and Therapy**, Springer, v. 12, n. 12, p. 2637–2651, 2022. Citation on page [19](#).

LOEHFELM, T. W. Artificial intelligence for quality improvement in radiology. **Radiologic Clinics**, Elsevier, v. 59, n. 6, p. 1053–1062, 2021. Citation on page [19](#).

LÓPEZ, J. V. E.; OROZCO-ARROYAVE, J. R.; GOSZTOLYA, G. Assessing parkinson's disease from speech using fisher vectors. In: . [S.l.]: ISCA-INT SPEECH COMMUNICATION ASSOC, 2019. Citations on pages [55](#) and [98](#).

LOVELEEN, G.; MOHAN, B.; SHIKHAR, B. S.; NZ, J.; SHORFUZZAMAN, M.; MASUD, M. Explanation-driven hci model to examine the mini-mental state for alzheimer's disease. **ACM Transactions on Multimedia Computing, Communications and Applications**, ACM New York, NY, v. 20, n. 2, p. 1–16, 2023. Citation on page [54](#).

LUCEY, P.; COHN, J. F.; KANADE, T.; SARAGIH, J.; AMBADAR, Z.; MATTHEWS, I. The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression. In: IEEE. **2010 ieee computer society conference on computer vision and pattern recognition-workshops**. [S.l.], 2010. p. 94–101. Citation on page [76](#).

LUCEY, P.; COHN, J. F.; PRKACHIN, K. M.; SOLOMON, P. E.; MATTHEWS, I. Painful data: The unbc-mcmaster shoulder pain expression archive database. In: **2011 IEEE International Conference on Automatic Face Gesture Recognition (FG)**. [S.l.: s.n.], 2011. p. 57–64. Citation on page 76.

LUSTERMANS, D. R.; AMIRRAJAB, S.; VETA, M.; BREEUWER, M.; SCANNELL, C. M. Optimized automated cardiac mr scar quantification with gan-based data augmentation. **Computer Methods and Programs in Biomedicine**, v. 226, p. 107116, 2022. ISSN 0169-2607. Available: <<https://www.sciencedirect.com/science/article/pii/S0169260722004977>>. Citation on page 61.

LUZ, S.; HAIDER, F.; GARCIA, S. de la F.; FROMM, D.; MACWHINNEY, B. **Alzheimer's dementia recognition through spontaneous speech**. [S.l.]: Frontiers Media SA, 2021. 780169 p. Citation on page 19.

LV, C.; FAN, L.; LI, H.; MA, J.; JIANG, W.; MA, X. Leveraging multimodal deep learning framework and a comprehensive audio-visual dataset to advance parkinson's detection. **Biomedical Signal Processing and Control**, Elsevier, v. 95, p. 106480, 2024. Citations on pages 42, 45, and 48.

MAHLKNECHT, P.; SEPPI, K.; POEWE, W. The concept of prodromal parkinson's disease. **Journal of Parkinson's disease**, IOS Press, v. 5, n. 4, p. 681–697, 2015. Citations on pages 32 and 33.

MAO, X.; LI, Q.; XIE, H.; LAU, R. Y.; WANG, Z.; SMOLLEY, S. P. Least squares generative adversarial networks. In: **Proceedings of the IEEE international conference on computer vision**. [S.l.: s.n.], 2017. p. 2794–2802. Citation on page 135.

MARRAS, C. *et al.* Prevalence of parkinson's disease across north america. **NPJ Parkinson's disease**, Nature Publishing Group UK London, v. 4, n. 1, p. 21, 2018. Citation on page 70.

MARSILI, L.; AGOSTINO, R.; BOLOGNA, M.; BELVISI, D.; PALMA, A.; FABBRINI, G.; BERARDELLI, A. Bradykinesia of posed smiling and voluntary movement of the lower face in parkinson's disease. **Parkinsonism & related disorders**, Elsevier, v. 20, n. 4, p. 370–375, 2014. Citation on page 33.

MASI, I.; WU, Y.; HASSNER, T.; NATARAJAN, P. Deep face recognition: A survey. In: **IEEE. 2018 31st SIBGRAPI conference on graphics, patterns and images (SIBGRAPI)**. [S.l.], 2018. p. 471–478. Citation on page 52.

MAVADATI, S. M.; MAHOOR, M. H.; BARTLETT, K.; TRINH, P.; COHN, J. F. Disfa: A spontaneous facial action intensity database. **IEEE Transactions on Affective Computing**, IEEE, v. 4, n. 2, p. 151–160, 2013. Citation on page 76.

MAYYA, V.; KULKARNI, U.; SURYA, D. K.; ACHARYA, U. R. An empirical study of preprocessing techniques with convolutional neural networks for accurate detection of chronic ocular diseases using fundus images. **Applied Intelligence**, Springer, v. 53, n. 2, p. 1548–1566, 2023. Citations on pages 130 and 131.

MCDONALD, J. H. **Handbook of Biological Statistics**. 3rd. ed. Baltimore, Maryland, USA: Sparky House Publishing, 2014. Citations on pages 111 and 112.

- MEI, J.; DESROSIERS, C.; FRASNELLI, J. Machine learning for the diagnosis of parkinson's disease: a review of literature. **Frontiers in aging neuroscience**, Frontiers Media SA, v. 13, p. 633752, 2021. Citation on page 97.
- MENG, C.; HE, Y.; SONG, Y.; SONG, J.; WU, J.; ZHU, J.-Y.; ERMON, S. Sdedit: Guided image synthesis and editing with stochastic differential equations. **arXiv preprint arXiv:2108.01073**, 2021. Citation on page 122.
- MERGL, R.; MAVROGIORGOU, P.; HEGERL, U.; JUCKEL, G. Kinematical analysis of emotionally induced facial expressions: a novel tool to investigate hypomimia in patients suffering from depression. **Journal of Neurology, Neurosurgery & Psychiatry**, BMJ Publishing Group Ltd, v. 76, n. 1, p. 138–140, 2005. Citation on page 33.
- MIKOŁAJCZYK, A.; GROCHOWSKI, M. Data augmentation for improving deep learning in image classification problem. In: IEEE. **2018 international interdisciplinary PhD workshop (IIPhDW)**. [S.l.], 2018. p. 117–122. Citation on page 61.
- MIRZA, M.; OSINDERO, S. Conditional generative adversarial nets. **arXiv preprint arXiv:1411.1784**, 2014. Citations on pages 62 and 135.
- MITTAL, V.; SHARMA, R. Classification of parkinson disease based on analysis and synthesis of voice signal. **International Journal of Healthcare Information Systems and Informatics (IJHISI)**, IGI Global, v. 16, n. 4, p. 1–22, 2021. Citation on page 57.
- MONSHIPOURI, M.; ALIAHMAD, B.; OGRIN, R.; ELDER, K.; ANDERSON, J.; POLUS, B.; KUMAR, D. Thermal imaging potential and limitations to predict healing of venous leg ulcers. **Scientific Reports**, Nature Publishing Group UK London, v. 11, n. 1, p. 13239, 2021. Citations on pages 58, 122, 123, and 128.
- MONTAGNA, S.; MARIANI, S.; PENGO, M. F. A chatbot-based recommendation framework for hypertensive patients. In: IEEE. **2023 IEEE 36th International Symposium on Computer-Based Medical Systems (CBMS)**. [S.l.], 2023. p. 730–733. Citation on page 114.
- MORO-VELAZQUEZ, L.; GOMEZ-GARCIA, J. A.; ARIAS-LONDOÑO, J. D.; DEHAK, N.; GODINO-LLORENTE, J. I. Advances in parkinson's disease detection and assessment using voice and speech: A review of the articulatory and phonatory aspects. **Biomedical Signal Processing and Control**, Elsevier, v. 66, p. 102418, 2021. Citations on pages 56 and 99.
- MORO-VELÁZQUEZ, L.; GÓMEZ-GARCÍA, J. A.; GODINO-LLORENTE, J. I.; VILLALBA, J.; OROZCO-ARROYAVE, J. R.; DEHAK, N. Analysis of speaker recognition methodologies and the influence of kinetic changes to automatically detect Parkinson's Disease. **Applied Soft Computing Journal**, Elsevier B.V., v. 62, p. 649–666, 2018. ISSN 1568-4946. Available: <<http://dx.doi.org/10.1016/j.asoc.2017.11.001>>. Citation on page 110.
- MORO-VELAZQUEZ, L.; GOMEZ-GARCIA, J. A.; GODINO-LLORENTE, J. I.; VILLALBA, J.; RUSZ, J.; SHATTUCK-HUFNAGEL, S.; DEHAK, N. A forced gaussians based methodology for the differential evaluation of parkinson's disease by means of speech processing. **Biomedical Signal Processing and Control**, Elsevier, v. 48, p. 205–220, 2019. Citation on page 97.
- MOTIN, M. A.; PAH, N. D.; RAGHAV, S.; KUMAR, D. K. Parkinson's disease detection using smartphone recorded phonemes in real world conditions. **IEEE Access**, IEEE, v. 10, p. 97600–97609, 2022. Citation on page 97.

