

UNIVERSIDADE FEDERAL DO ESTADO DO RIO DE JANEIRO
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA

EDUARDO FURTADO DE SOUZA OLIVEIRA

**Data augmentation and feature engineering as a strategy to improve F-score
metric of hand tremor classifiers**

Rio de Janeiro, RJ - Brasil

Março de 2024

Catálogo informatizado pelo(a) autor(a)

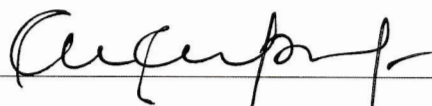
- | | |
|-----|---|
| 048 | <p>Oliveira, Eduardo Furtado de Souza
Data augmentation and feature engineering as a strategy
to improve F-score metric of hand tremor classifiers /
Eduardo Furtado de Souza Oliveira. -- Rio de Janeiro, 2024.
62 f</p> <p>Orientadora: Ana Cristina Bicharra Garcia.
Dissertação (Mestrado) - Universidade Federal do Estado
do Rio de Janeiro, Programa de Pós-Graduação em Informática,
2024.</p> <p>1. Artificial Intelligence in Medical Diagnostics. 2.
Hand Tremor Classification. 3. Parkinson's Disease Early
Diagnosis. I. Garcia, Ana Cristina Bicharra, orient. II.
Título.</p> |
|-----|---|

Data augmentation and feature engineering as a strategy to improve F-score metric of hand tremor classifiers

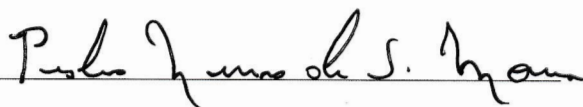
Eduardo Furtado de Souza Oliveira

DISSERTAÇÃO APRESENTADA COMO REQUISITO PARCIAL PARA OBTENÇÃO DO TÍTULO DE MESTRE PELO PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA DA UNIVERSIDADE FEDERAL DO ESTADO DO RIO DE JANEIRO (UNIRIO). APROVADA PELA COMISSÃO EXAMINADORA ABAIXO ASSINADA

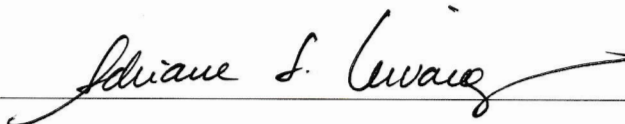
Aprovada por:



Ana Cristina Bicharra Garcia - UNIRIO



Pedro Nuno de Souza Moura - UNIRIO



Adriana Santarosa Vivacqua - UFRJ

RIO DE JANEIRO, RJ - BRASIL

Março de 2024

Abstract

Artificial intelligence is increasingly being adopted in medical diagnostics, offering tools that play a critical role in facilitating early diagnoses. This timely detection is essential for ensuring patients receive the appropriate care and treatment, and is especially evident in the analysis and diagnosis of movement disorders, notably Parkinson’s disease. One key challenge is ensuring the quality of the data used to train these algorithms, particularly when using 2D videos that are key to enabling in-home assessments for elderly patients or those with mobility problems. Our research addresses the critical issue of low accuracy in existing AI-driven diagnostic models for hand tremor disorders. Despite ongoing efforts to enhance these models, a significant challenge remains: the quality of the training data. Limited data availability, poorly captured videos, and the use of a narrow set of features, hinder the model’s learning efficiency and accuracy. Our hypothesis proposes that the underlying cause of this low accuracy is not just the complexity of hand tremor disorders, but also these fundamental limitations in the training data itself. Building on the foundational work of (YANG et al., 2022), we propose a comprehensive approach to improve the quality of training data for tremor classification with our proposed PClassifier. By implementing preprocessing techniques for data augmentation, we aim to enrich the dataset both in terms of quantity and diversity. Additionally, we expand the feature set used by the diagnostic models to include a wider array of statistical measures extracted from the tapping signal, thereby increasing the entropy of information available for model training. Centered on 2D video analysis for the non-invasive and accessible diagnosis of hand tremors, our PClassifier approach resulted in a statistically significant increase in the model’s F1 score, validating our hypothesis and underscoring the effectiveness of our methodological enhancements.

Keywords: Artificial Intelligence in Medical Diagnostics; Hand Tremor Classification; Parkinson’s Disease Early Diagnosis; Video Data Preprocessing.

List of Figures

Figure 1 – Word cloud of keywords from studies on AI diagnosis of hand tremors, illustrating key research themes.	17
Figure 2 – PRISMA Flowchart depicting the literature selection process for AI-assisted diagnosis of hand tremor disorders.	18
Figure 3 – Overview of the literature found during our review on AI-assisted for hand tremor diagnosis using videos.	20
Figure 4 – Tapping rate illustration pinpointing local minimas	34
Figure 5 – Tapping frozen state, indicating periods of little movement during the task	34
Figure 6 – Tapping amplitude with a declining trend observed throughout the task	34
Figure 7 – Participant L8 performing the finger tapping task with the hand near their face while a black box covers their identity.	36
Figure 8 – Instance from participant L31 where a black box appears over the hand movements, despite the absence of a face in the frame, resulting in unintended obstruction of crucial data.	37
Figure 9 – Overview of the PClassifier for AI-assisted diagnosis of hand tremor disorders, highlighting data augmentation techniques and the incorporation of new statistical features.	39
Figure 10 – Example of video from participant L40 after applying rotation and mirroring techniques for data augmentation.	41
Figure 11 – Frame from participant L40 showing original video quality (left) and the result after preprocessing it with CLAHE (right).	47
Figure 12 – Landing page of the video collection platform	48
Figure 13 – Login page of the platform	49
Figure 14 – Homepage when the user is logged in	49
Figure 15 – Doctors can add new participants by inputting their information, such as age and pathology	49
Figure 16 – List of all participants in the platform	50

Figure 17 – User page of the participant, where doctors can visualize their information, edit and save new information, and have access to their history of recorded videos	50
Figure 18 – The list videos page showing all videos recorded on the platform	51
Figure 19 – Before recording a new video, an instruction page is shown to the participant informing best practices on how to perform the finger tapping task	51
Figure 20 – The record page of the platform. When recording, a countdown allows the participant to know for how long they need to perform the task . .	52

List of Tables

Table 1 – Goals and tasks used in reviewed papers.	28
Table 2 – Datasets participants, raters, quality and preprocess techniques used by selected papers.	29
Table 3 – Features and extraction methods used by selected papers.	30
Table 4 – Model techniques, validation, metrics and performance reported in studies.	31
Table 5 – Average precision, recall, and F1-score for each of our experiments. . . .	61
Table 6 – Detailed precision, recall, and F1-score results by fold in cross-validation for the left hand.	61
Table 7 – Detailed precision, recall, and F1-score results by fold in cross-validation for the right hand.	62

Contents

1	Introduction	9
1.1	<i>Background and Motivation</i>	9
1.2	<i>Problem Statement</i>	12
1.3	<i>Structure of the Dissertation</i>	13
2	Literature Review	15
2.1	<i>Review Methodology</i>	15
2.2	<i>Review Results</i>	19
2.2.1	Objective of Research	21
2.2.2	Videos Tasks	21
2.2.3	Dataset Overview	22
2.2.4	Feature Engineering	22
2.2.5	Model Techniques and Metrics	24
2.3	<i>Discussion of Reviewed Literature</i>	26
3	Research Methodology	32
3.1	<i>Pose Extraction</i>	33
3.2	<i>Feature Engineering</i>	33
3.3	<i>Modeling Architecture</i>	34
3.4	<i>Cleaning the Dataset</i>	36
4	Proposed Approach	38
4.1	<i>Original Results Reported by Yang et al.</i>	39
4.2	<i>Replicating Yang et al. Methodology</i>	40
4.3	<i>Data Augmentation</i>	40
4.4	<i>Implementing New Features</i>	41
4.5	<i>Combining Techniques</i>	43
4.6	<i>Statistical Analysis of Results</i>	43
4.7	<i>Discussion of the Results</i>	44
5	Lessons Learned	46
5.1	<i>Preprocessing Videos Using CLAHE</i>	46

5.2	<i>Ethics Committee Process</i>	47
6	Conclusion	53
6.1	<i>Future Work</i>	53
	Bibliography	55
A	Experiments Results	61

1 Introduction

1.1 Background and Motivation

Tremor is notably the most common movement disorder encountered in outpatient neurology practices (PUSCHMANN; WSZOLEK, 2011). This condition is characterized by rhythmic oscillation of a body part, most often affecting the hands (JANKOVIC, 1980). The causes of involuntary tremors are diverse, ranging from Parkinson’s disease (PD) (BAUMANN, 2012) and Essential Tremor (ET) (LOUIS; FERREIRA, 2010). Other types include Enhanced Physiologic Tremor and Orthostatic Tremor, among others, each distinguished by unique frequencies and the body regions they affect (RANA; CHOU, 2015).

Clinically, tremor primarily affects the upper limbs in at least 95% of patients (ELBLE, 2013), but may also occur less frequently in other body parts, such as the head, face, trunk, lower limbs, and voice (ELBLE, 2013). Involuntary tremors, documented for centuries (PARKINSON, 2002), are not limited to ET and PD, but also manifest in disorders such as Enhanced Physiologic Tremor and Orthostatic Tremor, varying in frequency and body region affected (RANA; CHOU, 2015). While tremors significantly define ET, the disease also encompasses debilitating non-motor characteristics like depression, anxiety, fatigue, and sleep disturbances (LOUIS; BENITO-LEÓN; AND, 2007; SENGUL et al., 2014), with depressive symptoms affecting many patients yet often overlooked (LOUIS et al., 2012).

Parkinson’s disease, a commonly known neurodegenerative disorder, is typically diagnosed through patient history and clinical examinations (ARMSTRONG; OKUN, 2020). Patients with PD frequently struggle with movement difficulties, including tremors, stiffness, and slowness, and also face psychological challenges such as depression and anxiety. Clinical tests often reveal two major symptoms: bradykinesia (a decrease in movement speed), rigidity (involuntary resistance to passive movement of a joint by an examiner), or both of them (ARMSTRONG; OKUN, 2020).

Essential Tremor (ET) is another common movement disorder, characterized by its varying prevalence across studies. In 2018, the International Parkinson and Movement Disorder Society issued a consensus statement on tremor classification (BHATIA et al., 2017), which defined ET as a condition primarily involving bilateral upper limb action

tremor persisting for at least three years. Additionally, ET may also present with tremors in other locations, such as the head, voice, or lower limbs, lacking other neurological signs, such as dystonia, ataxia, or parkinsonism (BHATIA et al., 2017).

A substantial increase in the prevalence of movement disorders is expected between 2010 and 2050 (BACH et al., 2011), primarily due to population aging and longer life expectancies. In 2010, PD affected 0.4% to 0.5% of the total population, with projections suggesting an increase to approximately 0.8% in Europe and Canada, and 0.6% in the United States (BACH et al., 2011). ET is also a widespread movement disorder, with its prevalence varying significantly across different age groups and geographies. The pooled prevalence of ET in the general population is approximately 1.3% (LOUIS; MCCREARY, 2021). Notably, the prevalence of ET escalates with age, increasing in older populations. For individuals aged 65 years and above, the prevalence is 5.79%, observed to increase by 74% for each decade of age (LOUIS; MCCREARY, 2021).

No known cures exist for ET or PD, as current treatment focuses on symptom management (ABBOUD; AHMED; FERNANDEZ, 2011; BAUMANN, 2012). Such treatments encompass drugs like Propranolol and Primidone, surgical interventions for Deep Brain Stimulation and Focused Ultrasound, and Botulinum toxin injections (WELTON et al., 2021). Early diagnosis is crucial in the effective management of tremor-related disorders, playing a pivotal role in ensuring patients receive the appropriate treatment at the right time (LOCATELLI et al., 2020). The literature indicates that AI, when trained with sufficient data, can perform comparably to medical experts in diagnosing these conditions (SHEN et al., 2019). This is particularly significant, as AI can aid in the early detection and accurate classification of various tremor disorders, thereby facilitating timely and targeted interventions. Furthermore, AI's integration into healthcare extends to telemedicine, where it enhances accessibility and convenience for patients (BECK et al., 2017). By enabling remote diagnosis and monitoring, AI-powered telemedicine solutions can significantly improve patient care, especially for those with mobility challenges or in remote locations.

AI and machine learning techniques have shown considerable promise in early detection efforts, offering new avenues for addressing this growing challenge. (PEREIRA et al., 2019) review notes the diversity of methodologies employed in recent studies, including image and signal analysis, the use of smartphone devices, virtual and augmented reality, and web-based applications. Signal analysis, obtained through on-body sensors, stands

out as the most utilized method, highlighting the pivotal role of sensor technology in PD research. This review also acknowledges the widespread adoption of machine learning across these studies, emphasizing its essential function in automating PD diagnosis.

Further, the potential of smartphones and tablets in the future of PD research is underscored, with the development of e-health research kits poised to improve patient monitoring and disease understanding. The ability to monitor patients in their home environments offers a promising direction for actively understanding PD, marking a shift towards more accessible and continuous care options. It reinforces the critical role of AI and machine learning in advancing PD diagnostics and patient care, pointing towards an evolving landscape where technology plays a central role. The convergence of these tools paves the way for innovative solutions in the fight against the disease, spotlighting the need for continued research and development in this promising field. (BELIĆ et al., 2019) also conducted a review, covering a selection of studies published over a twelve-year period, to assess the implementation of machine learning algorithms in the diagnosis and assessment of PD. Their examination specifically focused on kinematic data derived from the movements of both upper and lower extremities, as well as whole-body movements. The studies collectively underscore the promising capabilities of machine learning algorithms in early PD diagnostics, leveraging data from globally accessible and cost-effective devices. Such innovations hold the potential to revolutionize PD diagnostics by making them more accessible and precise.

However, despite these advancements, (BELIĆ et al., 2019) identify several prevailing challenges that could impede the full realization of AI's potential in this field. Notably, the validation efforts for these diagnostic tools are often hampered by small sample sizes, the diverse nature of the instruments used, and the heterogeneity among participant groups concerning disease stage and symptoms. Additionally, inconsistencies in outcome measures, validity assessment methods, and the reporting of results further complicate the landscape.

To overcome these hurdles and harness AI's full capabilities in PD diagnostics and management, (BELIĆ et al., 2019) advocate for a more collaborative approach within the scientific community. This would involve the sharing and merging of datasets, as well as the alignment of data collection protocols to standardize research methodologies and enhance the validity of sensor-derived assessments of PD features. Such concerted efforts are crucial for advancing the field, enabling more robust and universally applicable diagnostic tools

that can accurately assess and monitor the progression of PD and other neurodegenerative movement disorders.

In addition to the broader motivations that has guided our study, there is also a personal reason this study. My mother lives with Essential Tremor, and witnessing her daily struggles and triumphs has been a profound influence. The tremors impact on her daily life is noticeable, and activities such as using a fork or tying shoelaces, among other simple tasks, can be challenging. However, I’ve also witnessed the positive impact of AI in her life. For instance, she struggles with typing texts on her phone, but the speech-to-text feature has been a game-changer for her. This technology makes it easier for her to communicate, showing me firsthand how AI can assist in her daily life. This personal experience was a major reason why I chose to pursue my master’s degree in this subject. I discussed this with my advisor since our first meeting, expressing my desire to explore how AI could help people like my mom. My advisor’s unwavering support has been a great source of motivation, and I’m truly grateful for it.

