A Systematic Literature Review of Artificial Intelligence Applications for Diagnosing Hand Tremor Disorders Through Video Analysis

Eduardo Furtado[a] and Ana Cristina Bicharra Garcia[b]
Federal University of the State of Rio de Janeiro, Department of Applied Informatics, Rio de Janeiro, 22290-255, Brazil

Keywords: Hand Tremors, Artificial Intelligence, Video Analysis, Machine Learning, Non-Invasive Diagnosis.

Abstract: In neurodegenerative disorders, accurate diagnosis of hand tremors serves as a cornerstone for effective management and treatment plans. With the burgeoning advances in Artificial Intelligence and machine learning, substantial promise exists for devising robust and reliable diagnostic methodologies. This paper presents a systematic literature review analyzing 17 key studies that have employed machine-learning techniques to diagnose hand tremors. The scrutiny is multidimensional, elucidating the primary research objectives, patient tasks during studies, distinct features utilized by the machine learning models, and various validation techniques applied. The aim is to offer a synthesized research landscape, identifying recurring methodologies and techniques. Moreover, we seek to underscore gaps and potential avenues for future investigations. Through this systematic examination, we endeavor to contribute to the scholarly discourse, aiding the focused and coherent advancement of machine learning-based diagnostic models within this critical healthcare domain.

1 INTRODUCTION

Tremor stands out as the most prevalent involuntary movement disorder, represented by rhythmic oscillation of a body part, most commonly observed in the hands (Jankovic, 1980). Numerous underlying causes for involuntary tremors exist, such as Parkinson's disease (PD) (Baumann, 2012), Essential Tremor (ET) (Louis and Ferreira, 2010), Enhanced Physiologic Tremor, and Orthostatic Tremor, each presenting with its own distinctive frequency and potentially affecting different body regions (Rana and Chou, 2015).

Parkinson's disease, a prevalent neurodegenerative disorder, is primarily diagnosed through patient history and clinical examinations. Patients often experience movement challenges like tremors, stiffness and slowness, accompanied by psychological issues such as depression and anxiety. Clinical tests typically reveal bradykinesia (slowness of movement and speed) and rigidity (Armstrong and Okun, 2020).

Predominantly, the clinical syndrome of tremor is most pronounced in the upper limbs, impacting at least 95% of all patients (Elble, 2013). It can also manifest, albeit less commonly, in other body parts including the head, face, trunk, lower limbs, and voice (Elble, 2013). The significant impact of involuntary tremulous motion on an individual's life has been documented for centuries (Parkinson, 2002). Presently, with no known cure, the treatment for tremors remains focused on managing symptoms (Abboud et al., 2011) (Baumann, 2012).

In a period of rapid advances in medical science, integrating artificial intelligence (AI) into medical diagnostics offers new possibilities. Diagnosing movement disorders, especially Parkinson's disease, has always been challenging due to the subtle and varied symptoms and their progression. Traditional diagnostic methods, while critical, can sometimes lead to delayed diagnoses, obstructing the timely start of the best treatment for patients.

Thus, early diagnosis is crucial in managing these disorders, ensuring that patients receive the right treatment as soon as possible (Locatelli et al., 2020). Existing studies indicate that AI can match the performance of medical experts when given enough data for model training (Shen et al., 2019). Additionally, AI has proven useful in telemedicine, improving treatment access and convenience for patients (Beck et al., 2017).

Our motivation for this systematic literature review lies in evaluating current applications of AI-assisted diagnosis of tremor utilizing simple hand videos. Given that hand tremors are prevalent in

[a] https://orcid.org/0009-0005-7994-9044
[b] https://orcid.org/0000-0002-3797-5157

Furtado, E. and Garcia, A.
A Systematic Literature Review of Artificial Intelligence Applications for Diagnosing Hand Tremor Disorders Through Video Analysis.
DOI: 10.5220/0012385400003636

In *Proceedings of the 16th International Conference on Agents and Artificial Intelligence (ICAART 2024) - Volume 3*, pages 707-717
ISBN: 978-989-758-680-4; ISSN: 2184-433X

numerous movement disorders, establishing a non-intrusive and easily accessible means of identifying and assessing them could significantly streamline preliminary diagnostic processes. We seek to explore how current research employs AI, particularly through non-complex video technology, to diagnose and evaluate hand tremors, ensuring the feasibility and accessibility of such approaches for potential use in telemedicine settings and beyond. By examining the depth of existing studies, we aim to contribute a coherent understanding and critical assessment of the current state of AI applications in this domain.