MUELLER, C.; BALLARD, C.; CORBETT, A.; AARSLAND, D. The prognosis of dementia with lewy bodies. **The Lancet Neurology**, Elsevier, v. 16, n. 5, p. 390–398, 2017. Citation on page 54.

MULDOON, J. Timescaled monitoring to manage venous leg ulcers and chronic oedema. **British Journal of Community Nursing**, MA Healthcare London, v. 28, n. Sup9, p. S24–S30, 2023. Citation on page 121.

MULLEN, M.; LOOMIS, C. *et al.* Differentiating facial weakness caused by bell's palsy vs. acute stroke. **JEMS**, 2014. Citations on pages 87 and 88.

MUNSIF, M.; SAJJAD, M.; ULLAH, M.; TAREKEGN, A. N.; CHEIKH, F. A.; TSAKANIKAS, P.; MUHAMMAD, K. Optimized efficient attention-based network for facial expressions analysis in neurological health care. **Computers in Biology and Medicine**, Elsevier, v. 179, p. 108822, 2024. Citations on pages 42, 45, and 47.

NAEINI, S. A.; SIMMATIS, L.; JAFAR, D.; GUARIN, D. L.; YUNUSOVA, Y.; TAATI, B. Automated temporal segmentation of orofacial assessment videos. In: **IEEE. 2022 IEEE-EMBS international conference on biomedical and health informatics (BHI)**. [S.l.], 2022. p. 01–06. Citation on page 31.

NAGELS, L.; GAUDRAIN, E.; VICKERS, D.; HENDRIKS, P.; BAŞKENT, D. Prelingually deaf children with cochlear implants show better perception of voice cues and speech in competing speech than postlingually deaf adults with cochlear implants. **Ear and hearing**, LWW, v. 45, n. 4, p. 952–968, 2024. Citation on page 57.

NAGRANI, A.; YANG, S.; ARNAB, A.; JANSEN, A.; SCHMID, C.; SUN, C. Attention bottlenecks for multimodal fusion. **Advances in Neural Information Processing Systems**, v. 34, 2021. Citation on page 52.

NATEKIN, A.; KNOLL, A. Gradient boosting machines, a tutorial. **Frontiers in neurorobotics**, Frontiers Media SA, v. 7, p. 21, 2013. Citation on page 100.

NEWMAN-TOKER, D. E.; MOY, E.; VALENTE, E.; COFFEY, R.; HINES, A. L. Missed diagnosis of stroke in the emergency department: a cross-sectional analysis of a large population-based sample. **Diagnosis**, De Gruyter, v. 1, n. 2, p. 155–166, 2014. Citation on page 74.

NGO, Q. C.; MOTIN, M. A.; PAH, N. D.; DROTÁR, P.; KEMPSTER, P.; KUMAR, D. Computerized analysis of speech and voice for parkinson's disease: A systematic review. **Computer Methods and Programs in Biomedicine**, Elsevier, p. 107133, 2022. Citations on pages 97 and 114.

NGO, Q. C.; OGRIN, R.; KUMAR, D. K. Computerised prediction of healing for venous leg ulcers. **Scientific Reports**, Nature Publishing Group UK London, v. 12, n. 1, p. 17962, 2022. Citations on pages 58, 122, 123, and 128.

NGUYEN-HUYNH, M. N.; JOHNSTON, S. C. Evaluation and management of transient ischemic attack: an important component of stroke prevention. **Nature Clinical Practice Cardiovascular Medicine**, Nature Publishing Group UK London, v. 4, n. 6, p. 310–318, 2007. Citation on page 74.

NOVA, K. Generative ai in healthcare: advancements in electronic health records, facilitating medical languages, and personalized patient care. **Journal of Advanced Analytics in Healthcare Management**, v. 7, n. 1, p. 115–131, 2023. Citation on page 20.

NOVOTNY, M.; TYKALOVA, T.; RUZICKOVA, H.; RUZICKA, E.; DUSEK, P.; RUSZ, J. Automated video-based assessment of facial bradykinesia in de-novo parkinson's disease. **NPJ digital medicine**, Nature Publishing Group UK London, v. 5, n. 1, p. 98, 2022. Citations on pages [61](#) and [72](#).

ODENA, A.; OLAH, C.; SHLENS, J. Conditional image synthesis with auxiliary classifier gans. In: PMLR. **International conference on machine learning**. [S.l.], 2017. p. 2642–2651. Citation on page [135](#).

OGRIN, R.; MOTIN, M. A.; ALIAHMAD, B.; ELDER, K.; ANDERSON, J.; KUMAR, D. Can thermal imaging technique be used to predict the healing status of a venous leg ulcer? **The International Journal of Lower Extremity Wounds**, SAGE Publications Sage CA: Los Angeles, CA, v. 22, n. 1, p. 85–92, 2023. Citations on pages [122](#) and [123](#).

OH, S.-i.; OH, K.-W.; KIM, H.-J.; PARK, J.-S.; KIM, S. H. Impaired perception of emotional expression in amyotrophic lateral sclerosis. **Journal of Clinical Neurology**, Korean Neurological Association, v. 12, n. 3, p. 295–300, 2016. Citations on pages [31](#) and [74](#).

OJALA, T.; PIETIKAINEN, M.; MAENPAA, T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. **IEEE Transactions on pattern analysis and machine intelligence**, IEEE, v. 24, n. 7, p. 971–987, 2002. Citation on page [39](#).

OLIVEIRA, G. C.; NGO, Q. C.; PASSOS, L. A.; PAPA, J. P.; JODAS, D.; KUMAR, D. Tabular data augmentation for video-based detection of hypomimia in parkinson's disease. **Computer Methods and Programs in Biomedicine**, Elsevier, p. 107713, 2023. Citations on pages [42](#), [45](#), [47](#), [49](#), [51](#), [74](#), [82](#), [97](#), [119](#), and [120](#).

OLIVEIRA, G. C.; NGO, Q. C.; PASSOS, L. A.; OLIVEIRA, L. S.; PAPA, J. P.; KUMAR, D. Facial expressions to identify post-stroke: A pilot study. **Computer Methods and Programs in Biomedicine**, Elsevier, p. 108195, 2024. Citations on pages [119](#) and [120](#).

OLIVEIRA, G. C.; PAH, N. D.; NGO, Q. C.; PAPA, J. P.; KUMAR, D. Nestneuro: Leveraging chatbots for vocal screening. In: IEEE. **2024 IEEE 37th International Symposium on Computer-Based Medical Systems (CBMS)**. [S.l.], 2024. p. 182–185. Citation on page [50](#).

OMAROV, B.; NARYNOV, S.; ZHUMANOV, Z. Artificial intelligence-enabled chatbots in mental health: A systematic review. **Computers, Materials & Continua**, v. 74, n. 3, 2023. Citation on page [114](#).

OMOREGBE, N. A.; NDAMAN, I. O.; MISRA, S.; ABAYOMI-ALLI, O. O.; DAMAŠEVIČIUS, R.; DOGRA, A. Text messaging-based medical diagnosis using natural language processing and fuzzy logic. **Journal of Healthcare Engineering**, Hindawi Limited, v. 2020, p. 1–14, 2020. Citation on page [114](#).

OROZCO-ARROYAVE, J. R.; ARIAS-LONDOÑO, J. D.; VARGAS-BONILLA, J. F.; GONZALEZ-RÁTIVA, M. C.; NÖTH, E. New spanish speech corpus database for the analysis of people suffering from parkinson's disease. In: **LREC**. [S.l.: s.n.], 2014. p. 342–347. Citations on pages [55](#) and [98](#).

OROZCO-ARROYAVE, J. R.; VÁSQUEZ-CORREA, J. C.; VARGAS-BONILLA, J. F.; ARORA, R.; DEHAK, N.; NIDADAVOLU, P. S.; CHRISTENSEN, H.; RUDZICZ, F.; YANCHEVA, M.; CHINAEI, H. *et al.* Neurospeech: An open-source software for parkinson's

speech analysis. **Digital Signal Processing**, Elsevier, v. 77, p. 207–221, 2018. Citation on page 100.