1.2 Problem Statement

The primary challenge this research aims to tackle is the low F1 scores observed in classifiers for AI-aided diagnosis of hand tremor disorders. The F1 score, an essential measure that combines precision and recall, indicates the current models’ limited effectiveness in classifying tremor severity accurately. This issue directly impacts the reliability of diagnostic outcomes and requires methodological enhancements to improve diagnostic accuracy.

Our Hypothesis: The low F1 scores stem from two main factors: the small size of the training dataset, limiting the model’s learning capacity, and the low entropy of the input layer, resulting from a restricted set of features and potential issues in the quality of data capture. Expanding the dataset with new instances required obtaining approval from an ethics committee, as well as establishing a partnership with medical professionals to accurately acquire and label the data. Given the time constraints associated with these processes, we have deferred this approach to future work. We propose that applying data augmentation to expand the dataset and enrich the input layer with a wider array of statistical features will significantly improve the classifiers’ F1 scores. These steps are

expected to increase the diversity and quality of information the model processes, enhancing its diagnostic performance. Additionally, trying to get more data for training presents challenging obstacles. The main issues are getting approval from ethics committees, which has a lengthy process, although it is necessary to look after the privacy and safety of those involved, and finding enough participants for the study. These challenges show how hard it is to simply collect more data to make the model better, and point out why we need to look for other ways to improve how the model learns.

Null Hypothesis: The proposed enhancements, with the data augmentation techniques and the addition of extra statistical features to the input layer, will not result in a statistically significant improvement in the F1 scores of the classifiers.

Our approach to testing these hypotheses involves a quantitative analysis using a t-test to evaluate the statistical significance of the observed improvements in F1 scores after implementing our enhancements. A statistically significant increase in the F1 scores compared to the baseline would allow us to reject the null hypothesis, confirming that our proposed modifications positively affect the model’s performance.

1.3 Structure of the Dissertation

The structure of this dissertation is organized into several chapters, each focusing on a critical aspect of our research. Following this introductory chapter, a literature review was performed to conduct a thorough examination of the current landscape for diagnosing hand tremors using video analysis. The review in this dissertation culminated in the generation of a paper that was submitted to, accepted, and presented at the International Conference on Agents and Artificial Intelligence (ICAART) in 2024, held in Rome, Italy (FURTADO; GARCIA, 2024).

The methodology chapter provides a detailed description of the finger-tapping dataset, a pivotal task in clinical diagnosis and prevalent in the literature, which was collected and shared by (YANG et al., 2022). This dataset, available for academic use under a license agreement, serves as the foundation for our analysis. Despite the lack of publicly available code from (YANG et al., 2022), we meticulously replicated their methodology based on the description in their publication, thereby independently validating their findings.

In the proposed approach chapter, we focus on the development and implementation of our PClassifier, using data augmentation and the introduction of new statistical features extracted from the tremor signal as primary methods for improving diagnostic performance. After performing the experiments, we executed a statistical analysis, employing a t-test to evaluate the impact of our PClassifier on the model's F1 score. This rigorous evaluation served to underscore the effectiveness of our approach in enhancing the model's ability to classify hand tremor severity with greater performance and reliability.

Concluding the study, the main findings were summarized, highlighting the contribution to the field of AI in medical diagnostics using videos of hand tremors. Acknowledging the study's limitations, in the future studies section of the dissertation we address one of the biggest hurdles in developing AI assisted tools, the lack of available data in the field. We introduce a platform developed to aid clinicians in tracking patient disease progression and to collect valuable research data. We showcase how the platform can be particularly useful for monitoring hand tremor disorders, enabling clinicians to make informed decisions based on the observed progression. Simultaneously, it serves as a valuable research tool, gathering data that can deepen our understanding of tremor-related diseases and treatment efficacy.

2 Literature Review

The advent of artificial intelligence in medical diagnostics has represented an important advancement in the detection and management of complex disorders. The broad application of AI across various medical fields, significantly enhances therapeutic development, early detection, disease diagnosis, and referral management (SHEN et al., 2019). The current state of AI development exhibits diagnostic performances on par with those of medical experts, particularly in image recognition fields (SHEN et al., 2019), a domain critical to our investigation of hand tremor disorders through 2D video analysis.

As we embark on exploring the potential of AI-assisted diagnosis through 2D video analysis, it becomes imperative to situate our research within the broader scientific literature. This chapter aims to systematically review the existing literature, shedding light on the evolution of diagnostic methodologies, the integration of AI technologies, and the specific challenges encountered in the domain of hand tremor disorders. By critically assessing past and current research findings, we establish a comprehensive background of available literature. Our objective is to discern patterns, identify gaps, and understand the trajectory of technological advancements in this field.

2.1 Review Methodology

In our pursuit to investigate AI-assisted diagnosis of hand tremor disorders through 2D video analysis, a systematic literature review was conducted according to (KITCHENHAM; BRERETON, 2013). This review involved a comprehensive search across six prominent academic databases: ACM Digital Library, IEEE Xplore, IOS Press, PubMed, ScienceDirect and Springer. We utilized a specific search string designed to encompass a wide range of relevant studies:

hand AND video AND (tremor OR bradykinesia) AND (classification OR diagnosis OR detection OR identification) AND ("machine learning" OR "artificial intelligence")

This search string, while broad in its scope, was essential to ensure comprehensive coverage of the field, albeit accepting that it would yield a number of off-topic results. The

intention was to be as inclusive as possible, to avoid overlooking studies that could offer valuable insights into the subject matter.

The review was structured around five key research questions listed below. These questions were critical in guiding the evaluation and synthesis of the literature, ensuring a focused exploration of the field. They were instrumental in identifying the methodologies, challenges, and opportunities present in the current research landscape surrounding the use of AI for diagnosing hand tremors via video analysis.

1. Objective of the Research: What is the primary aim of each study (e.g., classifying the type or severity of tremor)?
2. Required Tasks recorded in the video: What specific video tasks are analyzed in the studies, such as finger tapping or hand pronation/supination?
3. Dataset Overview: What are the characteristics of the datasets employed in terms of participant numbers, tremor classes, data accessibility, and any preprocessing techniques?
4. Feature Engineering: Which features are extracted from the videos for analysis, and how are they processed or engineered?
5. Model Techniques and Metrics: What machine learning techniques are utilized, what metrics are used for evaluation, and how are the findings validated?

Our inclusion criteria for the review encompassed studies that were published between 2018 and 2023 and were written in English. The studies selected specifically focused on AI-assisted diagnosis of tremor disorders using 2D videos of hands, highlighting the potential of AI in visual diagnosis. Additionally, we concentrated on research that utilized simple video recording devices, such as smartphones and webcams. This focus was intentional, aiming to underscore methodologies that are practical and applicable in various settings, including telemedicine and environments with limited resources. By emphasizing simple video capture methods, the review sought to identify approaches that are both effective in diagnosing tremor disorders and feasible for widespread implementation.

Conversely, our exclusion criteria to refine our research scope, excluded studies that incorporated more sophisticated technologies, such as 3D video capture, accelerometers, or other specialized sensors. This decision was based on our goal to maintain a focus on accessible and broadly implementable 2D video analysis methods. Additionally, research that centered on non-video data for diagnosis, like handwriting analyses or alternative

diagnostic tests not directly related to hand observation (for example, gait analysis, speech evaluation, head tremor, and facial expressions), was not considered. This approach was chosen to ensure the review remained concentrated on the visual analysis of hand tremors. Moreover, studies that merged hand video analysis with other data types, including gait or accelerometer data, were excluded to keep the investigation focused on the unique contributions of video analysis alone. Furthermore, duplicates, book chapters, and papers from low-impact journals were also excluded.

To visually synthesize the thematic essence of the research landscape surrounding AI-assisted diagnosis of hand tremor disorders, we generated a graphical representation with a word cloud constructed from the keywords of each study included in our literature review. By aggregating these keywords, the word cloud illustrates the frequency and prominence of terms, offering insights into the core topics, methodologies, and focal areas of the current research domain. The size of each term within the word cloud corresponds to its prevalence across the literature, highlighting the most central themes and technologies that define this field of study.



Figure 1 – Word cloud of keywords from studies on AI diagnosis of hand tremors, illustrating key research themes.

This rigorous and targeted selection approach was adopted to ensure our review was concentrated on methodologies that are not only low-cost but also easy to deploy. By focusing on solutions that utilize simpler hardware, our review aligns with the research objective of enhancing accessibility and practicality in diagnostic techniques. The aim was to explore and analyze approaches that are most likely to be readily adopted in diverse clinical and non-clinical settings, thereby contributing to a more inclusive and pragmatic

understanding of current trends and advancements in the field of AI-assisted hand tremor diagnosis.

The initial search across databases yielded a wide array of papers. We streamlined this pool by first reviewing titles and abstracts for relevance, and then reading the full texts of promising studies. This process led to the selection of 17 papers that closely align with our research questions and objectives.

We also provide a PRISMA flowchart that illustrates our literature selection process, from an initial wide-ranging search to the final choice of 17 pertinent studies on AI in diagnosing hand tremor disorders using 2D video analysis. We began by pooling results from the different databases, then systematically narrowed down these findings by excluding off-topic papers, closely reviewing abstracts, and conducting full-text analyses. Below, the flowchart outlines this structured approach, highlighting our methodical strategy to select the most relevant researches for our study.

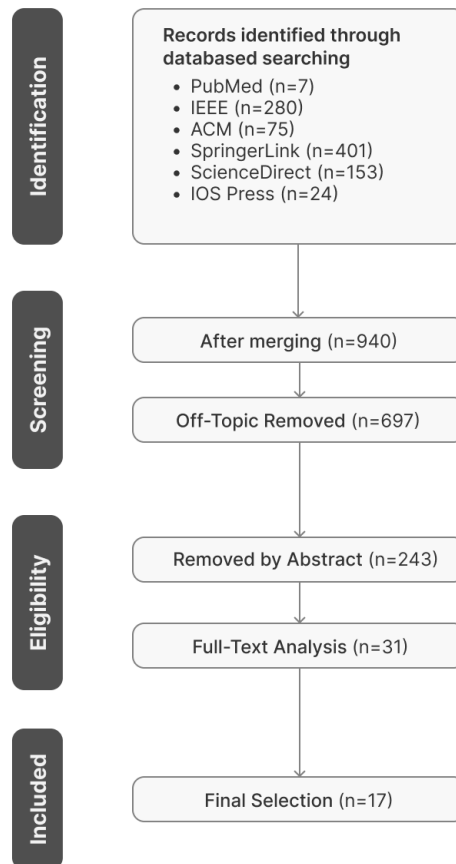


Figure 2 – PRISMA Flowchart depicting the literature selection process for AI-assisted diagnosis of hand tremor disorders.

2.2 Review Results

In this dissertation, a more condensed version of the literature review presented in our full paper is provided. While the detailed analysis in the paper delves into the intricacies of each study, here we offer a distilled interpretation that highlights the most notable patterns and gaps in the current research landscape. To encapsulate the key findings of this review, Figure 3 serves as a visual summary, mapping out the primary objectives, datasets, feature extraction methods, and modeling approaches from the studies we surveyed. It presents a clear and organized overview, allowing for a quick grasp of the critical elements that define the current state of research in this field. Tables 1, 2, 3, 4 are also provided summarizing the key findings related to our research questions, serving as a reference of the reviewed studies.

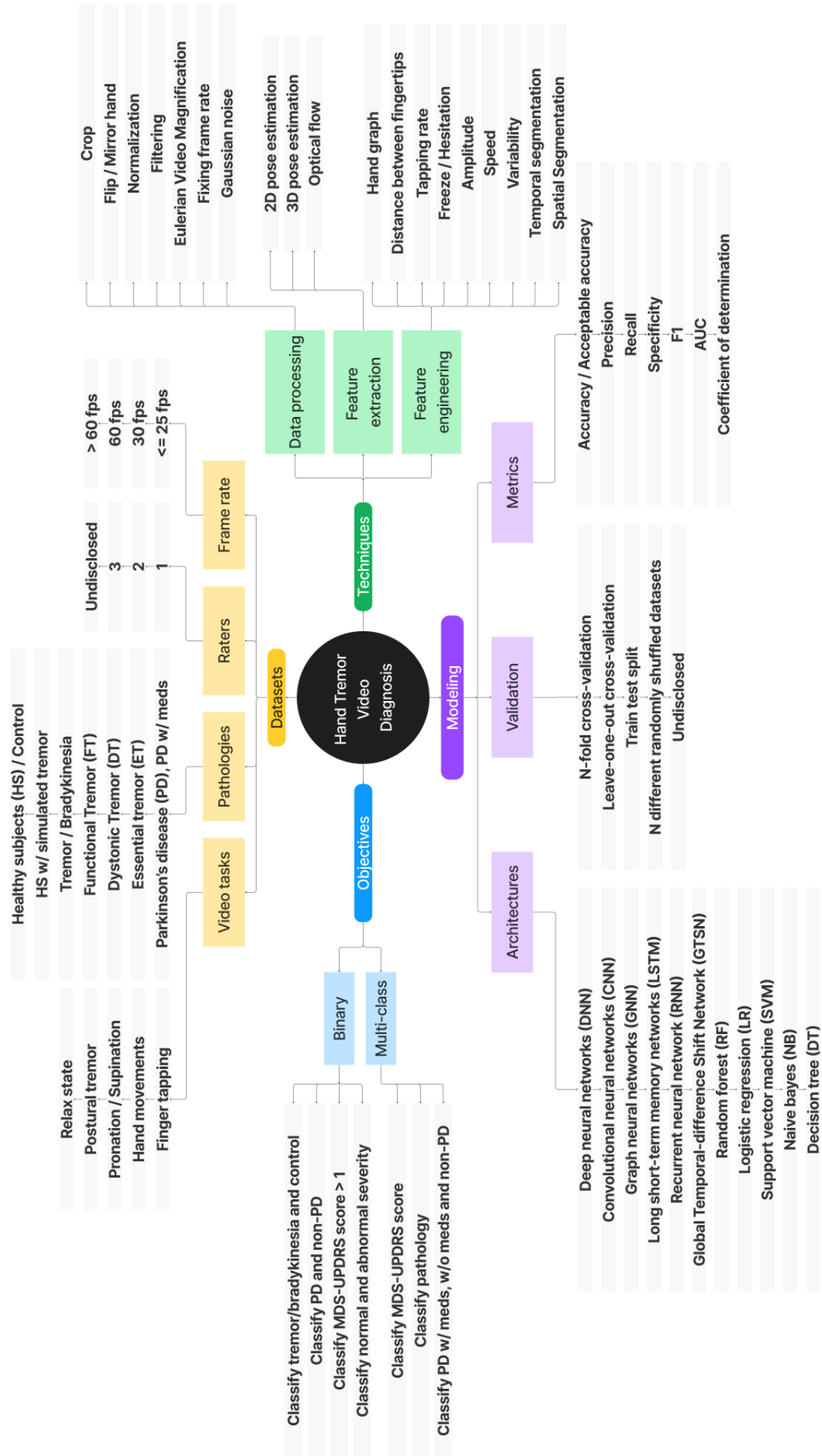


Figure 3 – Overview of the literature found during our review on AI-assisted for hand tremor diagnosis using videos.