2 RESEARCH METHODOLOGY

Our systematic literature review aimed to find studies on AI-assisted diagnosis of hand tremor disorders using 2D video analysis. We searched six databases: IEEE Xplore, PubMed, ACM Digital Library, ScienceDirect, Springer, and IOS Press using the following search string to capture relevant papers:

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

Although this string produced many off-topic results, we chose a broad approach to ensure we did not miss potentially relevant studies.

Five research questions were used to direct our study of the literature on artificial intelligence-assisted diagnosis of hand tremors.

1. Objective of Research: What is the primary aim of each study (e.g., classifying the type or severity of tremor)?
2. Video Tasks: 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 pre-processing 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?

The review includes studies from 2018 to 2023 that are written in English and concentrate on AI-assisted diagnosis of tremor disorders through the use of 2D videos of hands, specifically from widely used devices like smartphones and webcams. This focus on simple videos was chosen to highlight methodologies that are not only feasible but also straightforward to implement in various settings, including telemedicine and other resource-limited environments.

In contrast, we excluded studies utilizing 3D video technologies, accelerometers, and other sensors, or those examining non-video data such as handwriting images. We also excluded studies focusing on alternative diagnostic tests not related to viewing the hands, such as gait analysis, speech evaluation, head tremor, facial expressions and any combination of those with hand tremor videos as well (e.g. used hand videos with gait videos to provide the diagnostic, or used hand videos with accelerometer data). Further, duplicates, book chapters, and papers published in low-impact journals were also excluded from our corpus. This rigorous selection approach was adopted to focus on methodologies that are both low-cost and easy to deploy, aligning with our research objective of accessibility and practicality in diagnostic techniques, thus focusing on solutions using simpler hardware.

The initial search yielded a wide array of papers. We first examined titles and abstracts to check for clear relevance, and then read the full text of short-listed papers to confirm their applicability to our research questions and objectives. Through this rigorous selection process, 17 papers were identified as fulfilling our criteria and were thus included in this review. These selected studies, which vary in methodology, datasets, and research objectives, will be thoroughly reported and discussed in the following section.

3 RESULTS

In this section, we present the findings of our systematic literature review. Our approach to providing the results combines 4 tables which directly summarize and address our research questions, providing a succinct overview of data from the reviewed papers. All tables are annexed at the end of the paper for further reference.

An in-depth analysis will explore notable patterns and gaps, providing a thoughtful interpretation of the current landscape of research in the field. Figure 1 offers a concise overview of our literature review outcomes. It maps out the main objectives, datasets, feature extraction methods, and modeling approaches from the surveyed studies.