PACHADE, S.; PORWAL, P.; THULKAR, D.; KOKARE, M.; DESHMUKH, G.; SAHASRABUDDHE, V.; GIANCARDIO, L.; QUELLEC, G.; MÉRIAUDEAU, F. Retinal fundus multi-disease image dataset (rfmid): A dataset for multi-disease detection research. **Data**, Multi-disciplinary Digital Publishing Institute, v. 6, n. 2, p. 14, 2021. Citation on page 132.

PAH, N. D.; INDRAWATI, V.; KUMAR, D. K. Voice Features of Sustained Phoneme as COVID-19 Biomarker. **IEEE Journal of Translational Engineering in Health and Medicine**, IEEE, v. 10, n. July, p. 1–9, 2022. ISSN 21682372. Citation on page 113.

PAH, N. D.; MOTIN, M. A.; KEMPSTER, P.; KUMAR, D. K. Detecting effect of levodopa in parkinson's disease patients using sustained phonemes. **IEEE JTEHM**, IEEE, v. 9, p. 1–9, 2021. Citations on pages 56, 99, 106, 108, and 115.

PAH, N. D.; MOTIN, M. A.; KUMAR, D. K. Phonemes based detection of parkinson's disease for telehealth applications. **Scientific Reports**, Nature Publishing Group UK, v. 12, n. 1, p. 1–9, 2022. ISSN 20452322. Available: <<https://doi.org/10.1038/s41598-022-13865-z>>. Citations on pages 112 and 113.

\_\_\_\_\_. Phonemes based detection of parkinson's disease for telehealth applications. **Scientific Reports**, Nature Publishing Group UK London, v. 12, n. 1, p. 9687, 2022. Citations on pages 57, 97, 106, 114, and 115.

PAH, N. D.; MOTIN, M. A.; OLIVEIRA, G. C.; KUMAR, D. K. The change of vocal tract length in people with parkinson's disease. In: IEEE. **2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)**. [S.l.], 2023. p. 1–4. Citations on pages 57 and 97.

\_\_\_\_\_. The change of vocal tract length in people with parkinson's disease. In: IEEE. **2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)**. [S.l.], 2023. p. 1–4. Citation on page 115.

PAL, M. Random forest classifier for remote sensing classification. **International journal of remote sensing**, Taylor & Francis, v. 26, n. 1, p. 217–222, 2005. Citations on pages 63 and 64.

PALMERINI, L.; ROCCHI, L.; MELLONE, S.; VALZANIA, F.; CHIARI, L. Feature selection for accelerometer-based posture analysis in parkinson's disease. **IEEE Transactions on Information Technology in Biomedicine**, v. 15, n. 3, p. 481–490, 2011. Citation on page 97.

PAN, Z.; YU, W.; YI, X.; KHAN, A.; YUAN, F.; ZHENG, Y. Recent progress on generative adversarial networks (gans): A survey. **IEEE access**, IEEE, v. 7, p. 36322–36333, 2019. Citation on page 127.

PARRA-DOMINGUEZ, G. S.; SANCHEZ-YANEZ, R. E.; GARCIA-CAPULIN, C. H. Facial paralysis detection on images using key point analysis. **Applied Sciences**, MDPI, v. 11, n. 5, p. 2435, 2021. Citations on pages 31 and 85.

PASSOS, L. A.; JODAS, D.; RIBEIRO, L.; MOREIRA, T.; PAPA, J. P. O<sup>2</sup>pf: Oversampling via optimum-path forest for breast cancer detection. In: IEEE. **2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS)**. [S.l.], 2020. p. 498–503. Citation on page 63.

PASSOS, L. A.; JODAS, D. S.; RIBEIRO, L. C.; AKIO, M.; SOUZA, A. N. D.; PAPA, J. P. Handling imbalanced datasets through optimum-path forest. **Knowledge-Based Systems**, Elsevier, v. 242, p. 108445, 2022. Citation on page 63.

PASSOS, L. A.; PAPA, J. P.; ADEEL, A. Canonical cortical graph neural networks and its application for speech enhancement in future audio-visual hearing aids. **arXiv preprint arXiv:2206.02671**, 2022. Citation on page 52.

PASSOS, L. A.; PAPA, J. P.; SER, J. D.; HUSSAIN, A.; ADEEL, A. **Multimodal Audio-Visual Information Fusion using Canonical-Correlated Graph Neural Network for Energy-Efficient Speech Enhancement**. arXiv, 2022. Available: <<https://arxiv.org/abs/2202.04528>>. Citation on page 52.

PASZKE, A.; GROSS, S.; MASSA, F.; LERER, A.; BRADBURY, J.; CHANAN, G.; KILLEEN, T.; LIN, Z.; GIMELSHEIN, N.; ANTIGA, L. *et al.* Pytorch: An imperative style, high-performance deep learning library. **Advances in neural information processing systems**, v. 32, 2019. Citation on page 135.

PEAD, E.; MEGAW, R.; CAMERON, J.; FLEMING, A.; DHILLON, B.; TRUCCO, E.; MACGILLIVRAY, T. Automated detection of age-related macular degeneration in color fundus photography: a systematic review. **survey of ophthalmology**, Elsevier, v. 64, n. 4, p. 498–511, 2019. Citation on page 141.

PEDREGOSA, F.; VAROQUAUX, G.; GRAMFORT, A.; MICHEL, V.; THIRION, B.; GRISEL, O.; BLONDEL, M.; PRETTENHOFER, P.; WEISS, R.; DUBOURG, V. *et al.* Scikit-learn: Machine learning in python. **the Journal of machine Learning research**, JMLR. org, v. 12, p. 2825–2830, 2011. Citation on page 63.

\_\_\_\_\_. Scikit-learn: Machine learning in python. **the Journal of machine Learning research**, JMLR. org, v. 12, p. 2825–2830, 2011. Citations on pages 78, 101, and 116.

PEGOLO, E.; VOLPE, D.; CUCCA, A.; RICCIARDI, L.; SAWACHA, Z. Quantitative evaluation of hypomimia in parkinson’s disease: A face tracking approach. **Sensors**, MDPI, v. 22, n. 4, p. 1358, 2022. Citations on pages 41, 44, and 47.

PELL, M. D.; LEONARD, C. L. Facial expression decoding in early parkinson’s disease. **Cognitive Brain Research**, Elsevier, v. 23, n. 2-3, p. 327–340, 2005. Citation on page 33.

PEREIRA, T. A.; POPIM, R. C.; PASSOS, L. A.; PEREIRA, D. R.; PEREIRA, C. R.; PAPA, J. P. ComplexWoundDB: A database for automatic complex wound tissue categorization. In: **2022 29th International Conference on Systems, Signals and Image Processing (IWSSIP)**. [S.l.: s.n.], 2022. CFP2255E-ART, p. 1–4. Citation on page 121.

PEREZ, L.; WANG, J. The effectiveness of data augmentation in image classification using deep learning. **arXiv preprint arXiv:1712.04621**, 2017. Citation on page 59.

PÉREZ-TORO, P. A.; VÁSQUEZ-CORREA, J. C.; STRAUSS, M.; OROZCO-ARROYAVE, J. R.; NÖTH, E. Natural language analysis to detect parkinson’s disease. In: **SPRINGER. Text, Speech, and Dialogue: 22nd International Conference, TSD 2019, Ljubljana, Slovenia, September 11–13, 2019, Proceedings 22**. [S.l.], 2019. p. 82–90. Citation on page 97.

PETERSON, G. E.; BARNEY, H. L. Control Methods Used in a Study of the Vowels. **Journal of the Acoustical Society of America**, v. 24, n. 2, p. 175–184, 1952. ISSN NA. Citation on page 110.

PHOKAEWVARANGKUL, O.; VATEEKUL, P.; SURANGSRIRAT, D.; BHIDAYASIRI, R. Deep learning approach for parkinson's disease classification from facial expressions: A pilot feasibility study. In: WILEY 111 RIVER ST, HOBOKEN 07030-5774, NJ USA. **MOVEMENT DISORDERS**. [S.l.], 2020. v. 35, p. S660–S660. Citation on page 61.

PILLAI, A. S. Utilizing deep learning in medical image analysis for enhanced diagnostic accuracy and patient care: Challenges, opportunities, and ethical implications. **Journal of Deep Learning in Genomic Data Analysis**, v. 1, n. 1, p. 1–17, 2021. Citation on page 19.

PISANSKI, K.; CARTEI, V.; MCGETTIGAN, C.; RAINE, J.; REBY, D. Voice modulation: a window into the origins of human vocal control? **Trends in cognitive sciences**, Elsevier, v. 20, n. 4, p. 304–318, 2016. Citation on page 56.