2.2.1 Objective of Research

The objectives of the research, as addressed in our first research question, reveal a predominant focus in the selected studies on utilizing AI for the classification of MDS-UPDRS scores, evident in the work of (GUO et al., 2022), (LI et al., 2022), (LI et al., 2021), (YANG et al., 2022), (CHEN et al., 2021), (LU et al., 2021), (ZHAO; LI, 2022), (VIGNOUD et al., 2022), (LIU et al., 2023), and (LIU et al., 2019). The MDS-UPDRS (Movement Disorder Society-Sponsored Revision of the Unified Parkinson’s Disease Rating Scale) (GOETZ et al., 2008) utilizes a scoring system with five categories (0: normal, 1: slight, 2: mild, 3: moderate, and 4: severe), widely employed in clinical assessments of motor and non-motor symptoms associated with the disease. Other approaches include (LIN et al., 2020) who concentrated on distinguishing between Bradykinesia and Healthy Subjects, while (WANG et al., 2021) explored the classification of Tremor and Non-Tremor instances. (CHANG et al., 2019) targeted the differentiation between Normal and Abnormal tremor severity. In a comprehensive approach, (ZHANG et al., 2022) not only investigated classifications of PD and non-PD but also differentiated between various tremor types, including PD, Essential Tremor (ET), Functional Tremor (FT), Dystonic Tremor (DT), and Healthy Subjects. Similarly, (WONG et al., 2019) employed dual-objective methodologies, targeting classifications among different PD conditions and non-PD subjects, as well as classifying MDS-UPDRS scores.

2.2.2 Videos Tasks

The studies reviewed in this dissertation employed a variety of hand movement tasks to facilitate the analysis and classification of tremor types and severities. A prominent task observed across multiple studies of (LI et al., 2022), (LI et al., 2021), (YANG et al., 2022), (CHANG et al., 2019), (CHEN et al., 2021), (LU et al., 2021), (MONJE et al., 2021), (ALI et al., 2020), (WONG et al., 2019), (ZHAO; LI, 2022), (VIGNOUD et al., 2022), (LIU et al., 2019), was finger tapping, which involves participants touching the tips of their index finger and thumb together. Another common task observed was related to general hand movements, particularly opening and closing the hands, as seen in studies conducted by (GUO et al., 2022), (LIN et al., 2020), (CHEN et al., 2021), (MONJE et al., 2021), (ALI

et al., 2020), (ZHAO; LI, 2022), (VIGNOUD et al., 2022), and (LIU et al., 2019). Other studies from (CHEN et al., 2021), (MONJE et al., 2021), (ALI et al., 2020), (VIGNOUD et al., 2022), (LIU et al., 2019), opted for a combination of tasks to enhance the richness of their analysis. These included pronation/supination tasks, where participants turn their palms up and down, often in conjunction with other hand movements.

2.2.3 Dataset Overview

Variability among raters affects the consistency of findings across studies. (YANG et al., 2022), (CHEN et al., 2021), and (LIU et al., 2023) each utilized three raters, whereas studies by (CHANG et al., 2019), (LU et al., 2021), (GUO et al., 2022), (LI et al., 2022), (WANG et al., 2021), and (ZHANG et al., 2022) reported using one or did not specify any. Such differences in rater involvement can pose challenges to the reliability and generalizability of results, particularly critical in medical research where the precision of labels is crucial.

The studies of (LI et al., 2022), (YANG et al., 2022), (LIN et al., 2020), (WANG et al., 2021), (CHEN et al., 2021), (MONJE et al., 2021), (WONG et al., 2019), (ZHAO; LI, 2022), (VIGNOUD et al., 2022), and (LIU et al., 2019) commonly employed video cropping as a preprocessing step, indicating a need to focus on specific regions of interest and remove extraneous data. Additionally, techniques like normalization, used in (LI et al., 2022) and (LU et al., 2021), and various filtering methods, such as the Savitzky-Golay filter in (LI et al., 2021), were prominent. These methods enhance data quality by ensuring consistent input and removing noise.

Moreover, (LIU et al., 2023)’s use of Eulerian Video Magnification (EVM) to amplify subtle video data variations showcases the innovative approaches in data processing for tremor analysis. These diverse preprocessing techniques across the studies reflect the evolving methodologies in the field, aiming to improve data analysis for tremor assessment.

2.2.4 Feature Engineering

In the feature extraction techniques employed across the reviewed studies, 2D Pose Estimation emerges as a commonly used method. This approach is prevalent in studies of

(GUO et al., 2022), (LI et al., 2022), (LI et al., 2021), (YANG et al., 2022), (LIN et al., 2020), (WANG et al., 2021), (CHANG et al., 2019), (CHEN et al., 2021), (ZHANG et al., 2022), (LU et al., 2021), and (MONJE et al., 2021), where keypoints are extracted from video data, varying in number across studies. While most studies focused on 2D Pose Estimation, others like (ZHAO; LI, 2022) and (VIGNOUD et al., 2022) experimented with 3D Pose Estimation or a combination of both 2D and 3D techniques. The pose estimation phase is often performed employing off-the-shelf algorithms (e.g. MediaPipe (ZHANG et al., 2020) and OpenPose (SIMON et al., 2017)).

Diverse algorithms and networks for feature extraction are also evident. For example, (LIN et al., 2020) utilized HandSegNet and PoseNet, and (CHEN et al., 2021) employed SHG with OpenPose, indicating varied approaches in this domain.

In contrast to pose estimation, alternative methodologies such as CNN and Optical Flow were employed by studies like (ALI et al., 2020) and (WONG et al., 2019), showcasing the range of techniques explored for capturing movement data. (LIU et al., 2023)’s use of a temporal difference module for Optical Flow further illustrates the innovative approaches in feature extraction for tremor analysis.

The studies reviewed also reveal a focus on tapping-related features, such as the one-dimensional sequence data of tapping distance in (LI et al., 2022) and tapping rate, frozen times, and amplitude variation in (YANG et al., 2022). (WONG et al., 2019) further expands this with features like tapping frequency and peak-to-peak variability.

Hand graphs and geometric relations are also widely used. For instance, (GUO et al., 2022) and (LU et al., 2021) explore hand movement using hand graphs with 21 keypoints, while (VIGNOUD et al., 2022) and (CHANG et al., 2019) focus on spatial aspects like the distance between thumb and index. (WANG et al., 2021) and (LIU et al., 2019) also emphasize spatial and geometric features, highlighting changes in hand movement and distances between fingertips.

Additionally, some studies introduce task-specific features. (LIN et al., 2020) looks at stability and self-similarity, whereas (CHEN et al., 2021) and (LIU et al., 2023) explore context-specific and temporal features, respectively. This variety in feature engineering, ranging from tapping dynamics to spatial and geometric analyses, illustrates the diverse approaches used to understand and capture movement data in tremor assessment.

2.2.5 Model Techniques and Metrics

Techniques and Architectures

The reviewed studies demonstrate a broad spectrum of models and strategies for classification algorithms, from classic machine learning to specific deep learning architectures. As the evolving landscape of artificial intelligence and machine learning continuously introduces new methodologies, it allows for researchers to experiment with different techniques to capture the nuances of tremor movements and classify it more accurately.

Convolutional Neural Networks (CNNs) feature prominently in (LI et al., 2022) and (LI et al., 2021), the latter using a three-stream CNN for diverse data analysis. Deep Neural Networks (DNNs) are the chosen model architecture in (YANG et al., 2022) and (CHANG et al., 2019), while Recurrent Neural Networks (RNNs) with LSTM units are used by (LIN et al., 2020) for sequential data, and a two-channel LSTM by (ZHAO; LI, 2022). Graph Convolutional Networks (GCNs) are highlighted in (GUO et al., 2022) and (ZHANG et al., 2022), the latter incorporating a Spatial Attention Mechanism with a Graph Neural Network.

Support Vector Machine (SVM) models appear in various studies, including (WANG et al., 2021), (CHANG et al., 2019), and (ALI et al., 2020). (WANG et al., 2021) also explores CNN-LSTM and LSTM models. Ensemble methods, such as Random Forest and Gradient Boosting Decision Trees, along with logistic regression, are applied in (CHEN et al., 2021). Probabilistic classifiers and ensemble methods, including Naïve Bayes and Random Forest, are used by (MONJE et al., 2021), (WONG et al., 2019), and (VIGNOUD et al., 2022).

Unique approaches are seen in (LU et al., 2021) using a Temporal Convolutional Neural Network, and (LIU et al., 2023) introduction of the Global Temporal-difference Shift Network, indicating a diverse range of methods to address different aspects of tremor data analysis.

Validation Approaches

Cross-Validation is a prominent method for model validation, with 5-Fold Cross-Validation being particularly prevalent, as employed in (GUO et al., 2022), (LI et al., 2022),

(CHEN et al., 2021), (ZHANG et al., 2022), and (LIU et al., 2023). Alternatively, (LIN et al., 2020) and (WANG et al., 2021) utilize 10-Fold Cross-Validation for more detailed validation, albeit at a higher computational cost, while (LI et al., 2021) and (MONJE et al., 2021) opt for 4-Fold Cross-Validation.

Leave-One-Out Cross-Validation (LOOCV), suitable for smaller datasets due to its computational intensity, is used by (CHANG et al., 2019), (ALI et al., 2020), and (WONG et al., 2019). Meanwhile, (VIGNOUD et al., 2022) employs validation using 100 randomly shuffled datasets, offering an alternative approach.

Notably, (YANG et al., 2022) only implements a hold-out method performing a train-test split, though with limited details, and (LU et al., 2021) and (ZHAO; LI, 2022) do not specify their validation methodologies. The absence of detailed validation methodologies hinders the ability to assess the generalizability of the models and complicates efforts of reproducing the research.

Metrics Selected and Reported Performance

In evaluating the performance of models across the reviewed studies, accuracy stands out as the most frequently chosen metric. It is either used singularly or in combination with other metrics in studies by (GUO et al., 2022), (LI et al., 2022), (LI et al., 2021), (WANG et al., 2021), (CHANG et al., 2019), (CHEN et al., 2021), (ZHANG et al., 2022), (ALI et al., 2020), (WONG et al., 2019), (ZHAO; LI, 2022), and (LIU et al., 2023). It is important to note that while accuracy is a commonly used metric, high accuracy might not always indicate optimal performance (VALVERDE-ALBACETE; PELAEZ-MORENO, 2014), especially in cases of imbalanced classes. In such scenarios, models might display high accuracy while failing to classify minority classes correctly.

Precision and recall are also commonly selected. This combination is evident in studies of (GUO et al., 2022), (LI et al., 2022), (LI et al., 2021), (YANG et al., 2022), (LIN et al., 2020), (WANG et al., 2021), (LU et al., 2021), and (LIU et al., 2023). sensitivity (or recall) and specificity are utilized in (ZHANG et al., 2022), (MONJE et al., 2021), (WONG et al., 2019), and (ZHAO; LI, 2022), highlighting their focus on both positive and negative classifications. The F1 Score is employed in several studies of (GUO et al., 2022), (LI et al., 2022), (LI et al., 2021), (YANG et al., 2022), (LIN et al., 2020), (WANG et

al., 2021), (ZHANG et al., 2022), and (LIU et al., 2023), reflecting its utility in balancing precision and recall. Additionally, the area under the curve (AUC) is used in (GUO et al., 2022), (LU et al., 2021), (MONJE et al., 2021), (WONG et al., 2019), and (LIU et al., 2023), providing a measure of model performance across different threshold settings.

In the studies conducted by (GUO et al., 2022) and (LI et al., 2021), the reported general accuracy rates are around 70%, and notably high acceptable accuracies, where the difference between the predicted score and the corresponding clinical score is less than 1, exhibiting results approaching 100%. Both (LI et al., 2022) and (WANG et al., 2021) present consistent performance, with accuracies recorded at 79.7% and 80.6%, respectively. In terms of F1-scores, (YANG et al., 2022) reports a high score of 88%, while (LIN et al., 2020) achieves a slightly lower score of 77.78%. The complexities and varied accuracy rates in multiclass classification are explored by (CHANG et al., 2019) and (ZHANG et al., 2022).

2.3 Discussion of Reviewed Literature

The literature review reveals a crucial aspect concerning the availability of datasets used in these studies. A significant number of datasets, as in (GUO et al., 2022), (LI et al., 2021), and others, are either not publicly available or their availability status is unclear. This limitation poses a challenge for the broader research community, hindering the ability to replicate and build upon existing studies. Certain datasets are sometimes accessible upon request, yet they often come with restrictions to protect participant privacy. For instance, (LIU et al., 2023) requires proof of medical research relevance and a signed agreement, creating barriers for researchers outside the medical domain.

The removal of the TIM-Tremor dataset^{1,2} from the internet in recent years, utilized in (WANG et al., 2021) and (ZHANG et al., 2022), further exemplifies the delicate balance between open data access and privacy protection in health-related research.

Given these challenges, data augmentation emerges as a viable solution to address the scarcity of video data. Techniques like video rotation and mirroring can create new instances from existing footage that are not computationally expensive to generate (CAULI; RECUPERO, 2022), thereby expanding the dataset size available for training. This

¹ <https://doi.org/10.4121/UUID:522D14ED-3019-4206-B49E-A4E674B6440A>

² <https://research.tudelft.nl/en/datasets/technology-in-motion-tremor-dataset-tim-tremor>

approach can enhance the research scope, particularly in scenarios where data availability is a constraint.

In the body of research we reviewed, a common theme emerges regarding the preprocessing of data. Most studies primarily employ basic preprocessing, where they crop the original videos around the participant’s hands. However, some studies venture into more advanced preprocessing methods. For instance, (LIU et al., 2023) adopts Eulerian Video Magnification (EVM), an innovative technique that enhances subtle movements in video data, revealing intricate details about hand tremors that might be overlooked with simpler methods. Optical flow, used in (WONG et al., 2019) and (LIU et al., 2023), focuses on the movement aspect, especially the tremors, while disregarding static elements in the background. Such methodologies can potentially improve the quality of data by isolating and emphasizing the primary subject of interest – the tremor.

Despite the prevalent inclination towards minimal preprocessing, these examples highlight the potential benefits of adopting more sophisticated approaches. Enhanced preprocessing could uncover more detailed and nuanced data, which could lead to the development of more accurate and reliable AI models for diagnosing and analyzing hand tremors.

In the following chapter, we will delve into the dataset and methodology of (YANG et al., 2022), which we have chosen to replicate. Overcoming one of the greatest challenges in this field, we successfully gained access to the PDMotorDB dataset, therefore we will thoroughly examine the methodology proposed by (YANG et al., 2022), exploring the rationale behind their techniques and our replication approach. Access to this dataset is pivotal, as it not only enables us to validate and expand upon the original findings but also provides a rare opportunity to closely study a substantial and relevant collection of data in the field of AI-assisted diagnosis.

Table 1 – Goals and tasks used in reviewed papers.

Paper	Objective	Tasks
(GUO et al., 2022)	Classify MDS-UPDRS score	Hand movements
(LI et al., 2022)	Classify MDS-UPDRS score	Finger tapping
(LI et al., 2021)	Classify MDS-UPDRS score	Finger tapping
(YANG et al., 2022)	Classify MDS-UPDRS score	Finger tapping
(LIN et al., 2020)	Classify Bradykinesia / HS	Hand movements
(WANG et al., 2021)	Classify Tremor / Non-Tremor	Various tasks
(CHANG et al., 2019)	Classify Normal / Abnormal tremor severity	Finger tapping, relax state
(CHEN et al., 2021)	Classify MDS-UPDRS score	Finger tapping, hand movements and pronation/supination
(ZHANG et al., 2022)	(1) Classify PD and non-PD; (2) Classify PD, ET, FT, DT, HS	Various tasks
(LU et al., 2021)	Classify MDS-UPDRS score	Finger tapping
(MONJE et al., 2021)	Classify PD / HS	Finger tapping, hand movements and pronation/supination
(ALI et al., 2020)	(1) Classify PD and non-PD; (2) Classify PD meds, PD no meds, non-PD	Finger tapping, hand movements, pronation/supination and postural tremor
(WONG et al., 2019)	(1) Classify MDS-UPDRS ≤ 1 and MDS-UPDRS > 1 ; (2) Classify PD and non-PD	Finger tapping
(ZHAO; LI, 2022)	Classify MDS-UPDRS score	Finger tapping and hand movements
(VIGNOUD et al., 2022)	Classify MDS-UPDRS score	Finger tapping, hand movements and pronation/supination
(LIU et al., 2023)	Classify MDS-UPDRS score	Postural tremor
(LIU et al., 2019)	Classify MDS-UPDRS score	Finger tapping, hand movements and pronation/supination

Table 2 – Datasets participants, raters, quality and preprocess techniques used by selected papers.