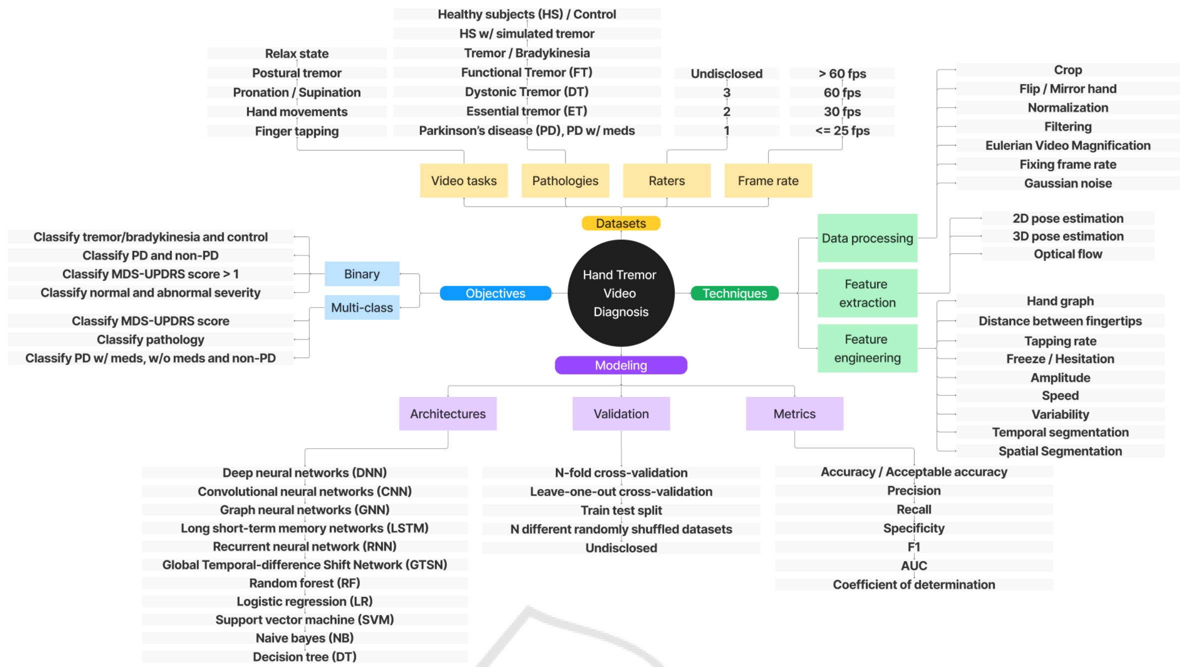

Figure 1: Overview of the literature review on AI-assisted for hand tremor diagnosis using videos.

3.1 Objectives

The majority of the selected studies, specifically (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 and Li, 2022), (Vignoud et al., 2022), (Liu et al., 2023), and (Liu et al., 2019), were dedicated to utilizing AI to classify the MDS-UPDRS score. Some researchers expanded their focus to other aspects of tremor analysis, with (Lin et al., 2020) aiming to distinguish between Bradykinesia and Healthy Subjects (HS), (Wang et al., 2021) seeking to classify Tremor and Non-Tremor instances, and (Chang et al., 2019) focusing on distinguishing between Normal and Abnormal tremor severity. A multifaceted approach was taken by (Zhang et al., 2022), investigating classification of PD and non-PD in one scenario and categorizing between Parkinson's Disease (PD), Essential Tremor (ET), Functional Tremor (FT), Dystonic Tremor (DT), and HS in another. A similar dual-objective methodology was utilized by (Ali et al., 2020), targeting classifications among PD with medication, PD without medication, and non-PD, while (Wong et al., 2019) sought to both classify MDS-UPDRS scores and differentiate between PD and non-PD subjects.

3.2 Tasks

Various tasks involving hand movements were employed in the studies to facilitate the classification and analysis of different tremor types and severities. A popular task was finger tapping, where participants touch the tip of their index with the tip of their thumb, utilized by (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 and Li, 2022), (Vignoud et al., 2022), (Liu et al., 2019). Another frequently observed task was related to generic hand movements, with participants opening and closing their hands, which was leveraged by studies conducted by (Guo et al., 2022), (Lin et al., 2020), (Chen et al., 2021), (Monje et al., 2021), (Ali et al., 2020), (Zhao and Li, 2022), (Vignoud et al., 2022), and (Liu et al., 2019). Certain studies opted for a mixture of tasks to enrich their analysis and model's predictive capabilities. For instance, pronation/supination, where participants extend their arms and turn their palm up and down, was often combined with other tasks (Chen et al., 2021), (Monje et al., 2021), (Ali et al., 2020), (Vignoud et al., 2022), (Liu et al., 2019). Additionally, (Chang et al., 2019) employed a relaxed state task and (Liu et al., 2023) utilized postural tremor as a distinctive task. Although other tasks like postural stability and gait were utilized by (Yang et al., 2022) and (Lu et al., 2021), and rest tremor

Table 1: Goals and Tasks used by selected 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 and 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

of other body parts by (Liu et al., 2023), these were outside the primary focus of our review as we concentrated on tasks involving only on the results using hands.

3.3 Datasets

The number of participants and raters in the datasets of these studies brings out some interesting observations about the current state of hand tremor diagnosis research:

Studies exhibit varied participant numbers and statuses: (Guo et al., 2022), (Li et al., 2022), and (Li et al., 2021) feature 120-174 PD participants without disclosing rater numbers. (Yang et al., 2022) involves a substantial PD participant count and three raters.

Datasets diverge in health status focus: (Lin et al., 2020), (Chang et al., 2019), and (Lu et al., 2021) explore Bradykinesia, HS, and PD participants, with (Zhang et al., 2022) adding diverse conditions and a broad video pool.

Mixed participant statuses appear in (Monje et al., 2021), (Wong et al., 2019), and (Vignoud et al., 2022), while (Ali et al., 2020) segregates PD participants by medication status, also involving HS individuals.