\_\_\_\_\_. Voice Modulation: A Window into the Origins of Human Vocal Control ? **Trends in Cognitive Sciences**, Elsevier Ltd, v. 20, n. 4, p. 304–318, 2016. ISSN 1364-6613. Available: <<http://dx.doi.org/10.1016/j.tics.2016.01.002>>. Citation on page 113.

PISANSKI, K.; FRACCARO, P. J.; TIGUE, C. C.; O'CONNOR, J. J.; RÖDER, S.; ANDREWS, P. W.; FINK, B.; DEBRUINE, L. M.; JONES, B. C.; FEINBERG, D. R. Vocal indicators of body size in men and women: a meta-analysis. **Animal Behaviour**, Elsevier, v. 95, p. 89–99, 2014. Citations on pages 56 and 115.

POEWE, W.; SEPPI, K.; TANNER, C. M.; HALLIDAY, G. M.; BRUNDIN, P.; VOLKMANN, J.; SCHRAG, A.-E.; LANG, A. E. Parkinson disease. **Nature reviews Disease primers**, Nature Publishing Group, v. 3, n. 1, p. 1–21, 2017. Citation on page 33.

POSTUMA, R. B.; BERG, D.; STERN, M.; POEWE, W.; OLANOW, C. W.; OERTEL, W.; OBESO, J.; MAREK, K.; LITVAN, I.; LANG, A. E. *et al.* Mds clinical diagnostic criteria for parkinson's disease. **Movement disorders**, Wiley Online Library, v. 30, n. 12, p. 1591–1601, 2015. Citations on pages 32 and 54.

QIANG, J.; WU, D.; DU, H.; ZHU, H.; CHEN, S.; PAN, H. Review on facial-recognition-based applications in disease diagnosis. **Bioengineering**, MDPI, v. 9, n. 7, p. 273, 2022. Citation on page 20.

RADFORD, A.; METZ, L.; CHINTALA, S. Unsupervised representation learning with deep convolutional generative adversarial networks. **arXiv preprint arXiv:1511.06434**, 2015. Citation on page 135.

RAJAVEL, R.; RAVICHANDRAN, S. K.; HARIMOORTHY, K.; NAGAPPAN, P.; GOBICHET-TIPALAYAM, K. R. Iot-based smart healthcare video surveillance system using edge computing. **Journal of ambient intelligence and humanized computing**, Springer, v. 13, n. 6, p. 3195–3207, 2022. Citation on page 30.

RAJNOHA, M.; MEKYSKA, J.; BURGET, R.; ELIASOVA, I.; KOSTALOVA, M.; REKTOROVÁ, I. Towards identification of hypomimia in parkinson's disease based on face recognition methods. In: IEEE. **2018 10th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)**. [S.l.], 2018. p. 1–4. Citations on pages 36, 61, and 72.

RANA, A.; DUMKA, A.; SINGH, R.; PANDA, M. K.; PRIYADARSHI, N. A computerized analysis with machine learning techniques for the diagnosis of parkinson's disease: Past studies and future perspectives. **Diagnostics**, MDPI, v. 12, n. 11, p. 2708, 2022. Citation on page 97.

RANJAN, P.; LASALA, A.; RUSCELLI, A. L.; SAHU, S. K.; GUARÍN, D. L.; MOCCIA, S.; CASTOLDI, P.; MICERA, S.; BANDINI, A. Facial asymmetry classification in neurological disorders: Integrating computer vision and machine learning for improved patient care. In: **IEEE. 2024 IEEE 8th Forum on Research and Technologies for Society and Industry Innovation (RTSI)**. [S.l.], 2024. p. 196–201. Citation on page [32](#).

RAZZOUKI, A. F.; JEANCOLAS, L.; MANGONE, G.; SAMBIN, S.; CHALANÇON, A.; GOMES, M.; LEHÉRICY, S.; CORVOL, J.-C.; VIDAILHET, M.; ARNULF, I.; EL-YACOUBI, M. A.; PETROVSKA-DELACRÉTAZ, D. Early-stage parkinson's disease detection based on optical flow and video vision transformer. In: **2024 16th International Conference on Human System Interaction (HSI)**. [S.l.: s.n.], 2024. p. 1–6. Citations on pages [42](#), [45](#), and [48](#).

REEVE, A.; SIMCOX, E.; TURNBULL, D. Ageing and parkinson's disease: why is advancing age the biggest risk factor? **Ageing research reviews**, Elsevier, v. 14, p. 19–30, 2014. Citation on page [32](#).

REFAEILZADEH, P.; TANG, L.; LIU, H. Cross-validation. **Encyclopedia of database systems**, Springer, v. 5, p. 532–538, 2009. Citation on page [64](#).

RICCIARDI, L.; ANGELIS, A. D.; MARSILI, L.; FAIMAN, I.; PRADHAN, P.; PEREIRA, E.; EDWARDS, M.; MORGANTE, F.; BOLOGNA, M. Hypomimia in parkinson's disease: an axial sign responsive to levodopa. **European Journal of Neurology**, Wiley Online Library, v. 27, n. 12, p. 2422–2429, 2020. Citation on page [33](#).

RICCIARDI, L.; VISCO-COMANDINI, F.; ERRO, R.; MORGANTE, F.; BOLOGNA, M.; FASANO, A.; RICCIARDI, D.; EDWARDS, M. J.; KILNER, J. Facial emotion recognition and expression in parkinson's disease: an emotional mirror mechanism? **PloS one**, Public Library of Science San Francisco, CA USA, v. 12, n. 1, p. e0169110, 2017. Citation on page [33](#).

RIZEK, P.; KUMAR, N.; JOG, M. S. An update on the diagnosis and treatment of parkinson disease. **Cmaj**, Can Med Assoc, v. 188, n. 16, p. 1157–1165, 2016. Citation on page [32](#).

ROMBACH, R.; BLATTMANN, A.; LORENZ, D.; ESSER, P.; OMMER, B. High-resolution image synthesis with latent diffusion models. In: **Proceedings of the IEEE/CVF conference on computer vision and pattern recognition**. [S.l.: s.n.], 2022. p. 10684–10695. Citation on page [122](#).

RUIZ, N.; LI, Y.; JAMPANI, V.; PRITCH, Y.; RUBINSTEIN, M.; ABERMAN, K. Dreambooth: Fine tuning text-to-image diffusion models for subject-driven generation. In: **Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition**. [S.l.: s.n.], 2023. p. 22500–22510. Citation on page [122](#).

RUSZ, J.; CMEJLA, R.; RUZICKOVA, H.; RUZICKA, E. Quantitative acoustic measurements for characterization of speech and voice disorders in early untreated Parkinson's disease. **The Journal of the Acoustical Society of America**, v. 129, n. 1, p. 350–367, 2011. Citation on page [109](#).

\_\_\_\_\_. Quantitative acoustic measurements for characterization of speech and voice disorders in early untreated parkinson's disease. **The journal of the Acoustical Society of America**, AIP Publishing, v. 129, n. 1, p. 350–367, 2011. Citations on pages [56](#) and [99](#).

RUSZ, J.; HLAVNIČKA, J.; NOVOTNÝ, M.; TYKALOVÁ, T.; PELLETIER, A.; MONTPLAISIR, J.; GAGNON, J.-F.; DUŠEK, P.; GALBIATI, A.; MARELLI, S. *et al.* Speech biomarkers in rapid eye movement sleep behavior disorder and parkinson disease. **Annals of neurology**, Wiley Online Library, v. 90, n. 1, p. 62–75, 2021. Citation on page 97.

SAKAR, B. E.; ISENKUL, M. E.; SAKAR, C. O.; SERTBAS, A.; GURGEN, F.; DELIL, S.; APAYDIN, H.; KURSUN, O. Collection and Analysis of a Parkinson Speech Dataset With Multiple Types of Sound Recordings. **IEEE Journal of Biomedical and Health Informatics**, v. 17, n. 4, p. 828–834, 2013. Available: <<https://www.ncbi.nlm.nih.gov/pubmed/25055311>>. Citation on page 110.

SAKAR, B. E.; SERBES, G.; SAKAR, C. O. Analyzing the effectiveness of vocal features in early telediagnosis of parkinson's disease. **PloS one**, Public Library of Science San Francisco, CA USA, v. 12, n. 8, p. e0182428, 2017. Citations on pages 97 and 108.