Paper	Participants	Raters	Quality	PreProcess
(GUO et al., 2022)	174 PD	N/A	8 seconds (30 fps) 1280 x 720 or 1920 x 1080	N/A
(LI et al., 2022)	120 PD	N/A	10 or more taps (30 fps) 1280 x 720	Crop, normalization
(LI et al., 2021)	157 PD	N/A	150 frames (30 fps) 1280 x 720	Flip left hand and Savitzky-Golay filter
(YANG et al., 2022)	368 PD (left hand), 298 PD (right hand)	3	5 seconds (25 fps) 1920 x 1080	Crop, low pass filtering
(LIN et al., 2020)	94 Bradykinesia + 83 HS	1	10 to 15s (240 fps) 1280 x 720	Crop, mean filter
(WANG et al., 2021)	189 Tremor and 176 Non-tremor (all videos)	N/A	3 seconds (30 fps) 1920 x 1080	Crop
(CHANG et al., 2019)	106 PD	1	300 frames (30 fps) 1280 x 720	N/A
(CHEN et al., 2021)	149 PD	3	N/A	Crop, Fourier filtering
(ZHANG et al., 2022)	105 PD, 182 ET, 88 FT, 204 DT, 60 HS (all videos)	N/A	100 frames	N/A
(LU et al., 2021)	34 PD	1	4 to 30s (30 fps)	Normalization, Gaussian noise
(MONJE et al., 2021)	22 PD + 20 HS + (6 PD 6 HS for val)	N/A	12 seconds (30 fps) 640 x 426	Crop (using Single Shot Multi-Box Detector - SSD), Normalization, Butterworth filter
(ALI et al., 2020)	87 PD meds + 119 PD no meds + 139 HS	N/A	Mean of 9.7s (15 fps) 256 x 256	Fixed frame rate at 15 fps
(WONG et al., 2019)	20 PD + 15 HS	2	10 seconds (60 fps) 1920 x 1080	Crop (CNN)
(ZHAO; LI, 2022)	12 HS (simulating all severity levels)	N/A	500 frames (30 fps) 640 x 480	Crop
(VIGNOUD et al., 2022)	36 PD + 11 HS	2	N/A (30 / 60 fps) 1280 x 720	Crop, Savitzky-Golay filter
(LIU et al., 2023)	130 PD	3	7 to 14s (30 fps) 1920 x 1080	Eulerian Video Magnification (EVM)
(LIU et al., 2019)	60 PD	2	N/A (25 fps)	Crop, Savitzky-Golay filter

Table 3 – Features and extraction methods used by selected papers.

Paper	Features	Extraction methods
(GUO et al., 2022)	Hand graph with 21 keypoints	2D PE (Pose Estimation) (MMPose) - 21 KP (keypoints) (+OpenPose for ROI)
(LI et al., 2022)	One-dimensional sequence data of tapping distance	2D PE (Mediapipe) - 2 KP (index and thumb tips)
(LI et al., 2021)	Pose, Motion and Geometry features from hand graph	2D PE (OpenPose) - 21 KP
(YANG et al., 2022)	Tapping rate, Tapping frozen times, Tapping amplitude variation	2D PE (MMPose) - 2 KP index and thumb tips
(LIN et al., 2020)	Stability (consistency of rhythm), completeness (of actions spatially) and self-similarity (stable periodic motion)	2D PE (HandSegNet + PoseNet) - 21 KP
(WANG et al., 2021)	Change in distance of hand movement and frequency of motion directional changes	2D PE (MediaPipe) - 21 KP
(CHANG et al., 2019)	Distance between keypoints, velocity and acceleration	2D PE (OpenPose) - 9 KP
(CHEN et al., 2021)	Slowing, Amplitude, Amplitude decrement, Hesitation/freeze, Interruption, Incompleteness of performing task	2D PE (SHG - Stacked Hourglass network + OpenPose) - 21 KP
(ZHANG et al., 2022)	Graph with 7 upper body keypoints	2D PE (OpenPose) - 7 upper body KP
(LU et al., 2021)	Hand graph with 21 KP	2D PE (OpenPose)
(MONJE et al., 2021)	Amplitude, Speed, Fatigue	2D PE (OpenPose)
(ALI et al., 2020)	Temporal Segmentation, Spatial Segmentation, Motion Magnification	CNN + Fast Fourier Transform + Deep Neural Network based Magnified Features
(WONG et al., 2019)	Tapping frequency, Energy spectral density (amplitude), Variability of peaks, Jitter, Peak-to-peak variability	Optical Flow
(ZHAO; LI, 2022)	128 dimensional feature vector every 10 frames	3D PE (HandSegNet + PoseNet + PosePrior) - 21 KP
(VIGNOUD et al., 2022)	Distance between the thumb and index, averaged distance between each fingertip and the wrist point, azimuthal angle from spherical coordinates of the tip of the thumb	Pose Estimations (2D: DeepLabCut, 2D+3D: HandGraphCNN)
(LIU et al., 2023)	Temporal features	Temporal difference module for Optical Flow (Pose estimation only to crop body parts (OpenPose) - 21 KP)
(LIU et al., 2019)	FT: Euclidean distance between thumb and index tip; HC: Average of Euclidean distances of each fingertip and palm; Pro/Supination: Difference between horizontal coordinates of thumb and pinky	MobileNet (and ShuffleNet tested)

Table 4 – Model techniques, validation, metrics and performance reported in studies.

Paper	Techniques	Validation	Metrics	Performance
(GUO et al., 2022)	Tree-structure-guided GCN	5-fold CV	Accuracy, Acceptable Accuracy, Precision, Recall, F1 and AUC	Accuracy of 73.71% and an acceptable accuracy of 99.20%
(LI et al., 2022)	CNN	5-fold CV	Accuracy, Precision, Recall and F1	Accuracy of 79.7%
(LI et al., 2021)	Skeleton-based three-stream fine-grained CNN with Markov chain fusion	4-fold CV	Accuracy, Acceptable Accuracy, Precision, Recall, and F1	Accuracy of 72.4% and an acceptable accuracy of 98.3%
(YANG et al., 2022)	DNN	Train Test split	Precision, Recall and F1	F1-score of 88%, 84% on left finger tapping, right finger tapping
(LIN et al., 2020)	Stacked RNN with LSTM	10-fold CV	Recall, Precision, Accuracy and F1	F1-score of 77.78%
(WANG et al., 2021)	CNN-LSTM, LSTM, SVM	10-fold CV	Accuracy, Recall, Precision and F1	Accuracy 80.6%
(CHANG et al., 2019)	DNN, SVM	Leave-One-Out CV	Accuracy	Binary: Accuracy 78.01% and 80.60% in right and left hand; Multiclass: 72.20% and 71.10% for right and left hand
(CHEN et al., 2021)	RF, LR, SVM, GBDT	5-fold CV	Accuracy	Accuracy of 84.1%
(ZHANG et al., 2022)	Graph Neural Network with Spatial Attention Mechanism	5-fold CV	Accuracy, Sensitivity, Specificity and F1	Binary: accuracy of 90.9% and an F1-score of 90.6%; Multiclass: accuracy 73.3% and F1-score 70.7%
(LU et al., 2021)	Temporal convolutional neural network (TCNN)	N/A	F1, AUC, Precision, Balanced Accuracy	Macro-average AUC of 0.69
(MONJE et al., 2021)	LR, NB, RF	4-fold CV	AUC, Sensitivity, Specificity	AUC 0.81
(ALI et al., 2020)	SVM	Leave-One-Out CV	Accuracy	Binary: accuracy 91.8%; Multiclass: accuracy 73.5%
(WONG et al., 2019)	NB, LR, SVM-L, SVM-R	Leave-One-Out CV	Accuracy, Sensitivity, Specificity and AUC	Bradykinesia test accuracy of 79% and PD test accuracy 63%
(ZHAO; LI, 2022)	Two-channel LSTM	N/A	Sensitivity, Specificity and Accuracy	95.7% precision, 95.8% sensitivity and 92.8% specificity
(VIGNOUD et al., 2022)	LR, DT	100 shuffled datasets	Coefficients of determination	Coefficients of determination for the tapping of 0.609 and hand movements of 0.701
(LIU et al., 2023)	Global Temporal-difference Shift Network (GTSN)	5-fold CV	F1, AUC, Precision, Recall and Accuracy	Binary: 93.7%; Multiclass: 84.9% accuracy
(LIU et al., 2019)	RBF-SVM (L-SVM, RF and KNN)	5-fold CV	Precision, Recall, F1 and Accuracy	89.7% accuracy

3 Research Methodology

One of the greatest obstacles in our research was acquiring a suitable dataset for analysis. The challenge of finding a comprehensive and accessible dataset is a common hurdle in the artificial intelligence field, and this difficulty is amplified when dealing with medical data, as is the case of hand tremors diseases. Such datasets not only need to be rich in content but also ethically sourced and shared, balancing the necessity of detailed information with the privacy of participants.

Fortunately, our research overcame this challenge by securing access to the PDMotorDB dataset, developed by (YANG et al., 2022). It includes videos of the finger-tapping task performed by Parkinson’s Disease (PD) patients, totaling 368 and 298 recordings for their left and right hands, respectively. These videos were recorded at the Beijing Tiantan Hospital and have been manually cropped from the original recordings, with faces blurred to ensure patient privacy. The PDMotorDB dataset is available for academic use, provided a signed license agreement, on their GitHub page¹. By utilizing this dataset, our research aims to replicate and validate the methodologies previously developed, contributing to a deeper understanding of PD diagnostics and potentially aiding in the development of more effective treatment approaches.

While (YANG et al., 2022) generously provided access to their dataset, the developed code for the methodology was not publicly available. To address this, we developed our own code, using Python as our programming language of choice, and replicated their original methodology. This process required a detailed examination and interpretation of their published work to reconstruct each step of their methodology accurately. Writing our own code to mirror (YANG et al., 2022) approach tested our understanding of their methods and allowed us to independently validate their findings and potentially explore new dimensions within the same framework. This involved two key steps: first, we developed Python scripts to process and extract hand keypoints data from the raw video files. Then, we implemented (YANG et al., 2022) model architecture in Python as well, utilizing TensorFlow (ABADI et al., 2015) for the construction and training of the model.

Furthermore, all the code developed for this study is open source and has been made available in the supplementary files of the dissertation for a more in-depth analysis. By sharing our code, we hope to contribute to the ongoing development and refinement

¹ <https://github.com/pddata/PDMotorDB>

of methodologies in the field of AI-assisted diagnosis of Parkinson’s Disease and other movement disorders.

3.1 Pose Extraction

Our replication encompassed their detailed three-stage process for tremor severity classification. The first stage involved pose estimation, where we used the MediaPipe Hands (ZHANG et al., 2020) library to extract keypoints for the hand from the video data. This step is crucial for capturing the movement pattern’s raw data, which will be used to classify the tremor severity. Hand pose extraction algorithms generally extract 21 keypoints for each hand, representing each finger joint and the palm, however in this study, only two are needed, to represent the tip of the index finger and the tip of the thumb.

These two key points are used to calculate the distance between them during each frame of the finger-tapping task video, converting the measured distances into a time-series signal. This signal provides a dynamic representation of the patient’s hand movement patterns. After, a normalization process for these distance values is applied, scaling them between 0 and 1. A value of 0 indicated minimal distance, suggesting the fingers were close together, whereas a value of 1 signified the greatest distance achieved during the task. This normalization process was crucial for standardizing the data across various videos, compensating for differences in how close the hand was to the camera in each recording. Furthermore, to reduce noise and enhance the clarity of the tremor patterns in the signal, a Butterworth filter is also applied to the time series. The Butterworth filter is a signal processing method designed to have a flat frequency response in the passband, making it ideal for preserving the original shape of signals while filtering out unwanted frequencies. It balances between signal smoothing and retention of important signal features, allowing frequencies below a certain cutoff point to pass through unchanged, while reducing the amplitude of frequencies above this point.

3.2 Feature Engineering

In their research, (YANG et al., 2022) collaborated with clinical professionals to ensure that the features extracted from the tremor data accurately represented the real-world symptoms of Parkinson’s Disease. Their approach to feature extraction was tailored

to capture the specific motor challenges PD patients face during the finger-tapping task. They focused on three key metrics:

- **Tapping Rate:** This feature measures how frequently the tapping occurs. It is calculated by identifying the low points in the signal’s waveform, which correspond to each tap. The number of these points is then divided by the total duration of the task, standardized to five seconds in all videos, providing the rate of tapping.

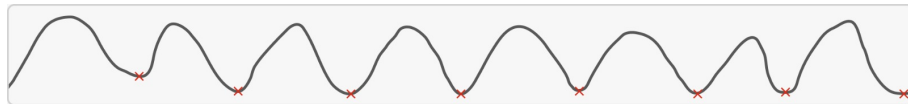


Figure 4 – Tapping rate illustration pinpointing local minimas

- **Tapping Frozen Times:** This metric identifies moments when the patient’s movements show a temporary freeze, indicative of periods where there is hesitation and little to no movement. By calculating the first derivative of the waveform, they could identify intervals with minimal change, signifying these frozen moments.



Figure 5 – Tapping frozen state, indicating periods of little movement during the task

- **Tapping Amplitude Variation:** This feature assesses the consistency of the tapping amplitude, reflecting the patient’s ability to maintain uniformity in their movements. Variability in amplitude can be indicative of their motor challenges. To quantify this, the standard deviation of the tapping amplitude is calculated, providing a measure of how much the amplitude varies over the course of the task.

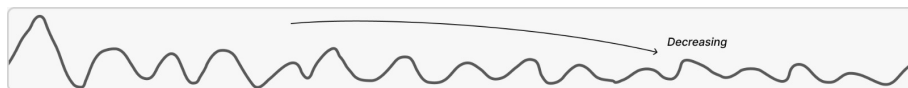


Figure 6 – Tapping amplitude with a declining trend observed throughout the task

3.3 Modeling Architecture

To replicate their proposed architecture, we used TensorFlow (ABADI et al., 2015) to create the model and train it on the created dataset with the finger tapping signal. In order to provide an accurate and comprehensive understanding, we include a direct quote

from (YANG et al., 2022) below, which details the specific architecture of the model they developed.