(Zhao and Li, 2022) limits to a smaller HS cohort but simulates varied tremor severities, lacking rater detail. (Liu et al., 2023) and (Liu et al., 2019) assure ground truth verification with multiple raters and a satisfactory PD participant count.

Rater variability impacts reliability across studies, with (Yang et al., 2022), (Chen et al., 2021), and (Liu et al., 2023) using three, and others like (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) specifying one or none. This inconsistency may challenge the robustness and applicability of findings, especially in a medical context where labeling accuracy is paramount.

In a sizable portion of the papers, like (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 and Li, 2022), (Vignoud et al., 2022), and (Liu et al., 2019), cropping is a com-

Table 2: Datasets Participants, Raters, Quality and Pre-process techniques used by selected papers.

Paper	**Participants**	**Raters**	**Quality**	**Pre-Process**
(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 MultiBox 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 and 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

mon pre-processing step. This suggests a widespread necessity to focus on the region of interest and remove irrelevant data or background noise, while (Monje et al., 2021) used a different approach using a Single Shot MultiBox Detector (SSD) to crop their region of interest.

Also, normalization, used in (Li et al., 2022), (Lu et al., 2021), and (Monje et al., 2021), allows participants to perform the tasks with their hands close or far from the camera without compromising the model's input. Another notable trend is the utilization of various filtering techniques. The Savitzky-Golay filter, applied in (Li et al., 2021), (Vignoud et al., 2022), and (Liu et al., 2019), or low pass and Fourier filtering, employed in (Yang et al., 2022) and (Chen et al., 2021) process the extract data from pose algorithms

Table 3: Features and Extraction methods used by selected papers.

Paper	**Features**	**Extraction methods**
(Guo et al., 2022)	Hand graph with 21 keypoints	2D Pose Estimation (MMPose) - 21 keypoints (+OpenPose for ROI)
(Li et al., 2022)	One-dimensional sequence data of tapping distance	2D Pose Estimation (Mediapipe) - 2 keypoints index and thumb tips
(Li et al., 2021)	Pose, Motion and Geometry features from hand graph	2D Pose Estimation (OpenPose) - 21 keypoints
(Yang et al., 2022)	Tapping rate, Tapping frozen times, Tapping amplitude variation	2D Pose Estimation (MMPose) - 2 keypoints index and thumb tips
(Lin et al., 2020)	Stability (to measure consistency of rhythm), completeness (of actions spatially) and self-similarity (stable periodic motion)	2D Pose estimation (HandSegNet + PoseNet) - 21 keypoints
(Wang et al., 2021)	Change in distance of hand movement (DIST features) and frequency of motion directional changes (MDC features)	2D Pose Estimation (MediaPipe) - 21 keypoints
(Chang et al., 2019)	Distance between keypoints, velocity and acceleration	2D Pose Estimation (OpenPose) - 9 keypoints
(Chen et al., 2021)	Slowing, Amplitude, Amplitude decrement, Hesitation/freeze, Interruption, Incompetence of performing task	2D Pose Estimation (SHG - Stacked Hourglass network + OpenPose) - 21 keypoints and
(Zhang et al., 2022)	Graph with 7 upper body keypoints	2D Pose Estimation (OpenPose) - 7 upper body keypoints
(Lu et al., 2021)	Hand graph with 21 keypoints	2D Pose Estimation (OpenPose)
(Monje et al., 2021)	Amplitude, Speed, Fatigue	2D Pose Estimation (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 and Li, 2022)	128 dimensional feature vector every 10 frames	3D Pose Estimation (HandSegNet + PoseNet + PosePrior) - 21 keypoints
(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 keypoints)
(Liu et al., 2019)	Finger Tapping: Euclidean distance between tips of thumb and index finger; Hand Clasping: Average of Euclidean distances between each fingertip and palm; Hand Pro/Supination: Difference between horizontal coordinates of thumb and little finger	MobileNet (and ShuffleNet tested)

to remove noise and generate a smoother signal.

A distinctive approach was the application of Eulerian Video Magnification (EVM) in (Liu et al., 2023) where they amplify subtle variations in the video data, and are able to highlight minute motions in patients' tremors.

Table 4: Model Techniques, Validation, Metrics and Performance reported by selected papers.