SAKAR, C. O.; SERBES, G.; GUNDUZ, A.; TUNC, H. C.; NIZAM, H.; SAKAR, B. E.; TUTUNCU, M.; AYDIN, T.; ISENKUL, M. E.; APAYDIN, H. A comparative analysis of speech signal processing algorithms for Parkinson's disease classification and the use of the tunable Q-factor wavelet transform. **Applied Soft Computing Journal**, Elsevier B.V., v. 74, p. 255–263, 2019. ISSN 15684946. Available: <<https://doi.org/10.1016/j.asoc.2018.10.022>>. Citation on page 110.

SALEHINEJAD, H.; VALAEE, S.; DOWDELL, T.; COLAK, E.; BARFETT, J. Generalization of deep neural networks for chest pathology classification in x-rays using generative adversarial networks. In: IEEE. **2018 IEEE international conference on acoustics, speech and signal processing (ICASSP)**. [S.l.], 2018. p. 990–994. Citation on page 59.

SANHUEZA-GARRIDO, M.; ROJAS-ZEPEDA, C.; GARCÍA-FLORES, V. Use of the diadochokinetic index ddkccvp% in the detection of articulatory inaccuracies in parkinson's disease: a preliminary exploratory study. **Revista CEFAC**, SciELO Brasil, v. 25, p. e4723, 2023. Citation on page 56.

SCHIMMEL, M.; LEEMANN, B.; CHRISTOU, P.; KILIARIDIS, S.; HERRMANN, F.; MÜLLER, F. Quantitative assessment of facial muscle impairment in patients with hemispheric stroke. **Journal of Oral Rehabilitation**, Wiley Online Library, v. 38, n. 11, p. 800–809, 2011. Citation on page 82.

SCHIMMEL, M.; ONO, T.; LAM, O.; MÜLLER, F. Oro-facial impairment in stroke patients. **Journal of oral rehabilitation**, Wiley Online Library, v. 44, n. 4, p. 313–326, 2017. Citations on pages 82, 87, and 88.

SEABOLD, S.; PERKTOLD, J. Statsmodels: Econometric and statistical modeling with python. In: AUSTIN, TX. **Proceedings of the 9th Python in Science Conference**. [S.l.], 2010. v. 57, p. 10–25080. Citation on page 78.

SELVARAJU, R. R.; COGSWELL, M.; DAS, A.; VEDANTAM, R.; PARIKH, D.; BATRA, D. Grad-cam: Visual explanations from deep networks via gradient-based localization. In: **Proceedings of the IEEE international conference on computer vision**. [S.l.: s.n.], 2017. p. 618–626. Citation on page 140.

SEPPI, K.; CHAUDHURI, K. R.; COELHO, M.; FOX, S.; KATZENSCHLAGER, R.; LLORET, S. P.; WEINTRAUB, D.; SAMPAIO, C. the collaborators of the parkinson's disease update on

non-motor symptoms study group on behalf of the movement disorders society evidence-based medicine committee. update on treatments for nonmotor symptoms of parkinson's disease-an evidence-based medicine review. **Mov Disord**, v. 34, n. 2, p. 180–198, 2019. Citations on pages 32 and 33.

SHAKE, A.; OOSTEMA, A.; JONES, J. Bet 2: Missed diagnosis of ischaemic stroke in the emergency department. **Emergency Medicine Journal**, BMJ Publishing Group Ltd and the British Association for Accident . . . , v. 35, n. 12, p. 768–769, 2018. Citation on page 74.

SHORTEN, C.; KHOSHGOFTAAR, T. M. A survey on image data augmentation for deep learning. **Journal of Big Data**, Springer, v. 6, n. 1, p. 1–48, 2019. Citation on page 61.

SILBERGLEIT, A. K.; LEWITT, P. A.; PETERSON, E. L.; GARDNER, G. M. Quantitative Analysis of Voice in Parkinson Disease Compared to Motor Performance : A Pilot Study. **Journal of Parkinson's Disease**, v. 5, p. 517–524, 2015. Citation on page 109.

SIMONET, C.; SCHRAG, A.; LEES, A. J.; NOYCE, A. J. The motor prodromes of parkinson's disease: from bedside observation to large-scale application. **J Neurol**, p. 1–10, 2019. Available: <<https://www.ncbi.nlm.nih.gov/pubmed/31802219>>. Citation on page 109.

SIMONS, G.; PASQUALINI, M. C. S.; REDDY, V.; WOOD, J. Emotional and nonemotional facial expressions in people with parkinson's disease. **Journal of the International Neuropsychological Society**, Cambridge University Press, v. 10, n. 4, p. 521–535, 2004. Citation on page 34.

SONAWANE, B.; SHARMA, P. Review of automated emotion-based quantification of facial expression in parkinson's patients. **The Visual Computer**, Springer, v. 37, p. 1151–1167, 2021. Citations on pages 33, 39, 44, 46, 49, 51, 61, and 72.

SU, G.; LIN, B.; LUO, W.; YIN, J.; DENG, S.; GAO, H.; XU, R. Hypomimia recognition in parkinson's disease with semantic features. **ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)**, ACM New York, NY, v. 17, n. 3s, p. 1–20, 2021. Citations on pages 41, 44, 47, and 48.

SU, G.; LIN, B.; YIN, J.; LUO, W.; XU, R.; XU, J.; DONG, K. Detection of hypomimia in patients with parkinson's disease via smile videos. **Annals of Translational Medicine**, AME Publications, v. 9, n. 16, 2021. Citations on pages 41, 44, 47, and 48.

SUÁREZ-HERNÁNDEZ<sup>1</sup>, D.; SANTOS-ARCE<sup>1</sup>, S. R.; TORRES-RAMOS<sup>1</sup>, S.; SALIDORUIZ<sup>1</sup>, R. A. Amyotrophic lateral sclerosis detection using facial symmetry analysis with machine learning. In: SPRINGER NATURE. **XLVII Mexican Conference on Biomedical Engineering: Proceedings of CNIB 2024, November 7–9, 2024, Hermosillo, Sonora, México-Volume 1: Signal Processing And Bioinformatics**Congreso Nacional de Ingeniería Biomédica CNIBHermosillo, Mexico, November 6–8, 2024. [S.l.], 2024. p. 239. Citation on page 31.

SULLIVAN, G. M.; FEINN, R. Using Effect Size—or Why the P Value Is Not Enough . **Journal of Graduate Medical Education**, v. 4, n. 3, p. 279–282, 2012. ISSN 1949-8349. Citation on page 111.

SVEINBJORNSDOTTIR, S. The clinical symptoms of parkinson's disease. **Journal of neurochemistry**, Wiley Online Library, v. 139, p. 318–324, 2016. Citation on page 33.

TAN, Z.-H. *et al.* Vocal tract length perturbation for text-dependent speaker verification with autoregressive prediction coding. **IEEE Signal Processing Letters**, IEEE, v. 28, p. 364–368, 2021. Citation on page 115.

THAKOOR, K. A.; CARTER, A.; SONG, G.; WAX, A.; MOUSSA, O.; CHEN, R. W.; HENDON, C.; SAJDA, P. Enhancing portable oct image quality via gans for ai-based eye disease detection. In: SPRINGER. **International Workshop on Distributed, Collaborative, and Federated Learning**. [S.l.], 2022. p. 155–167. Citation on page 59.

TICKLE-DEGNEN, L.; LYONS, K. D. Practitioners' impressions of patients with parkinson's disease: the social ecology of the expressive mask. **Social Science & Medicine**, Elsevier, v. 58, n. 3, p. 603–614, 2004. Citation on page 34.

TJOA, E.; GUAN, C. A survey on explainable artificial intelligence (xai): Toward medical xai. **IEEE transactions on neural networks and learning systems**, IEEE, v. 32, n. 11, p. 4793–4813, 2020. Citation on page 52.

TURNER, R. E.; WALTERS, T. C.; MONAGHAN, J. J. M.; PATTERSON, R. D. A statistical, formant-pattern model for segregating vowel type and vocal-tract length in developmental formant data. **The Journal of the Acoustical Society of America**, v. 125, n. 4, p. 2374–2386, 2009. ISSN 0001-4966. Citation on page 110.

TYSNES, O.-B.; STORSTEIN, A. Epidemiology of parkinson's disease. **Journal of neural transmission**, Springer, v. 124, p. 901–905, 2017. Citation on page 70.

VAICIUKYNAS, E.; VERIKAS, A.; GELZINIS, A.; BACAUSKIENE, M. Detecting Parkinson's disease from sustained phonation and speech signals. **PloS one**, v. 12, n. 10, p. e0185613, 2017. Available: <<https://www.ncbi.nlm.nih.gov/pubmed/28982171>>. Citation on page 109.