"We use the data with three dimensions to train a classifier with a deep neural network (DNN). Our DNN classifier has three layers and each layer has 8 hidden nodes. We use adam learner with learning rate: 0.01 and weight decay: 1e-6. The number of the training epoch is 200 and the training loss is Softmax. Then we test the method on the test set. We report the precision, recall, f1-score, and confusion matrix of the DNN-based classifier. Type 0, type 1, type 2, and type 3 represent the normal, slight, mild, and moderate severity of MDS-UPDRS respectively." (YANG et al., 2022)

In our replication of the classification algorithm, the choice of training loss required an adaptation. While (YANG et al., 2022) mentioned the use of softmax as the training loss in their paper, this seemed atypical for our purposes since softmax is generally employed as an activation function in the output layer for classification tasks. For a training loss function, especially in a multiclass classification context like ours, categorical cross-entropy is a better fitting choice to measure the error between the predicted probabilities and the one-hot representation of the class labels. Therefore, in our implementation, we opted for categorical cross-entropy as our loss function to better align with the common practices in machine learning for handling problems with multiple classes.

Another adaptation we introduced was the implementation of 5-fold cross-validation (CV), diverging from the original approach where the authors only made use of the hold-out approach, with a train-test split. This decision was influenced by our review of the literature, where 5-fold CV emerged as the most popular validation technique. The choice of cross-validation is grounded in its ability to provide a more comprehensive evaluation of the model's performance. Unlike a single train-test split, cross-validation systematically divides the dataset into multiple subsets, ensuring that each data point is used for both training and validation at different iterations. This process significantly mitigates the risk of model overfitting and enhances the generalizability of the results. Overfitting occurs when a model is too closely tailored to the specifics of the training data, potentially failing to perform well on unseen data. Cross-validation counters this by repeatedly assessing the model across different data samples, thus offering a more reliable estimate of its performance in real-world scenarios. By incorporating 5-fold CV in our study, we aimed

not only to adhere to a rigorous validation standard seen in current literature, but also to ensure that our replication and subsequent findings are as robust and generalizable as possible.

3.4 *Cleaning the Dataset*

Another important adaptation in our research involved addressing quality issues with the dataset provided by (YANG et al., 2022). While they took necessary measures to protect participant privacy by adding a black box over faces in the videos, this approach presented unique challenges. Notably, there were instances where participants moved their hands over the covered area of their faces, leading to a loss of crucial hand movement data. Moreover, we observed occurrences where the black box erroneously appeared even in the absence of a face in the frame, inadvertently obscuring hand movements. Given that these instances resulted in incomplete or obscured data, which is vital for our analysis, we chose to exclude such videos from our training dataset, corresponding to approximately 10% of the original dataset. This step of cleaning the dataset by removing videos with obstructed hand movements is essential to ensure the integrity and reliability of our training data, allowing for a more accurate and effective analysis in our AI model. For the left hand, videos 8, 13, 31, 35, 36, 42, 75, 95, 98, 104, 105, 116, 139, 170, 188, 189, 191, 193, 198, 214, 217, 228, 231, 233, 236, 244, 251, 253, 255, 258, 262, 263, 271, 282, 284, 294, 305, 307, 312, 338, 343 and 368 were removed, while for the right hand, videos from participants 17, 34, 35, 41, 54, 60, 63, 71, 74, 85, 88, 105, 118, 138, 139, 146, 159, 178, 194, 195, 197, 203, 212, 214 and 243 are excluded.



Figure 7 – Participant L8 performing the finger tapping task with the hand near their face while a black box covers their identity.



Figure 8 – Instance from participant L31 where a black box appears over the hand movements, despite the absence of a face in the frame, resulting in unintended obstruction of crucial data.

Having detailed the three steps of (YANG et al., 2022) methodology and successfully implemented them using our own code, we are now prepared to delve into the next critical phase of our research. In the upcoming section, we will thoroughly discuss the various experiments conducted, and the results obtained from these endeavors. This exploration will include an in-depth analysis of the findings, comparing them against the benchmarks set by (YANG et al., 2022), and evaluating the effectiveness of our replication and experiments.

4 Proposed Approach

Our proposal introduces a significant improvement in AI-driven diagnostics for tremor disorders based on the methodology originally developed by (YANG et al., 2022) through our proposed PClassifier model. This model emphasizes two critical enhancements: firstly, the application of data augmentation techniques to broaden the diversity and volume of our training dataset, specifically rotating and mirroring the videos, to increase our dataset by a factor of eight. This expansion aims to enhance the model’s learning capacity by introducing a broader variety of data representations, thereby improving its generalization ability across unseen data.

Secondly, we believe that relying solely on three features, as in the original model, may restrict the model’s learning capability. To address this, we incorporate additional statistical features derived from the tapping signal into the model. These features are intended to enrich the entropy of the information fed into the model, ensuring a more complex and nuanced learning process. By broadening the feature set, we aim to provide the deep learning model with a more comprehensive understanding of tremor characteristics, potentially leading to more accurate and reliable classification outcomes.

This enhanced approach, combining data augmentation and the integration of a richer feature set in our PClassifier, represents our effort to advance the performance of AI-driven diagnostics in the domain of tremor disorder classification.

To showcase our approach, Figure 9 outlines our PClassifier, highlighting the two improvements added, for the augmented dataset and the new statistical features from the tapping signal, illustrating the differences between our approach and the original framework by (YANG et al., 2022).

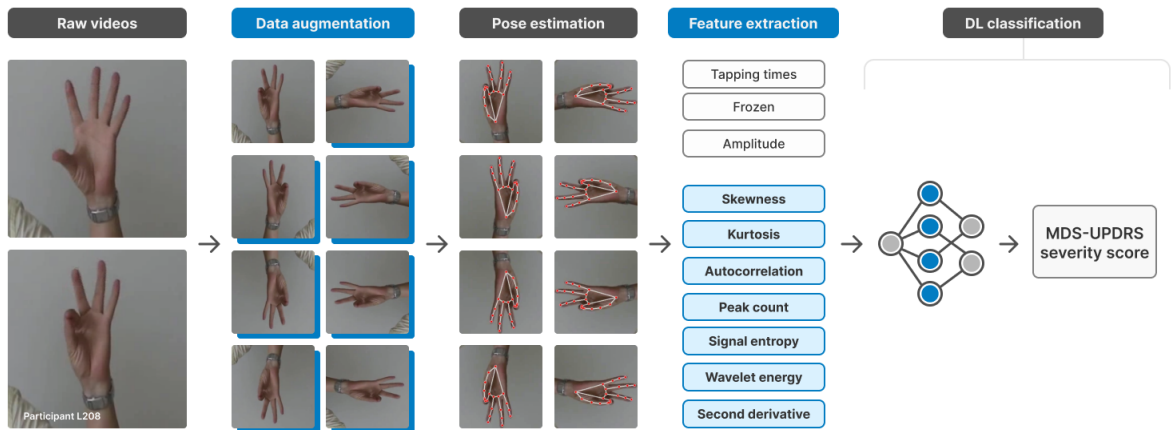


Figure 9 – Overview of the PClassifier for AI-assisted diagnosis of hand tremor disorders, highlighting data augmentation techniques and the incorporation of new statistical features.

4.1 Original Results Reported by Yang et al.

Before exploring our replication of (YANG et al., 2022) methodology, it is essential to review the results they reported in their paper. For the left-hand finger tapping task, (YANG et al., 2022) achieved averages of 88% in precision, 89% in recall, and 88% in f1-score. For the right hand, the results were slightly lower with precision, recall, and f1-score at 85%, 85%, and 84% respectively. These figures indicate a satisfactory performance of their method in analyzing hand tremors, suggesting its potential effectiveness in tremor analysis. The different results for the left and right hand are attributed to the difference in the data distribution between the two datasets.

However, as previously mentioned, (YANG et al., 2022) did not provide specific details on how they validated their models, only sharing these average performance metrics. This lack of detailed validation information poses a challenge to fully understanding the robustness of their methodology. It also highlights the importance of our experiment to replicate and validate their approach under our implementation.

In our study, we aimed to closely replicate (YANG et al., 2022) method with the necessary adaptations we previously discussed, such as the choice of the training loss function, dealing with obscured data in the videos, and the usage of 5-fold cross-validation, the most commonly used technique in the reviewed literature. Our objective was to see if

we could replicate their reported results, thereby establishing a baseline for the effectiveness of their method.

4.2 Replicating Yang et al. Methodology

After implementing their approach with our adaptations, we evaluated the performance of our model. For the left-hand dataset of the finger tapping task, our model achieved an average of 85% precision, 84% recall, and 83% in f1-score. For the right hand, the results showed a slight decrease, with the model attaining 83% precision, 80% recall, and 80% in f1-score. These results are somewhat lower than those reported by (YANG et al., 2022), which could be attributed to the adaptations we made in our methodology or the lack of a specified validation method by the original authors, in contrast to our usage of cross-validation to ensure a more robust validation of our model. This process reduces the risk of overfitting the model, and also helps in assessing the model’s ability to generalize, which is crucial when dealing with real-world medical scenarios where data variability is inherent. A model that generalizes well is more likely to maintain its performance and accuracy when exposed to new, unseen data, making it more reliable for practical applications. Despite this, our replication attempt demonstrates the robustness of the original methodology and provides a solid baseline for further exploration and refinement in our subsequent experiments.

4.3 Data Augmentation

In our subsequent experiment, we turned our focus to data augmentation, a technique commonly used in deep learning and especially helpful when dealing with video images, to expand the size of the available dataset and improve model generalization (CAULI; RECUPERO, 2022). We applied two primary techniques to our video dataset: rotation and mirroring. Specifically, we rotated the videos by 90, 180, and 270 degrees and also created mirrored versions of each original and rotated video, as shown in Figure 10. This process effectively increased our dataset eightfold.

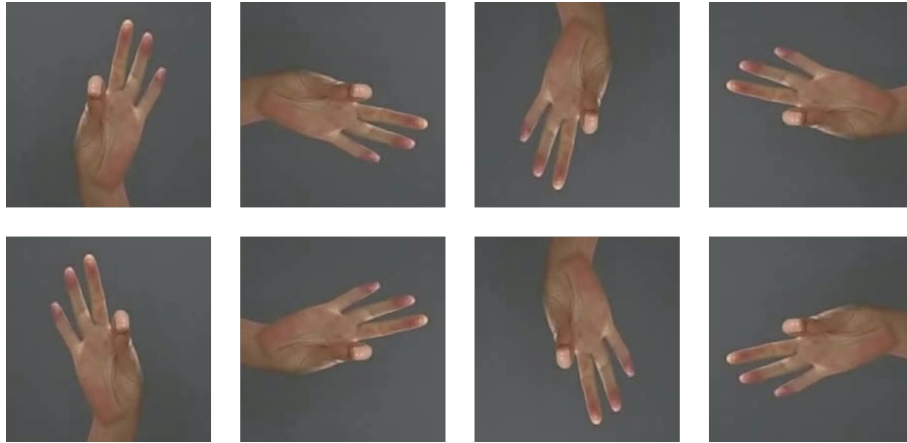


Figure 10 – Example of video from participant L40 after applying rotation and mirroring techniques for data augmentation.

By expanding our dataset through these augmentations, our aim was to enhance the model’s ability to generalize across various scenarios, reducing the risk of overfitting and potentially improving its diagnostic accuracy. Data augmentation is a well-established practice in deep learning, particularly when dealing with limited datasets, as it can significantly aid in training more robust models (CAULI; RECUPERO, 2022).

However, applying those two techniques alone did not show a significant improvement in the model’s performance. For the left hand finger tapping task, the augmented data led to the model achieving 86% precision, 84% recall, and 85% in f1-score. For the right hand, the results were slightly different, with precision at 83%, recall at 85%, and f1-score at 84%. While data augmentation is generally effective in enhancing model performance by providing more varied training data, our experiment indicated that these techniques alone might not be sufficient for significant improvement in the context of PD tremor analysis.

4.4 *Implementing New Features*

In our next trial, in pursuit of enhancing the classifier model’s performance, we conducted an experiment where we expanded upon the original set of features used by (YANG et al., 2022). Recognizing that their deep neural network (DNN) model was initially limited to just three features, we hypothesized that such a complex model could benefit from a broader range of features to capture more nuances of hand tremors.

We concentrated our efforts on selecting features that could be directly extracted from the finger tapping signal. This approach was strategic, ensuring that we did not

deviate significantly from the original methodology, which would have represented a more substantial alteration to the code. By doing so, we maintained the core essence of (YANG et al., 2022) approach while introducing enhancements that could potentially yield more detailed insights. For this experiment, we introduced the following new features: skewness, kurtosis, autocorrelation, peak count, signal entropy, wavelet energy, and the second derivative.

With the skewness, we have a feature that measures the asymmetry of the signal distribution, and the kurtosis can assess its peakedness, indicating more pronounced or smoother peaks. The autocorrelation measures the similarity of the signal with a delayed version of itself, which could help to identify the repetitive patterns and rhythms of the task. The peak count is complementary to the original feature that measures the local minima to identify the taps during the task, while the entropy measures the randomness or unpredictability in the signal, which can indicate irregular patterns. The wavelet energy provides information about both frequency and time, with the intensity distribution of the tremor signal across different frequencies. Finally, the second derivative of the signal provides information about the acceleration of the finger movements and is complementary to the tapping frozen feature that calculates the first derivative. The inclusion of these features aligns with their established efficacy in time-series analysis across various domains. Each feature is known for providing important statistical insights that are useful in distinguishing patterns within time-series data (ZIEMANN, 2021) (BAI; NG, 2005) (SHARDT; HUANG, 2013) (CHAOVALIT et al., 2011).

The results of this experiment indicated an improvement in the performance, although not yet a statistically significant one. For the left hand finger tapping task, the model equipped with new features achieved 87% precision, 88% recall, and 87% f1-score. For the right hand, the results were 85% precision, 84% recall, and 83% f1-score. Despite the theoretical advantages of enriching the model with a broader range of features to capture more nuances of the tremor patterns, these enhancements alone did not translate into a marked improvement in the model’s diagnostic capabilities. This outcome suggests that while the inclusion of additional features contributes to a more detailed analysis, it alone was not sufficient to achieve a significant leap in diagnostic accuracy.

4.5 *Combining Techniques*

In our final experiment, we combined the data augmentation techniques with the introduction of new features to evaluate their collective impact on the model’s performance. This approach aimed to capitalize on the strengths of both strategies, enhancing the dataset’s diversity through rotations and mirroring while also deepening the analysis with additional features like skewness, kurtosis, and autocorrelation. The rationale behind this combined methodology was to create a more robust model and better capture the intricate patterns of hand tremors more effectively. By integrating these strategies, we anticipated a synergy that might lead to a significant improvement in the model’s performance and reliability in classifying tremor patterns.

With this experiment, we observed a remarkable improvement from our baseline results. Originally, our baseline model, closely replicating (YANG et al., 2022) methodology, had achieved results in the low 80s for precision, recall, and f1-score. With our enhanced approach, the performance metrics for the left-hand finger tapping task increased into the low 90s, achieving 92% in precision, recall, and f1-score. Similarly, for the right hand, the model displayed a significant boost, reaching 92% precision, 91% recall, and 91% f1-score. This substantial enhancement from our initial baseline results demonstrates the effectiveness of this combined strategy. The integration of a more diverse dataset through augmentation techniques, along with the incorporation of a richer set of analytical features, proved to be instrumental in elevating the model’s accuracy and reliability. This outcome not only validates our approach but also exemplifies the potential for improvements in AI-driven diagnostic tools.

4.6 *Statistical Analysis of Results*

To evaluate the effectiveness of the enhancements in our PClassifier, we conducted a statistical analysis performing a paired t-test. We will compare the baseline results we obtained by replicating the original methodology from (YANG et al., 2022) against the results from our PClassifier, where we combined data augmentation with the new features we presented. We employed the paired t-test using the Scipy (VIRTANEN et al., 2020)

library to calculate the values. This involved comparing the f1-score from the baseline model with those from the enhanced model across each fold of our cross-validation.

Comparing the results we obtained by replicating the original methodology by (YANG et al., 2022) against our proposed PClassifier in the statistical evaluation of the F1 score, we observed significant outcomes, which serve as a comprehensive measure of the model’s diagnostic accuracy. Specifically, the analysis produced p-values of 0.002 for the F1 score concerning the left hand, and an even more compelling p-value of 0.0001 for the right hand’s F1 score. These values indicate statistically significant improvements in the model’s ability to classify tremor severity for both hands. This outcome allows us to confidently reject the null hypothesis, which posed that the enhancements to the dataset and the inclusion of a more comprehensive set of features would not result in a meaningful improvement in the model’s F1 score, highlighting the efficacy of our enhancements.

4.7 Discussion of the Results

Our findings have demonstrated a statistically significant improvement in the classifiers’ F1 scores, with p-values less than 0.05. This statistical threshold is a widely accepted benchmark for determining significance in scientific research. By achieving results that fall below this threshold, our study highlights the effectiveness of the proposed enhancements to the diagnostic model and refutes the null hypothesis.

The null hypothesis posed that the introduction of data augmentation techniques and the expansion of the input layer with a more comprehensive set of statistical features would not lead to any significant improvement in the model’s performance. However, the empirical evidence gathered from our experiments rejects this assumption. The observed p-values, indicative of a significant difference in F1 scores before and after the implementation of our PClassifier, supports our original hypothesis, and confirms that our methodological improvements have a tangible impact on enhancing diagnostic accuracy.

By rejecting the null hypothesis, we underscore the critical role of innovative data processing and feature enhancement techniques in overcoming the limitations of existing models, paving the way for more accurate and reliable diagnostic tools. A structured approach for preprocessing the dataset is often emphasized as a crucial step for enhancing machine learning model’s performance (MAHARANA; MONDAL; NEMADE, 2022). The

removal of noisy data to clean the dataset, the normalization to ensure that the variability in the data does not negatively impact the accuracy of predictions or introduce any biases, the data augmentation steps to increase the input space, and the engineering of new features significantly improve the model's predictive capabilities.

5 Lessons Learned

Throughout our research, we encountered a series of challenges and learning opportunities. This chapter is dedicated to distilling these experiences, focusing particularly on our initial methodological approach and the complexities of navigating institutional ethics approval.

5.1 *Preprocessing Videos Using CLAHE*

In our first experiment, we aimed to enhance the accuracy of pose estimation for our dataset, inspired by the challenges addressed in the study by (CHEN et al., 2021). In a similar condition in which a video dataset was recorded in a clinical environment, they faced issues in estimating poses of bedridden patients, where varied lighting conditions and other environmental factors led to inaccuracies. To improve the quality of their pose estimations, (CHEN et al., 2021) implemented CLAHE as a preprocessing step, enhancing video clarity under different lighting conditions.

Contrast Limited Adaptive Histogram Equalization (CLAHE) is an advanced method for improving the contrast in images (ZUIDERVELD; HECKBERT, 1994). The technique is an extension of Adaptive Histogram Equalization (AHE) (PIZER et al., 1987). AHE works by dividing the image into smaller tiles and applying histogram equalization to each of these tiles independently. This local approach allows AHE to adapt to specific regions in an image, enhancing local contrast and bringing out more details, especially in areas that are darker or lighter than the majority of the image. This results in enhancing the contrast and, consequently, the visibility of features within each local region of the image. However, one of the drawbacks of AHE is that it can overamplify noise in relatively homogeneous regions of an image, as it treats each tile independently (ZUIDERVELD; HECKBERT, 1994).

CLAHE addresses the issue of noise amplification through modifying the AHE process by introducing a contrast limiting step, which prevents the over-enhancement of contrast in an image, thus reducing the problem of amplifying noise. In CLAHE, after dividing the image into tiles and applying the histogram equalization, any histogram bin that exceeds a specified contrast limit is clipped. The excess from these bins is

then redistributed evenly across other bins, ensuring that the enhancement stays within acceptable limits. By limiting the contrast enhancement, CLAHE ensures that the resultant image is more natural and less prone to noise compared to standard AHE.

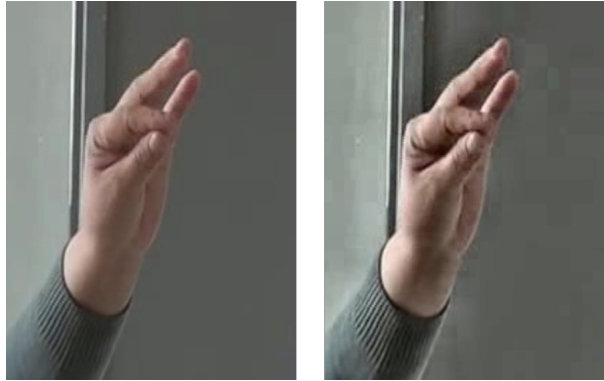


Figure 11 – Frame from participant L40 showing original video quality (left) and the result after preprocessing it with CLAHE (right).

In the context of our study, applying CLAHE to the hand tremor videos was intended to enhance the quality of these videos by improving the visibility of hand movements, which are essential for accurate pose estimation. However, despite the theoretical advantages of CLAHE in enhancing image quality, our experiment revealed that this did not necessarily translate into improved performance of the pose estimation algorithm in our specific case. For the left hand finger tapping task, the model achieved an average precision of 86%, recall of 82%, and an f1-score of 82%. For the right hand, the results were slightly different, with the model attaining a precision of 84%, recall of 82%, and an f1-score of 82%. While the application of CLAHE did lead to a marginal improvement in precision, especially for the left hand, it did not significantly enhance the overall performance of the model as we had hypothesized. This outcome suggests that while CLAHE improved the video quality, its impact on the accuracy of pose estimation and subsequent classification was limited.

5.2 Ethics Committee Process

In exploring practical applications and collaborations for our research, we engaged in discussions with a medical doctor from the movement disorder center at Hospital Universitário Gaffrée e Guinle (HUGG). The clinician expressed keen interest, particularly due to its potential to enhance patient assessment and tracking. Currently, while the finger tapping task is routinely performed with patients, these sessions are not recorded. This

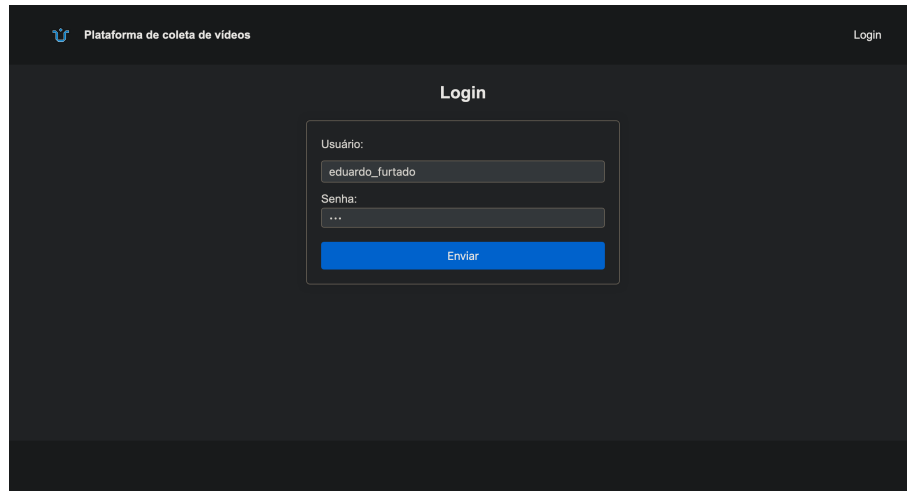
lack of documentation poses challenges in monitoring patients' progress over time, as it relies heavily on the doctor's memory and notes from previous assessments. This situation highlighted an opportunity to develop a platform that could aid in patient disease tracking while also collecting a valuable dataset for artificial intelligence research. The integration of this platform into their practice could offer a significant improvement, enabling the doctors to record each session, creating a valuable database of patient assessments. This not only aids in tracking the progression of a patient's condition over time but also allows for a more objective and detailed analysis during follow-up visits, while producing an important dataset that can be used to train models and help doctors with AI-assisted tools.

Due to the extensive process required for ethical committee approval, the implementation of this solution in a clinical setting was not feasible within the timeframe of this dissertation. Despite this, the platform represents a ready-to-use solution for future research and clinical application. To promote collaborative research and further development, the code for this platform, developed using Python with the Flask framework, along with all the code used in this dissertation, is made available for public use. This open-source approach encourages other researchers and clinicians to adapt and employ this tool in their own studies and practices.

To provide a clear understanding of the platform's interface and functionality, several screenshots have been included below. These images showcase the platform's user-friendly design, featuring aspects such as adding a participant, video recording, and listing participants and videos.



Figure 12 – Landing page of the video collection platform



Plataforma de coleta de vídeos

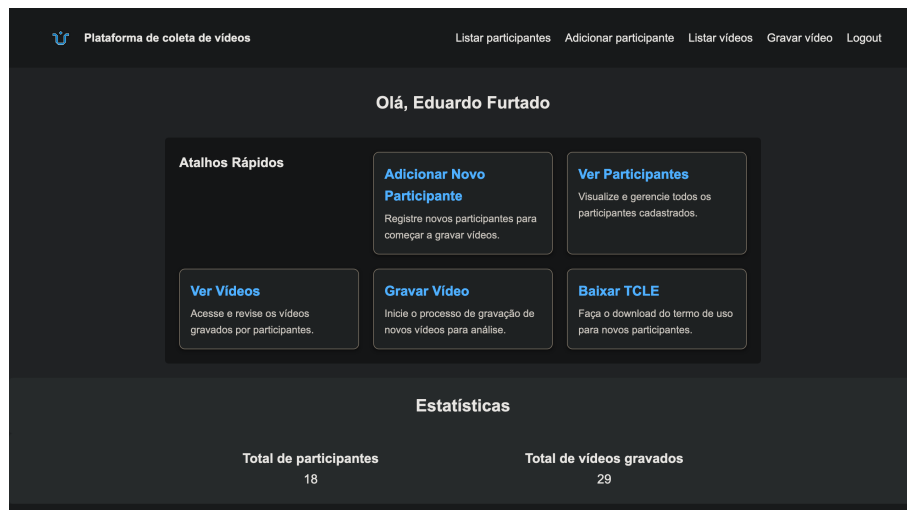
Login

Usuário:
eduardo_furtado

Senha:
...

Enviar

Figure 13 – Login page of the platform



Plataforma de coleta de vídeos

Listar participantes Adicionar participante Listar vídeos Gravar vídeo Logout

Olá, Eduardo Furtado

Atalhos Rápidos

Adicionar Novo Participante
Registre novos participantes para começar a gravar vídeos.

Ver Participantes
Visualize e gerencie todos os participantes cadastrados.

Ver Vídeos
Acesse e revise os vídeos gravados por participantes.

Gravar Vídeo
Inicie o processo de gravação de novos vídeos para análise.

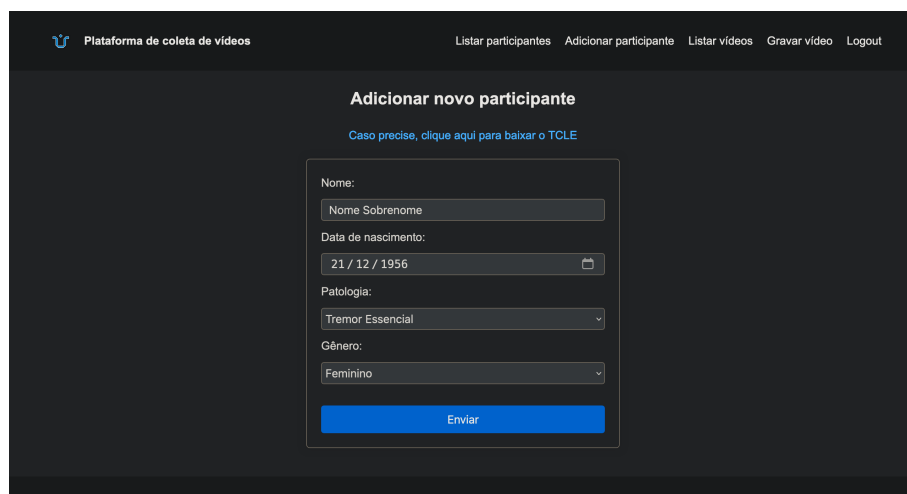
Baixar TCLE
Faça o download do termo de uso para novos participantes.

Estatísticas

Total de participantes
18

Total de vídeos gravados
29

Figure 14 – Homepage when the user is logged in



Plataforma de coleta de vídeos

Listar participantes Adicionar participante Listar vídeos Gravar vídeo Logout

Adicionar novo participante

Caso precise, clique aqui para baixar o TCLE

Nome:
Nome Sobrenome

Data de nascimento:
21 / 12 / 1956

Patologia:
Tremor Essencial

Gênero:
Feminino

Enviar

Figure 15 – Doctors can add new participants by inputting their information, such as age and pathology

Plataforma de coleta de vídeos

Listar participantes Adicionar participante Listar vídeos Gravar vídeo Logout

Lista de participantes

Busca por nome de participante

ID	Nome	Patologia	Data de nascimento	Gênero	Vídeos	Criado por	Ações
1	Eduardo Furtado	Tremor Essencial	27/12/2000	M	0	Eduardo Furtado	Visualizar
2	João Pessoa	Parkinson	07/02/2000	F	5	Eduardo Furtado	Visualizar
3	Pessoa Test	Outro	11/11/1984	O	4	Ana Cristina Bicharra Garcia	Visualizar
4	Maria Pessoa	Parkinson	22/02/1922	F	2	Eduardo Furtado	Visualizar
5	Pessoa Sauro	Outro	11/11/2001	O	0	Eduardo Furtado	Visualizar
11	Mary Doe	Tremor Essencial	11/11/1991	F	0	Ana Cristina Bicharra Garcia	Visualizar
16	Aurora Borealis	Parkinson	22/02/1952	O	3	Ana Cristina Bicharra Garcia	Visualizar
17	James Bond	Parkinson	01/01/1981	M	8	Eduardo Furtado	Visualizar
18	John Smith	Outro	02/01/2001	M	1	Eduardo Furtado	Visualizar

Figure 16 – List of all participants in the platform

Plataforma de coleta de vídeos

Listar participantes Adicionar participante Listar vídeos Gravar vídeo Logout

Participante

ID: 18
Nome: John Smith
Patologia: Outro
Data de nascimento: 02/01/2001
Gênero: M

Participante criado por Eduardo Furtado

[Editar](#) [Deletar](#)

Vídeos gravados pelo participante: 1

[Gravar vídeo para esse participante](#)



0:00 / 0:10

Avaliação MDS-UPDRS: 4
Comentários: good
Gravado em: 10/11/2023 11:58:33
Gravado por: Eduardo Furtado

Figure 17 – User page of the participant, where doctors can visualize their information, edit and save new information, and have access to their history of recorded videos