VALENZUELA, B.; AREVALO, J.; CONTRERAS, W.; MARTINEZ, F. A spatio-temporal hypomimic deep descriptor to discriminate parkinsonian patients. In: IEEE. **2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)**. [S.l.], 2022. p. 4192–4195. Citations on pages 41, 45, 47, 61, and 72.

VÁSQUEZ-CORREA, J. C.; ARIAS-VERGARA, T.; OROZCO-ARROYAVE, J. R.; ESKOFIER, B.; KLUCKEN, J.; NÖTH, E. Multimodal assessment of parkinson's disease: a deep learning approach. **IEEE journal of biomedical and health informatics**, IEEE, v. 23, n. 4, p. 1618–1630, 2018. Citations on pages 97 and 106.

VÁSQUEZ-CORREA, J. C.; OROZCO-ARROYAVE, J.; BOCKLET, T.; NÖTH, E. Towards an automatic evaluation of the dysarthria level of patients with parkinson's disease. **Journal of communication disorders**, Elsevier, v. 76, p. 21–36, 2018. Citations on pages 56, 99, and 100.

VETURI, Y. A.; WOOF, W.; LAZEBNIK, T.; MOGHUL, I.; WOODWARD-COURT, P.; WAGNER, S. K.; GUIMARÃES, T. A. C. de; VARELA, M. D.; LIEFERS, B.; PATEL, P. J. *et al.* Syntheye: Investigating the impact of synthetic data on artificial intelligence-assisted gene diagnosis of inherited retinal disease. **Ophthalmology Science**, Elsevier, v. 3, n. 2, p. 100258, 2023. Citation on page 131.

VISWANATHAN, R.; ARJUNAN, S. P.; BINGHAM, A.; JELFS, B.; KEMPSTER, P.; RAGHAV, S.; KUMAR, D. K. Complexity measures of voice recordings as a discriminative tool for parkinson's disease. **Biosensors**, MDPI, v. 10, n. 1, p. 1, 2019. Citation on page 115.

WAHL, B.; COSSY-GANTNER, A.; GERMANN, S.; SCHWALBE, N. R. Artificial intelligence (ai) and global health: how can ai contribute to health in resource-poor settings? **BMJ global health**, *BMJ Specialist Journals*, v. 3, n. 4, p. e000798, 2018. Citation on page 19.

WANG, G.; LI, W.; AERTSEN, M.; DEPREST, J.; OURSELIN, S.; VERCAUTEREN, T. Aleatoric uncertainty estimation with test-time augmentation for medical image segmentation with convolutional neural networks. **Neurocomputing**, Elsevier, v. 338, p. 34–45, 2019. Citation on page 62.

WANG, J.; PEREZ, L. *et al.* The effectiveness of data augmentation in image classification using deep learning. **Convolutional Neural Networks Vis. Recognit**, v. 11, p. 1–8, 2017. Citation on page 61.

WANG, M.; DENG, W. Deep face recognition: A survey. **Neurocomputing**, Elsevier, v. 429, p. 215–244, 2021. Citation on page 52.

WANG, Y.-X.; GIRSHICK, R.; HEBERT, M.; HARIHARAN, B. Low-shot learning from imaginary data. In: **Proceedings of the IEEE conference on computer vision and pattern recognition**. [S.l.: s.n.], 2018. p. 7278–7286. Citation on page 61.

WANG, Z.; BOVIK, A. C.; SHEIKH, H. R.; SIMONCELLI, E. P. Image quality assessment: from error visibility to structural similarity. **IEEE transactions on image processing**, IEEE, v. 13, n. 4, p. 600–612, 2004. Citation on page 125.

WANG, Z.; LIM, G.; NG, W. Y.; TAN, T.-E.; LIM, J.; LIM, S. H.; FOO, V.; LIM, J.; SINISTERRA, L. G.; ZHENG, F. *et al.* Synthetic artificial intelligence using generative adversarial network for retinal imaging in detection of age-related macular degeneration. **Frontiers in Medicine**, Frontiers, v. 10, p. 1184892, 2023. Citations on pages 130 and 131.

WIJESEKERA, L. C.; LEIGH, P. N. Amyotrophic lateral sclerosis. **Orphanet journal of rare diseases**, Springer, v. 4, p. 1–22, 2009. Citation on page 74.

WOLF, L.; HASSNER, T.; MAOZ, I. Face recognition in unconstrained videos with matched background similarity. In: IEEE. **CVPR 2011**. [S.l.], 2011. p. 529–534. Citation on page 39.

WONG, W. L.; SU, X.; LI, X.; CHEUNG, C. M. G.; KLEIN, R.; CHENG, C.-Y.; WONG, T. Y. Global prevalence of age-related macular degeneration and disease burden projection for 2020 and 2040: a systematic review and meta-analysis. **The Lancet Global Health**, Elsevier, v. 2, n. 2, p. e106–e116, 2014. Citation on page 130.

WU, P.; GONZALEZ, I.; PATSIS, G.; JIANG, D.; SAHLI, H.; KERCKHOFS, E.; VANDEKERCKHOVE, M. Objectifying facial expressivity assessment of parkinson’s patients: preliminary study. **Computational and mathematical methods in medicine**, Hindawi, v. 2014, 2014. Citation on page 40.

WU, Y.; JI, Q. Facial landmark detection: A literature survey. **International Journal of Computer Vision**, Springer, v. 127, n. 2, p. 115–142, 2019. Citation on page 88.

WU, Z.; PAN, S.; CHEN, F.; LONG, G.; ZHANG, C.; PHILIP, S. Y. A comprehensive survey on graph neural networks. **IEEE transactions on neural networks and learning systems**, IEEE, v. 32, n. 1, p. 4–24, 2020. Citation on page 83.

XI, Z.; CHEN, W.; GUO, X.; HE, W.; DING, Y.; HONG, B.; ZHANG, M.; WANG, J.; JIN, S.; ZHOU, E. *et al.* The rise and potential of large language model based agents: A survey. **arXiv preprint arXiv:2309.07864**, 2023. Citation on page 114.

XU, L.; SKOULARIDOU, M.; CUESTA-INFANTE, A.; VEERAMACHANENI, K. Modeling tabular data using conditional gan. **Advances in Neural Information Processing Systems**, v. 32, 2019. Citation on page 62.

XU, R.-S.; YUAN, M. Considerations on the concept, definition, and diagnosis of amyotrophic lateral sclerosis. **Neural Regeneration Research**, Medknow, v. 16, n. 9, p. 1723–1729, 2021. Citation on page 88.

XU, Z.; LV, D.; LI, H.; LI, H.; GAO, H. Application of reslstm in hypomimia video detection for parkinson's disease. In: IEEE. **2023 International Conference on New Trends in Computational Intelligence (NTCI)**. [S.l.], 2023. v. 1, p. 243–247. Citations on pages 42, 45, and 47.

YANG, J.; ZHANG, D.; FRANGI, A. F.; YANG, J.-y. Two-dimensional pca: a new approach to appearance-based face representation and recognition. **IEEE transactions on pattern analysis and machine intelligence**, IEEE, v. 26, n. 1, p. 131–137, 2004. Citation on page 43.

YANG, L.; ZHANG, Z.; SONG, Y.; HONG, S.; XU, R.; ZHAO, Y.; ZHANG, W.; CUI, B.; YANG, M.-H. Diffusion models: A comprehensive survey of methods and applications. **ACM Computing Surveys**, ACM New York, NY, USA, v. 56, n. 4, p. 1–39, 2023. Citation on page 122.

YANG, S.; WANG, F.; YANG, L.; XU, F.; LUO, M.; CHEN, X.; FENG, X.; ZOU, X. The physical significance of acoustic parameters and its clinical significance of dysarthria in Parkinson's disease. **Scientific Reports**, Nature Publishing Group UK, v. 10, n. 11776, p. 1–9, 2020. ISSN 2045-2322. Available: <<https://doi.org/10.1038/s41598-020-68754-0>>. Citations on pages 109 and 112.

YOLCU, G.; OZTEL, I.; KAZAN, S.; OZ, C.; PALANIAPPAN, K.; LEVER, T. E.; BUNYAK, F. Facial expression recognition for monitoring neurological disorders based on convolutional neural network. **Multimedia Tools and Applications**, Springer, v. 78, p. 31581–31603, 2019. Citations on pages 30 and 88.

ZHAM, P.; ARJUNAN, S. P.; RAGHAV, S.; KUMAR, D. K. Efficacy of Guided Spiral Drawing in the Classification of Parkinson's Disease. **IEEE Journal of Biomedical and Health Informatics**, v. 22, n. 5, p. 1648–1652, 2018. Citation on page 97.