Plataforma de coleta de vídeos						
Listar participantes Adicionar participante Listar vídeos Gravar vídeo Logout						
Lista de vídeos						
Participante	Patologia	Data	Laudo MDS-UPDRS	Comentários	Gravado por	Link
Maria Silva	Tremor Essencial	11/11/2023 17:46:23	3	fast	Ana Cristina Bicharra Garcia	Ver Vídeo
Maria Silva	Tremor Essencial	11/11/2023 17:40:56	4	-	Eduardo Furtado	Ver Vídeo
Bruna Almeida	Parkinson	10/11/2023 15:04:19	4	-	Ana Cristina Bicharra Garcia	Ver Vídeo
John Smith	Outro	10/11/2023 11:58:33	4	good	Eduardo Furtado	Ver Vídeo
James Bond	Parkinson	10/11/2023 11:48:35	3		Eduardo Furtado	Ver Vídeo
Aurora Borealis	Parkinson	10/11/2023 11:42:32	1	poor lighting	Eduardo Furtado	Ver Vídeo
James Bond	Parkinson	10/11/2023 11:37:07	2		Eduardo Furtado	Ver Vídeo
James Bond	Parkinson	10/11/2023 11:33:50	0	dark	Ana Cristina Bicharra Garcia	Ver Vídeo
James Bond	Parkinson	10/11/2023 11:16:36	4	light	Eduardo Furtado	Ver Vídeo
James Bond	Parkinson	09/11/2023 23:38:38	4	quit	Eduardo Furtado	Ver Vídeo
James Bond	Parkinson	09/11/2023 23:36:18	2	good	Eduardo Furtado	Ver Vídeo

Figure 18 – The list videos page showing all videos recorded on the platform

Plataforma de coleta de vídeos
[Listar participantes](#)
[Adicionar participante](#)
[Listar vídeos](#)
[Gravar vídeo](#)
[Logout](#)

Teste de Finger Tapping

Bata o dedo indicador no polegar como mostrado no exemplo abaixo
Abra o dedo o máximo que conseguir e faça o movimento o mais rápido possível

Evite mostrar seu rosto durante a gravação para preservar sua privacidade
Posicione a câmera de forma a capturar apenas suas mãos



Clique no botão ou aperte a barra de espaço para avançar

Continuar

Figure 19 – Before recording a new video, an instruction page is shown to the participant informing best practices on how to perform the finger tapping task

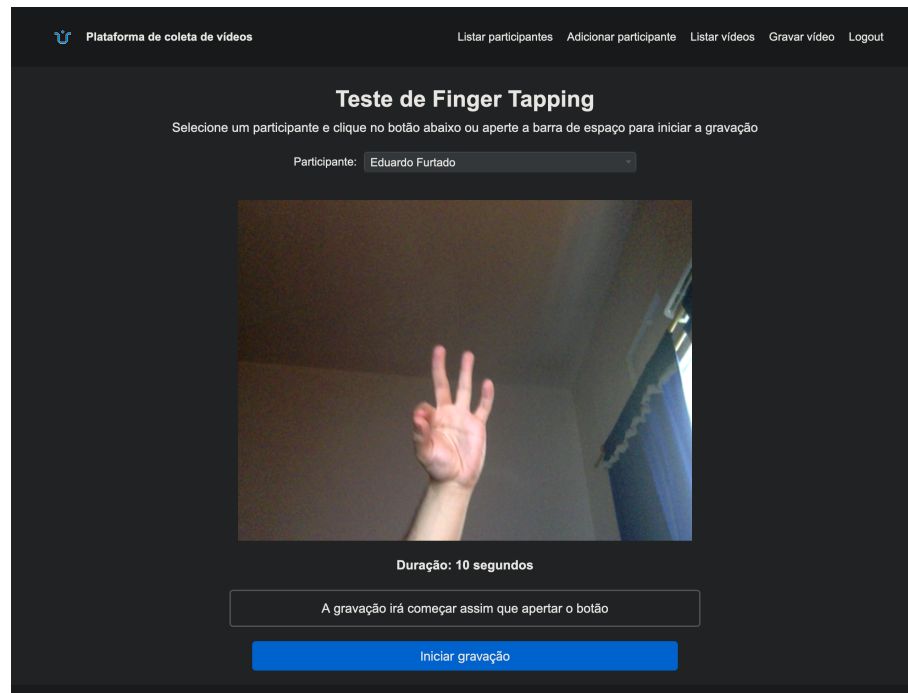


Figure 20 – The record page of the platform. When recording, a countdown allows the participant to know for how long they need to perform the task

The lessons learned from our initial approach and the challenges faced with the ethics committee have not only informed the direction of our current research but also provided a roadmap for navigating similar challenges in the future. By sharing these insights, we hope to contribute to the collective knowledge base and assist future researchers in their endeavors in the field of AI-assisted diagnostics.

6 Conclusion

This dissertation has contributed to the field of AI-assisted diagnosis of hand tremors in Parkinson’s Disease, beginning with a comprehensive literature review. The insights from this review, which was presented at the ICAART conference in Rome, laid the groundwork for our experimental research. This review helped identify patterns and gaps in current methodologies and provided direction for our subsequent experiments. The review highlighted the minimal use of preprocessing techniques in existing studies and the general lack of extensive data, driving the focus of our experimental work.

An important hurdle we overcame in our research was gaining access to clinical data used for hand tremor diagnosis, the PDMotorDB dataset. Despite the original authors not making their source code publicly available, we meticulously replicated their methodology based on the detailed descriptions in their paper. This replication not only served as a foundational step in our research, but also reinforced the importance of accessibility in scientific work. In the spirit of contributing to the broader research community, we have made our source code openly available, providing a resource for other researchers and potentially accelerating advancements in the field.

Our research journey has led to the development and testing of significant enhancements to the baseline AI-driven diagnostic model for hand tremor disorders. Central to our experimental work was the implementation of data augmentation and the incorporation of an expanded set of features derived from the tremor signal. This focused approach resulted in a marked improvement in the model’s F1 score, demonstrating the effectiveness of our PClassifier in enhancing diagnostic performance.

In conclusion, this dissertation contributes to the growing body of knowledge in AI-assisted medical diagnostics, particularly in the analysis of hand tremors in PD. Our research demonstrates that while individual enhancements may offer limited improvements, a combination of methods can significantly enhance the performance of diagnostic models.

6.1 Future Work

Considering the temporal nature of the finger tapping input data, it becomes necessary to evaluate architectures effective at processing sequential information. The

initial model, following Yang et al.’s dense network structure, may not optimally capture the intricacies of time-series data. Therefore, an exploration into 1D Convolutional Neural Networks (CNNs) or a Recurrent Neural Networks (RNNs), such as Long Short-Term Memory (LSTM) networks, presents a promising direction. These architectures are effective in analyzing temporal patterns and long-range dependencies, which could be a potential strategy to improve the model’s F1 score.

Another key area for future investigation involves analyzing the significance of the different features used in predicting tremor severity. SHapley Additive exPlanations (SHAP) (LUNDBERG; LEE, 2017) offers a robust framework designed to interpret the output of deep learning models by attributing prediction outcomes to individual features. It leverages the concept of Shapley values from cooperative game theory, which quantifies the contribution of each player towards the total payoff. By attributing Shapley values to the features, SHAP quantifies the impact of each variable on the model’s output, giving a better understanding of how individual features influence predictions. In our study, employing SHAP can highlight the specific attributes of finger tapping signals that most significantly contribute to the model’s results.

Finally, one of the greatest challenges we found in developing and replicating solutions in AI-assisted diagnosis for hand tremor disorders is the lack of data. Since most research in this field is constrained by small and often private datasets, there is a pressing need for more open and extensive data sources. This is a limitation that hampers the ability to test, validate, and improve proposed diagnostic tools. To address this critical gap, we developed a web-based platform specifically designed to facilitate the collection and organization of video data related to hand tremors. This platform aims to empower clinicians with a tool that allows in tracking and assessing the disorder progression for the patients.

While our initial focus in developing this platform was to record videos of the finger tapping task for hand tremor analysis, the platform can be easily adapted to record a variety of video tasks, catering to different diagnostic needs. By not limiting its use to hand tremor videos alone, and by making the code available to the whole research community, this platform stands as a multifunctional tool, capable of broadening its utility in medical practice and making it a valuable asset in tracking patient health and progression over time.

Bibliography

- ABADI, M. et al. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. 2015. Software available from tensorflow.org. Disponível em: <https://www.tensorflow.org/>. Citado 2 vezes nas páginas 32 and 34.
- ABBOUD, H.; AHMED, A.; FERNANDEZ, H. H. Essential tremor: Choosing the right management plan for your patient. *Cleveland Clinic Journal of Medicine*, Cleveland Clinic Journal of Medicine, v. 78, n. 12, p. 821–828, dez. 2011. Disponível em: <https://doi.org/10.3949/ccjm.78a.10178>. Citado na página 10.
- ALI, M. R. et al. Spatio-temporal attention and magnification for classification of parkinson's disease from videos collected via the internet. In: *2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020)*. IEEE, 2020. Disponível em: <https://doi.org/10.1109/fg47880.2020.00008>. Citado 9 vezes nas páginas 21, 22, 23, 24, 25, 28, 29, 30, and 31.
- ARMSTRONG, M. J.; OKUN, M. S. Diagnosis and treatment of parkinson disease. *JAMA*, American Medical Association (AMA), v. 323, n. 6, p. 548, fev. 2020. Disponível em: <https://doi.org/10.1001/jama.2019.22360>. Citado na página 9.
- BACH, J. et al. Projected numbers of people with movement disorders in the years 2030 and 2050. *Movement Disorders*, Wiley, v. 26, n. 12, p. 2286–2290, out. 2011. ISSN 1531-8257. Disponível em: <http://dx.doi.org/10.1002/mds.23878>. Citado na página 10.
- BAI, J.; NG, S. Tests for skewness, kurtosis, and normality for time series data. *Journal of Business amp; Economic Statistics*, Informa UK Limited, v. 23, n. 1, p. 49–60, jan. 2005. ISSN 1537-2707. Disponível em: <http://dx.doi.org/10.1198/073500104000000271>. Citado na página 42.
- BAUMANN, C. R. Epidemiology, diagnosis and differential diagnosis in parkinson's disease tremor. *Parkinsonism & Related Disorders*, Elsevier BV, v. 18, p. S90–S92, jan. 2012. Disponível em: [https://doi.org/10.1016/s1353-8020\(11\)70029-3](https://doi.org/10.1016/s1353-8020(11)70029-3). Citado 2 vezes nas páginas 9 and 10.
- BECK, C. A. et al. National randomized controlled trial of virtual house calls for parkinson disease. *Neurology*, Ovid Technologies (Wolters Kluwer Health), v. 89, n. 11, p. 1152–1161, ago. 2017. Disponível em: <https://doi.org/10.1212/wnl.0000000000004357>. Citado na página 10.
- BELIĆ, M. et al. Artificial intelligence for assisting diagnostics and assessment of parkinson's disease—a review. *Clinical Neurology and Neurosurgery*, Elsevier BV, v. 184, p. 105442, set. 2019. Disponível em: <https://doi.org/10.1016/j.clineuro.2019.105442>. Citado na página 11.
- BHATIA, K. P. et al. Consensus statement on the classification of tremors. from the task force on tremor of the international parkinson and movement disorder society. *Movement Disorders*, Wiley, v. 33, n. 1, p. 75–87, nov. 2017. Disponível em: <https://doi.org/10.1002/mds.27121>. Citado 2 vezes nas páginas 9 and 10.
- CAULI, N.; RECUPERO, D. R. Survey on videos data augmentation for deep learning models. *Future Internet*, MDPI AG, v. 14, n. 3, p. 93, mar. 2022. ISSN 1999-5903.

Disponível em: <http://dx.doi.org/10.3390/fi14030093>. Citado 3 vezes nas páginas 26, 40, and 41.

CHANG, C.-M. et al. Improving automatic tremor and movement motor disorder severity assessment for parkinson's disease with deep joint training. In: *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2019. Disponível em: <https://doi.org/10.1109/embc.2019.8857472>. Citado 10 vezes nas páginas 21, 22, 23, 24, 25, 26, 28, 29, 30, and 31.

CHAOVALIT, P. et al. Discrete wavelet transform-based time series analysis and mining. *ACM Computing Surveys*, Association for Computing Machinery (ACM), v. 43, n. 2, p. 1–37, jan. 2011. ISSN 1557-7341. Disponível em: <http://dx.doi.org/10.1145/1883612.1883613>. Citado na página 42.

CHEN, Y. et al. PD-net: Quantitative motor function evaluation for parkinson's disease via automated hand gesture analysis. In: *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*. ACM, 2021. Disponível em: <https://doi.org/10.1145/3447548.3467130>. Citado 10 vezes nas páginas 21, 22, 23, 24, 25, 28, 29, 30, 31, and 46.

ELBLE, R. J. What is essential tremor? *Current Neurology and Neuroscience Reports*, Springer Science and Business Media LLC, v. 13, n. 6, abr. 2013. Disponível em: <https://doi.org/10.1007/s11910-013-0353-4>. Citado na página 9.

FURTADO, E.; GARCIA, A. A systematic literature review of artificial intelligence applications for diagnosing hand tremor disorders through video analysis. In: *Proceedings of the 16th International Conference on Agents and Artificial Intelligence*. SCITEPRESS - Science and Technology Publications, 2024. Disponível em: <http://dx.doi.org/10.5220/0012385400003636>. Citado na página 13.

GOETZ, C. G. 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*, Wiley, v. 23, n. 15, p. 2129–2170, nov. 2008. Disponível em: <https://doi.org/10.1002/mds.22340>. Citado na página 21.

GUO, R. et al. A tree-structure-guided graph convolutional network with contrastive learning for the assessment of parkinsonian hand movements. *Medical Image Analysis*, Elsevier BV, v. 81, p. 102560, oct 2022. Disponível em: <https://doi.org/10.1016/j.media.2022.102560>. Citado 10 vezes nas páginas 21, 22, 23, 24, 25, 26, 28, 29, 30, and 31.

JANKOVIC, J. Physiologic and pathologic tremors. *Annals of Internal Medicine*, American College of Physicians, v. 93, n. 3, p. 460, set. 1980. Disponível em: <https://doi.org/10.7326/0003-4819-93-3-460>. Citado na página 9.

KITCHENHAM, B.; BRERETON, P. A systematic review of systematic review process research in software engineering. *Information and Software Technology*, Elsevier BV, v. 55, n. 12, p. 2049–2075, dez. 2013. ISSN 0950-5849. Disponível em: <http://dx.doi.org/10.1016/j.infsof.2013.07.010>. Citado na página 15.

LI, H. et al. Automated assessment of parkinsonian finger-tapping tests through a vision-based fine-grained classification model. *Neurocomputing*, Elsevier BV, v. 441,

p. 260–271, jun 2021. Disponível em: <https://doi.org/10.1016/j.neucom.2021.02.011>. Citado 10 vezes nas páginas 21, 22, 23, 24, 25, 26, 28, 29, 30, and 31.

LI, Z. et al. An automatic evaluation method for parkinson's dyskinesia using finger tapping video for small samples. *Journal of Medical and Biological Engineering*, Springer Science and Business Media LLC, v. 42, n. 3, p. 351–363, may 2022. Disponível em: <https://doi.org/10.1007/s40846-022-00701-y>. Citado 10 vezes nas páginas 21, 22, 23, 24, 25, 26, 28, 29, 30, and 31.

LIN, B. et al. Bradykinesia recognition in parkinson's disease via single RGB video. *ACM Transactions on Knowledge Discovery from Data*, Association for Computing Machinery (ACM), v. 14, n. 2, p. 1–19, apr 2020. Disponível em: <https://doi.org/10.1145/3369438>. Citado 10 vezes nas páginas 21, 22, 23, 24, 25, 26, 28, 29, 30, and 31.