ZHAM, P.; KUMAR, D.; VISWANATHAN, R.; WONG, K.; NAGAO, K. J.; ARJUNAN, S. P.; RAGHAV, S.; KEMPSTER, P. Effect of levodopa on handwriting tasks of different complexity in parkinson's disease: a kinematic study. **Journal of neurology**, Springer, v. 266, p. 1376–1382, 2019. Citation on page 97.

ZHAM, P.; POOSAPADI, S. A.; KEMPSTER, P.; RAGHAV, S.; NAGAO, K. J.; WONG, K.; KUMAR, D. Differences in levodopa response for progressive and non-progressive micrographia in parkinson's disease. **Frontiers in Neurology**, Frontiers Media SA, v. 12, p. 665112, 2021. Citation on page 97.

ZHAM, P.; RAGHAV, S.; KEMPSTER, P.; ARJUNAN, S. P.; WONG, K.; NAGAO, K. J.; KUMAR, D. K. A kinematic study of progressive micrographia in parkinson's disease. **Frontiers in neurology**, Frontiers Media SA, v. 10, p. 403, 2019. Citation on page 97.

ZHANG, J.; CORMODE, G.; PROCOPIUC, C. M.; SRIVASTAVA, D.; XIAO, X. Privbayes: Private data release via bayesian networks. **ACM Transactions on Database Systems (TODS)**, ACM New York, NY, USA, v. 42, n. 4, p. 1–41, 2017. Citations on pages 62 and 64.

ZHANG, L.; RAO, A.; AGRAWALA, M. Adding conditional control to text-to-image diffusion models. In: **Proceedings of the IEEE/CVF International Conference on Computer Vision**. [S.l.: s.n.], 2023. p. 3836–3847. Citation on page 122.

ZHANG, X.; YIN, L.; COHN, J. F.; CANAVAN, S.; REALE, M.; HOROWITZ, A.; LIU, P.; GIRARD, J. M. Bp4d-spontaneous: a high-resolution spontaneous 3d dynamic facial expression database. **Image and Vision Computing**, Elsevier, v. 32, n. 10, p. 692–706, 2014. Citation on page 76.

ZHAO, J.; MATHIEU, M.; LECUN, Y. Energy-based generative adversarial network. **arXiv preprint arXiv:1609.03126**, 2016. Citation on page 135.

ZHOU, A.; LI, S.; SRIRAM, P.; LI, X.; DONG, J.; SHARMA, A.; ZHONG, Y.; LUO, S.; KINDRATENKO, V.; HEINTZ, G. *et al.* Youtubepd: A multimodal benchmark for parkinson's disease analysis. **Advances in Neural Information Processing Systems**, v. 36, 2024. Citations on pages 42, 45, 50, and 51.

ZHOU, Y.; PANG, M.; HUANG, W.; WANG, B. Early diagnosing parkinson's disease via a deep learning model based on augmented facial expression data. In: IEEE. **ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)**. [S.l.], 2024. p. 1621–1625. Citations on pages 42, 45, 47, and 52.

ZHU, J.-Y.; PARK, T.; ISOLA, P.; EFROS, A. A. Unpaired image-to-image translation using cycle-consistent adversarial networks. In: **Proceedings of the IEEE international conference on computer vision**. [S.l.: s.n.], 2017. p. 2223–2232. Citation on page 127.

ZIMMERMAN, E. K.; ESLINGER, P. J.; SIMMONS, Z.; BARRETT, A. M. Emotional perception deficits in amyotrophic lateral sclerosis. **Cognitive and Behavioral Neurology**, LWW, v. 20, n. 2, p. 79–82, 2007. Citation on page 30.

---

## MEDIA COVERAGE

---

My research, **Facial Expressions to Identify Post-Stroke: A Pilot Study**, has garnered significant media attention, featured in major outlets including 9News, 7News, ABC, RMIT News, The Conversation, and Brazilian Forbes, achieving an **Altmetric attention score of 319**. Additionally, Figure 37 displays a screenshot of the websites of *The Conversation*, *Forbes*, and *RMIT News*. I developed an innovative AI-powered smartphone application that leverages facial expressions, particularly smiles, to rapidly detect signs of stroke. This groundbreaking tool has the potential to revolutionize emergency response by enabling paramedics to expedite critical treatments, thereby potentially saving countless lives.

Additionally, my research on **ALS Detection through Facial Analysis** (paper: *Video Assessment to Detect Amyotrophic Lateral Sclerosis*) has also received noteworthy media coverage. By analyzing subtle facial movements, this study provides a promising approach for ALS detection. Moreover, the **systematic literature review on facial expression for Parkinson's Disease** (paper: *Facial Expression Analysis in Parkinson's Disease Using Machine Learning: A Review*) highlights a novel approach to understanding and quantifying hypomimia, a reduced ability to express emotions that is often seen in Parkinson's patients.

### ***Facial Expressions to Identify Post-Stroke***

Below is a selection of prominent media mentions for my stroke-related research. These items demonstrate the broad reach of the study's impact, from mainstream news channels to specialized health outlets:

- **The Conversation** – <<https://theconversation.com/our-smartphone-screening-tool-could-help-detect-strokes-faster-and-lead-to-quicker-treatment-232710>>
- **Forbes** – <<https://forbes.com.br/forbes-tech/2024/08/conheca-brasileiro-que-desenvolveu-aplicativo-de-ia-capaz-de-revolucionar-diagnostico-de-avc/>>

Figure 37 – Screenshot showing, from left to right, the websites of *The Conversation*, *Forbes*, and *RMIT News*, which feature AI-based stroke detection research.



- **9News** – <<https://www.9news.com.au/national/new-scanning-tool-could-help-paramedics-identify-a-stroke-in-seconds/926627b9-d851-4278-89a8-e8476876b74a>>
- **7News** – <<https://www.youtube.com/watch?v=dvY6XY5aPZ8>>
- **RMIT News** – <<https://www.rmit.edu.au/news/all-news/2024/june/stroke-face-screening>>
- **Unesp** – <<https://www.fc.unesp.br/#!/noticia/1203/projeto-visa-estudo-e-desenvolvimento-de-modelos-inteligentes/>>
- **ABC** – <<https://www.abc.net.au/listen/programs/am/detecting-signs-of-stroke-with-a-phone/103989968>>
- **US News** – <<https://www.usnews.com/news/health-news/articles/2024-06-19/smartphone-face-screening-tool-could-help-paramedics-spot-stroke>>

- **Correio Braziliense** – <<https://www.correiobraziliense.com.br/ciencia-e-saude/2024/07/6882837-em-testes-aplicativo-com-ia-e-capaz-de-detectar-risco-de-avc.html>>

Moreover, a custom web scraper was employed to gather broadcast data from various online sources, and the consolidated results are presented in Table 44. This table provides a comprehensive overview of the media coverage, including the country of origin, media outlet name, broadcast type (e.g., article, radio, or video), and direct hyperlinks to each item. By bringing together an array of global sources—from Australia, Brazil, Germany, India, the United States, and more—the table underscores the extensive reach and significance of this pioneering stroke-detection technology.

Table 44 – Broadcast Links

<b>Country</b>	<b>Broadcasting Name</b>	<b>Media Type</b>	<b>Link</b>
Australia	The Conversation	Article	<a href="#">Link</a>
Australia	RMIT News	Article	<a href="#">Link</a>
Australia	Mirage News	Article	<a href="#">Link</a>
Australia	Hospital & Health	Article	<a href="#">Link</a>
Australia	9 News	Article	<a href="#">Link</a>
Australia	The Times	Article	<a href="#">Link</a>
Australia	Mirage News	Article	<a href="#">Link</a>
Australia	Yahoo News Australia	Article	<a href="#">Link</a>
Australia	iTWire	Article	<a href="#">Link</a>
Australia	Daily Telegraph	Article	<a href="#">Link</a>
Australia	2BS	Radio	<a href="#">Link</a>
Australia	ABC AM	Radio	<a href="#">Link</a>
Australia	ABC Sydney	Radio	<a href="#">Link</a>
Australia	ABC Canberra	Radio	<a href="#">Link</a>
Australia	4BC News Talk	Radio	<a href="#">Link</a>
Australia	2GB Radio	Radio	<a href="#">Link</a>
Australia	7NEWS Melbourne	Video	<a href="#">Link</a>
Australia	ABC News	Video	<a href="#">Link</a>
Brazil	Forbes	Article	<a href="#">Link</a>
Brazil	Olhar Digital	Article	<a href="#">Link</a>
Brazil	Super Interessante	Article	<a href="#">Link</a>
Brazil	Epoca Negocios (Globo)	Article	<a href="#">Link</a>
Brazil	Correio Braziliense	Article	<a href="#">Link</a>
Brazil	UOL	Article	<a href="#">Link</a>
Brazil	SBT News	Video	<a href="#">Link</a>
Brazil	FC Unesp	Article	<a href="#">Link</a>