LIU, W. et al. Vision-based estimation of MDS-UPDRS scores for quantifying parkinson's disease tremor severity. *Medical Image Analysis*, Elsevier BV, v. 85, p. 102754, abr. 2023. Disponível em: <https://doi.org/10.1016/j.media.2023.102754>. Citado 11 vezes nas páginas 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, and 31.

LIU, Y. et al. Vision-based method for automatic quantification of parkinsonian bradykinesia. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, Institute of Electrical and Electronics Engineers (IEEE), v. 27, n. 10, p. 1952–1961, oct 2019. Disponível em: <https://doi.org/10.1109/tnsre.2019.2939596>. Citado 7 vezes nas páginas 21, 22, 23, 28, 29, 30, and 31.

LOCATELLI, P. et al. Classification of essential tremor and parkinson's tremor based on a low-power wearable device. *Electronics*, MDPI AG, v. 9, n. 10, p. 1695, out. 2020. Disponível em: <https://doi.org/10.3390/electronics9101695>. Citado na página 10.

LOUIS, E. D.; BENITO-LEÓN, J.; AND, F. B.-P. Self-reported depression and anti-depressant medication use in essential tremor: cross-sectional and prospective analyses in a population-based study. *European Journal of Neurology*, Wiley, v. 14, n. 10, p. 1138–1146, out. 2007. Disponível em: <https://doi.org/10.1111/j.1468-1331.2007.01923.x>. Citado na página 9.

LOUIS, E. D.; FERREIRA, J. J. How common is the most common adult movement disorder? update on the worldwide prevalence of essential tremor. *Movement Disorders*, Wiley, v. 25, n. 5, p. 534–541, fev. 2010. Disponível em: <https://doi.org/10.1002/mds.22838>. Citado na página 9.

LOUIS, E. D. et al. Depressive traits in essential tremor: impact on disability, quality of life, and medication adherence. *European Journal of Neurology*, Wiley, v. 19, n. 10, p. 1349–1354, maio 2012. Disponível em: <https://doi.org/10.1111/j.1468-1331.2012.03774.x>. Citado na página 9.

LOUIS, E. D.; MCCREARY, M. How common is essential tremor? update on the worldwide prevalence of essential tremor. *Tremor and Other Hyperkinetic Movements*, Ubiquity Press, Ltd., v. 11, n. 1, 2021. ISSN 2160-8288. Disponível em: <http://dx.doi.org/10.5334/tohm.632>. Citado na página 10.

LU, M. et al. Quantifying parkinson's disease motor severity under uncertainty using MDS-UPDRS videos. *Medical Image Analysis*, Elsevier BV, v. 73, p. 102179, oct 2021.

Disponível em: <https://doi.org/10.1016/j.media.2021.102179>. Citado 10 vezes nas páginas 21, 22, 23, 24, 25, 26, 28, 29, 30, and 31.

LUNDBERG, S. M.; LEE, S.-I. A unified approach to interpreting model predictions. In: GUYON, I. et al. (Ed.). *Advances in Neural Information Processing Systems*. Curran Associates, Inc., 2017. v. 30. Disponível em: https://proceedings.neurips.cc/paper_files/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf. Citado na página 54.

MAHARANA, K.; MONDAL, S.; NEMADE, B. A review: Data pre-processing and data augmentation techniques. *Global Transitions Proceedings*, Elsevier BV, v. 3, n. 1, p. 91–99, jun. 2022. ISSN 2666-285X. Disponível em: <http://dx.doi.org/10.1016/j.gltp.2022.04.020>. Citado na página 44.

MONJE, M. H. G. et al. Remote evaluation of parkinson's disease using a conventional webcam and artificial intelligence. *Frontiers in Neurology*, Frontiers Media SA, v. 12, dec 2021. Disponível em: <https://doi.org/10.3389/fneur.2021.742654>. Citado 10 vezes nas páginas 21, 22, 23, 24, 25, 26, 28, 29, 30, and 31.

PARKINSON, J. An essay on the shaking palsy. *The Journal of Neuropsychiatry and Clinical Neurosciences*, American Psychiatric Association Publishing, v. 14, n. 2, p. 223–236, maio 2002. Disponível em: <https://doi.org/10.1176/jnp.14.2.223>. Citado na página 9.

PEREIRA, C. R. et al. A survey on computer-assisted parkinson's disease diagnosis. *Artificial Intelligence in Medicine*, Elsevier BV, v. 95, p. 48–63, abr. 2019. ISSN 0933-3657. Disponível em: <http://dx.doi.org/10.1016/j.artmed.2018.08.007>. Citado na página 10.

PIZER, S. M. et al. Adaptive histogram equalization and its variations. *Computer Vision, Graphics, and Image Processing*, Elsevier BV, v. 39, n. 3, p. 355–368, set. 1987. ISSN 0734-189X. Disponível em: [http://dx.doi.org/10.1016/S0734-189X\(87\)80186-X](http://dx.doi.org/10.1016/S0734-189X(87)80186-X). Citado na página 46.

PUSCHMANN, A.; WSZOLEK, Z. Diagnosis and treatment of common forms of tremor. *Seminars in Neurology*, Georg Thieme Verlag KG, v. 31, n. 01, p. 065–077, fev. 2011. ISSN 1098-9021. Disponível em: <http://dx.doi.org/10.1055/s-0031-1271312>. Citado na página 9.

RANA, A. Q.; CHOU, K. L. *Essential Tremor in Clinical Practice*. Springer International Publishing, 2015. Disponível em: <https://doi.org/10.1007/978-3-319-14598-3>. Citado na página 9.

SENGUL, Y. et al. Cognitive functions, fatigue, depression, anxiety, and sleep disturbances: assessment of nonmotor features in young patients with essential tremor. *Acta Neurologica Belgica*, Springer Science and Business Media LLC, v. 115, n. 3, p. 281–287, dez. 2014. Disponível em: <https://doi.org/10.1007/s13760-014-0396-6>. Citado na página 9.

SHARDT, Y. A.; HUANG, B. Statistical properties of signal entropy for use in detecting changes in time series data. *Journal of Chemometrics*, Wiley, v. 27, n. 11, p. 394–405, set. 2013. ISSN 1099-128X. Disponível em: <http://dx.doi.org/10.1002/cem.2533>. Citado na página 42.

SHEN, J. et al. Artificial intelligence versus clinicians in disease diagnosis: Systematic review. *JMIR Medical Informatics*, JMIR Publications Inc., v. 7, n. 3, p. e10010, ago. 2019. Disponível em: <https://doi.org/10.2196/10010>. Citado 2 vezes nas páginas 10 and 15.

SIMON, T. et al. Hand keypoint detection in single images using multiview bootstrapping. In: *CVPR*. [S.l.: s.n.], 2017. Citado na página 23.

VALVERDE-ALBACETE, F. J.; PELAEZ-MORENO, C. 100 *PLoS ONE*, Public Library of Science (PLOS), v. 9, n. 1, p. e84217, jan. 2014. ISSN 1932-6203. Disponível em: <http://dx.doi.org/10.1371/journal.pone.0084217>. Citado na página 25.

VIGNOUD, G. et al. Video-based automated assessment of movement parameters consistent with MDS-UPDRS III in parkinson's disease. *Journal of Parkinson's Disease*, IOS Press, v. 12, n. 7, p. 2211–2222, oct 2022. Disponível em: <https://doi.org/10.3233/jpd-223445>. Citado 9 vezes nas páginas 21, 22, 23, 24, 25, 28, 29, 30, and 31.

VIRTANEN, P. et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, v. 17, p. 261–272, 2020. Citado na página 43.

WANG, X. et al. Hand tremor detection in videos with cluttered background using neural network based approaches. *Health Information Science and Systems*, Springer Science and Business Media LLC, v. 9, n. 1, jul 2021. Disponível em: <https://doi.org/10.1007/s13755-021-00159-3>. Citado 10 vezes nas páginas 21, 22, 23, 24, 25, 26, 28, 29, 30, and 31.

WELTON, T. et al. Essential tremor. *Nature Reviews Disease Primers*, Springer Science and Business Media LLC, v. 7, n. 1, nov. 2021. Disponível em: <https://doi.org/10.1038/s41572-021-00314-w>. Citado na página 10.

WONG, D. C. et al. Supervised classification of bradykinesia for parkinson's disease diagnosis from smartphone videos. In: *2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS)*. IEEE, 2019. Disponível em: <https://doi.org/10.1109/cbms.2019.00017>. Citado 11 vezes nas páginas 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, and 31.

YANG, N. et al. Automatic detection pipeline for accessing the motor severity of parkinson's disease in finger tapping and postural stability. *IEEE Access*, Institute of Electrical and Electronics Engineers (IEEE), v. 10, p. 66961–66973, 2022. Disponível em: <https://doi.org/10.1109/access.2022.3183232>. Citado 25 vezes nas páginas 3, 13, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 35, 36, 37, 38, 39, 40, 41, 42, 43, and 44.

ZHANG, F. et al. *MediaPipe Hands: On-device Real-time Hand Tracking*. arXiv, 2020. Disponível em: <https://arxiv.org/abs/2006.10214>. Citado 2 vezes nas páginas 23 and 33.

ZHANG, H. et al. Pose-based tremor classification for parkinson's disease diagnosis from video. In: *Lecture Notes in Computer Science*. Springer Nature Switzerland, 2022. p. 489–499. Disponível em: https://doi.org/10.1007/978-3-031-16440-8_47. Citado 10 vezes nas páginas 21, 22, 23, 24, 25, 26, 28, 29, 30, and 31.

ZHAO, A.; LI, J. Two-channel lstm for severity rating of parkinson's disease using 3d trajectory of hand motion. *Multimedia Tools and Applications*, Springer Science and Business Media LLC, v. 81, n. 23, p. 33851–33866, abr. 2022. Disponível em:

<https://doi.org/10.1007/s11042-022-12659-9>. Citado 9 vezes nas páginas 21, 22, 23, 24, 25, 28, 29, 30, and 31.

ZIEMANN, V. Time series. In: _____. *Physics and Finance*. Springer International Publishing, 2021. p. 91–112. ISBN 9783030636432. Disponível em: http://dx.doi.org/10.1007/978-3-030-63643-2_8. Citado na página 42.

ZUIDERVELD, K.; HECKBERT, P. S. Graphics gems iv. *San Diego, CA, USA: Academic Press Professional, Inc*, p. 474–485, 1994. Citado na página 46.

A Experiments Results

Table 5 – Average precision, recall, and F1-score for each of our experiments.

Hand	Experiment	New Features	Avg Precision	Avg Recall	Avg F1
Left	Original	False	0.8484968206	0.8376572920	0.8334274236
Left	Clahe	False	0.8608688472	0.8163884159	0.8234131961
Left	Augmented	False	0.8621467256	0.8388482879	0.8482954993
Left	Original	True	0.8720100449	0.8863512719	0.8668243645
Left	Augmented	True	0.9221421194	0.9207986193	0.9208444846
Right	Original	False	0.8348825758	0.8031845238	0.7959436089
Right	Clahe	False	0.8458208579	0.8243750000	0.8189512871
Right	Augmented	False	0.8351740536	0.8518559111	0.8409977346
Right	Original	True	0.8509583574	0.8438095238	0.8290771266
Right	Augmented	True	0.9206954537	0.9126387479	0.9134632842

Table 6 – Detailed precision, recall, and F1-score results by fold in cross-validation for the left hand.

Hand	Exp.	N.F.	Precision Scores	Recall Scores	F1 Scores
Left	Original	F	[0.8262000768049154, 0.8090659340659341, 0.8336151368760065, 0.9287478825522304, 0.8448550724637682]	[0.8164018952062431, 0.8767419175027871, 0.8131521739130434, 0.8365942028985508, 0.8453962703962703]	[0.8044326241134752, 0.827417148069322, 0.8213514746123443, 0.8690012305168171, 0.8449346405228759]
Left	Clahe	F	[0.8377159977703457, 0.7928321678321678, 0.8480988000623344, 0.9415292353823088, 0.884168035030104]	[0.8876114827201784, 0.8057608695652174, 0.684927536231884, 0.8647727272727273, 0.8388694638694639]	[0.8586355960824046, 0.7845452419920506, 0.7256944444444445, 0.895093795093795, 0.8530969030969031]
Left	Aug.	F	[0.841411374236367, 0.8615127524950613, 0.8796532142950683, 0.845049409270493, 0.8831068779172608]	[0.8193346589688053, 0.8263631868433697, 0.8689122572887817, 0.8483609660080249, 0.8312703703484511]	[0.8294813995988017, 0.8407687204449829, 0.8737537326401945, 0.8460873846319668, 0.8513862591223167]
Left	Original	T	[0.7449805068226121, 0.8829085457271364, 0.8538919413919415, 0.9507692307692308, 0.9275000000000000]	[0.8183528428093645, 0.8744425863991081, 0.9048188405797102, 0.8874637681159421, 0.9466783216783217]	[0.7472804972804973, 0.8668862538427756, 0.8722925216598267, 0.9139557338086749, 0.9337068160597572]
Left	Aug.	T	[0.9207427959501557, 0.9306842438012117, 0.9186855035514354, 0.9285539423416037, 0.9120441114655193]	[0.9095389123209245, 0.9082766610586732, 0.9215002773310701, 0.9350902393181805, 0.9295870064074315]	[0.9147740266552126, 0.9181927128969006, 0.919453519102427, 0.9315141702600354, 0.9202879941695731]

Table 7 – Detailed precision, recall, and F1-score results by fold in cross-validation for the right hand.

Hand	Exp.	N.F.	Precision Scores	Recall Scores	F1 Scores
Right	Original	F	[0.8638392857142857, 0.8591666666666666, 0.8601190476190477, 0.7541666666666667, 0.837121212121212]	[0.7327380952380953, 0.8011904761904762, 0.7309523809523809, 0.8791666666666667, 0.8718750000000000]	[0.7759688514999293, 0.8203703703703703, 0.7706830206830206, 0.7791666666666667, 0.8335291353383458]
Right	Clahe	F	[0.7536858974358975, 0.938664596273292, 0.8913690476190477, 0.7888888888888889, 0.8564958592132506]	[0.7053571428571428, 0.8827380952380952, 0.7619047619047619, 0.9000000000000000, 0.8718750000000000]	[0.7218614718614719, 0.9034014969259556, 0.8042930057092134, 0.820248538011696, 0.844951923076923]
Right	Aug.	F	[0.8098439094343786, 0.8669604635828769, 0.8214829236015677, 0.8781138496101494, 0.7994691219775206]	[0.8439396375851136, 0.8714060632049483, 0.8267114936370465, 0.8902393287967832, 0.8269830323365159]	[0.8252091315197468, 0.8687406575179861, 0.8174560630735845, 0.8823738561417247, 0.8112089645366483]
Right	Original	T	[0.6842105263157895, 0.9307692307692308, 0.9191666666666667, 0.7958333333333333, 0.924812030075188]	[0.7702380952380952, 0.7535714285714286, 0.8494047619047619, 0.8958333333333334, 0.9500000000000000]	[0.7051282051282051, 0.8115079365079365, 0.8731481481481481, 0.8240384615384615, 0.9315628815628816]
Right	Aug.	T	[0.921849593495935, 0.936518443842538, 0.8928559688808798, 0.9282198886639677, 0.9240333738307209]	[0.9026683267061697, 0.9280441096210822, 0.9335910918651407, 0.8571674362599973, 0.9417227751756441]	[0.9113610605140658, 0.9319908718425152, 0.911430244094523, 0.8847313237221494, 0.9278029209847392]