*Continued on next page*

Table 44 – Continued from previous page

<b>Country</b>	<b>Broadcasting Name</b>	<b>Media Type</b>	<b>Link</b>
Brazil	Jornal UNESP	Video	<a href="#">Link</a>
Brazil	Jornal UNESP	Article	<a href="#">Link</a>
Brazil	Rede GenteTV	Article	<a href="#">Link</a>
Brazil	Anadem	Article	<a href="#">Link</a>
Brazil	Francisco Sales	Article	<a href="#">Link</a>
Brazil	Chico Sabe Tudo	Article	<a href="#">Link</a>
Brazil	MSN Brazil	Article	<a href="#">Link</a>
Germany	NetDoktor	Article	<a href="#">Link</a>
Global	Interesting Engineering	Article	<a href="#">Link</a>
Global	The Brighter Side News	Article	<a href="#">Link</a>
Global	New Atlas	Article	<a href="#">Link</a>
Global	News-Medical.net	Article	<a href="#">Link</a>
Global	MENAFN	Article	<a href="#">Link</a>
Global	Healthbanks	Article	<a href="#">Link</a>
Global	Eurekaalert	Article	<a href="#">Link</a>
Global	Newzchain	Article	<a href="#">Link</a>
Global	CNN Tech Zone	Article	<a href="#">Link</a>
Global	Econo Times	Article	<a href="#">Link</a>
Global	YouTube	Video	<a href="#">Link</a>
Global	YouTube	Video	<a href="#">Link</a>
Global	YouTube	Video	<a href="#">Link</a>
India	NDTV	Article	<a href="#">Link</a>
India	News9 Live	Article	<a href="#">Link</a>
India	MediBulletin	Article	<a href="#">Link</a>
India	Latestly	Article	<a href="#">Link</a>
Netherlands	Smartphone Magazine	Article	<a href="#">Link</a>
New Zealand	Foreign Affairs	Article	<a href="#">Link</a>
Philippines	MSN	Article	<a href="#">Link</a>
Switzerland	Simba for Kids	Article	<a href="#">Link</a>
UK	Eastern Eye	Article	<a href="#">Link</a>
USA	US News	Article	<a href="#">Link</a>
USA	VOA News	Article	<a href="#">Link</a>
USA	Tech Times	Article	<a href="#">Link</a>
USA	Drugs.com	Article	<a href="#">Link</a>
USA	Medical Design Development	Article	<a href="#">Link</a>
USA	Citizen Tribune	Article	<a href="#">Link</a>

Continued on next page

Table 44 – Continued from previous page

Country	Broadcasting Name	Media Type	Link
USA	WFMZ	Article	<a href="#">Link</a>
USA	Newsmax	Article	<a href="#">Link</a>
USA	KPVI	Article	<a href="#">Link</a>

### ***ALS Detection through Facial Analysis***

Building on the success of our stroke-detection study, this project (Video Assessment to Detect Amyotrophic Lateral Sclerosis) investigates how nuanced facial movements can be leveraged to detect ALS at an earlier stage. Below are representative media sources in Brazil covering the research:

- **Estadão** – [https://www.estadao.com.br/saude/pesquisadores-usam-ia-para-identificar-sinais-de-doenca-degenerativa-em-expressoes-faciais-nprm/?srsltid=AfmBOor8uXmjHj\\\_penalty\z@AA11bUIIjrD06hpZ1QIiYK2WSDlhQbxwj0HqHhU5Nt](https://www.estadao.com.br/saude/pesquisadores-usam-ia-para-identificar-sinais-de-doenca-degenerativa-em-expressoes-faciais-nprm/?srsltid=AfmBOor8uXmjHj\_penalty\z@AA11bUIIjrD06hpZ1QIiYK2WSDlhQbxwj0HqHhU5Nt)
- **CNN** – <https://www.cnnbrasil.com.br/tecnologia/ia-pode-ajudar-a-identificar-doencas-degenerativas-atraves-de-expressoes-faciais/>
- **Canaltech** – <https://canaltech.com.br/saude/videos-de-expressoes-faciais-podem-indicar-sinais-de-doenca-degenerativa/>
- **Uol** – <https://noticias.uol.com.br/ultimas-noticias/agencia-estado/2024/08/29/pesquisadores-usam-ia-para-identificar-sinais-de-doenca-degenerativa-em-expressoes-faciais.htm>

### ***Hypomimia in Parkinson Disease***

Finally, I have extended my facial analysis work to Parkinson’s disease, focusing on the condition known as *hypomimia* (reduced facial expressions). Below is one of the media mentions that highlights this evolving research path:

- **Kudos** – <https://www.growkudos.com/publications/10.1145%25252F3716818/reader>

By exploring these three areas—post-stroke detection, ALS identification, and Parkinson’s-related hypomimia—my research demonstrates the vast potential of AI-driven facial analysis to enhance clinical diagnoses across a range of neurological conditions, ultimately aiming to improve patient care and outcomes on a global scale.

---

## ETHICS APPROVAL LETTER

---

---

Ethics approval is a very important part of any research project. It shows that the study has been carefully reviewed to protect the rights, safety, and well-being of all participants. This approval demonstrates that the research follows ethical guidelines and standards, which builds trust with both the public and the academic community.

The Ethics Approval Letter confirms that an ethics committee has thoroughly examined the project. They looked at the research methods, potential risks, and benefits to ensure that everything is handled properly. This process not only meets the necessary legal and institutional requirements but also reassures readers that the study is being conducted responsibly.

In this letter, you will find details about the ethical standards that were followed, including the guidelines from the National Statement on Ethical Conduct in Human Research (NHMRC, 2007) and other relevant protocols. It also outlines the measures put in place to protect participants and ensure the integrity of the research.

For a complete understanding of these ethical safeguards, please refer to the full Ethics Approval Letter on the next page.

**STEM College**

College Human Ethics Advisory  
Network (CHEAN)  
Email: [humanethics@rmit.edu.au](mailto:humanethics@rmit.edu.au)  
Tel: [61 3] 9925 9160

**Notice of Approval**

Date: **8 September 2023**

Project number: 25163

Project title: *Automated Facial Expression for Parkinson Disease*

Risk classification: **Negligible/Low**

Principal investigator: Professor Dinesh Kumar

Status: **Approved**

Approval period: From: **08/09/2023** To: **09/09/2026**

The above application has been approved by the RMIT University CHEAN as it meets the requirements of the *National Statement on Ethical Conduct in Human Research* (NHMRC, 2007).

## Terms of approval:

**1. Responsibilities of principal investigator**

It is the responsibility of the above principal investigator to ensure that all other investigators and staff on a project are aware of the terms of approval and to ensure that the project is conducted as approved by CHEAN. Approval is valid only whilst the principal investigator holds a position at RMIT University.

**2. Amendments**

Approval must be sought from CHEAN to amend any aspect of a project. To apply for an amendment, submit a request for amendment form, via the Research Ethics Platform. Amendments must not be implemented without first gaining approval from CHEAN.

**3. Adverse events**

You should notify the CHEAN immediately (within 24 hours) of any serious or unanticipated adverse effects of their research on participants, and unforeseen events that might affect the ethical acceptability of the project.

**4. Annual reports**

Continued approval of this project is dependent on the submission of an annual report. Annual reports must be submitted by the anniversary of approval of the project for each full year of the project. If the project is of less than 12 months duration, then a final report only is required.

**5. Final report**

A final report must be provided within six months of the end of the project. The CHEAN must be notified if the project is discontinued before the expected date of completion.

**6. Monitoring**

Projects may be subject to an audit or any other form of monitoring by the CHEAN at any time.

**7. Retention and storage of data**

The investigator is responsible for the storage and retention of original data according to the requirements of the *Australian Code for the Responsible Conduct of Research (R22)* and relevant RMIT policies.

**8. Special conditions of approval**

Not applicable.

In any future correspondence please quote the project number and project title above.

Yours faithfully,

**Dr Lauren Saling**  
**Chair, STEM College Human Ethics Advisory Network**

Cc Student Investigator/s: Mr Guilherme Oliveira  
Other Investigator/s: Dr Quoc Cuong Ngo  
Dr Nemuel Daniel Pah  
Dr Barbara Polus