Paper	Techniques	Validation	Metrics	Performance
(Guo et al., 2022)	Tree-structure-guided graph convolutional network	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 Parkinson's test accuracy 63%
(Zhao and Li, 2022)	Two-channel LSTM	N/A	Sensitivity, Specificity and Accuracy	95.7% of the precision, 95.8% of the sensitivity and 92.8% of the specificity
(Vignoud et al., 2022)	LR, DT	100 randomly 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

3.4 Features

3.4.1 Feature Extraction

Looking through the approaches for feature extraction across these papers, there is a noticeable pattern of reliance on pose estimation, particularly 2D Pose Estimation, prevalent in various studies such as (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). The extraction of keypoints, ranging from 2 (tip of index finger and thumb) to 21 (all keypoints in the hand). A similar approach is the usage of 3D Pose Estimation by (Zhao and Li, 2022) and a mixed-method approach involving both 2D and 3D Pose Estimation by (Vignoud et al., 2022).

Differing algorithms and networks like HandSegNet + PoseNet used by (Lin et al., 2020), or SHG + OpenPose applied by (Chen et al., 2021), are also proposed methods for feature extraction.

In contrast to pose estimation, alternative methodologies include utilizing CNN, Fast Fourier Transform and Deep Neural Network-based Magnified Features by (Ali et al., 2020), Optical Flow by (Wong et al., 2019), or the temporal difference module for Optical Flow employed by (Liu et al., 2023), showcase varied approaches to understanding and capturing movement data.

3.4.2 Engineered Features

Several studies highlight the tapping-related features. For instance, (Li et al., 2022) focuses on the one-dimensional sequence data of tapping distance, while (Yang et al., 2022) considers tapping rate, frozen times, and amplitude variation. (Wong et al., 2019) explores a range of features like tapping frequency, energy spectral density, and peak-to-peak variability.

Additionally, hand graphs and geometric relations are also prominent. (Guo et al., 2022) and (Lu et al., 2021) use hand graphs with 21 keypoints, exploring hand anatomy and movement in a structured, geometrical format. (Vignoud et al., 2022) utilizes the distance between the thumb and index, providing a spatial perspective of hand movements, which is echoed by (Chang et al., 2019) who also employs distance between keypoints as a feature. Similarly, (Wang et al., 2021) considers changes in distance of hand movement, and (Liu et al., 2019) uses Euclidean distance between fingertips in different contexts, pointing towards a prevalent use of spatial and geometric features.

Amplitude and motion characteristics are common too. For instance, (Monje et al., 2021) uses amplitude as a feature, and (Ali et al., 2020) considers motion magnification.

Certain papers also introduce more task-specific features. (Lin et al., 2020) introduces the concept of stability, completeness, and self-similarity to gauge the rhythmic and spatial consistency of actions. (Chen et al., 2021) takes a more contextual approach, exploring features like slowing, amplitude decrement, and incompetence of performing a task. (Liu et al., 2023) emphasizes temporal features, signifying an interest in the time-related aspects of movements.

A few studies utilize more comprehensive and multidimensional feature sets. For instance, (Li et al., 2021) employs pose, motion, and geometry features derived from hand graphs, pointing towards an integrative approach that spans across spatial, temporal, and kinematic domains. (Zhang et al., 2022) crafts a graph with 7 upper body keypoints, indicating a broader, body-inclusive approach to understand and analyze motion.

3.5 Modeling

3.5.1 Techniques and Architectures

Graph Convolutional Networks (GCNs) are utilized by (Guo et al., 2022) for their ability to capture hierarchical relationships, and by (Zhang et al., 2022), which uses a Graph Neural Network with a Spatial Attention Mechanism to capture spatial dependencies.

Convolutional Neural Networks (CNNs) are used by (Li et al., 2022) to focus on spatial hierarchies in data, while (Li et al., 2021) uses a three-stream CNN with Markov chain fusion to capture various data facets. (Yang et al., 2022) and (Chang et al., 2019) opt for Deep Neural Networks (DNNs).

Recurrent Neural Networks (RNNs) are employed by (Lin et al., 2020) with LSTM units for managing sequential data. (Zhao and Li, 2022) uses a two-channel LSTM to handle multivariate sequential data. (Wang et al., 2021) explores CNN-LSTM, LSTM, and Support Vector Machine (SVM) models for their classification capabilities, with SVM also used by (Chang et al., 2019) and (Ali et al., 2020).

(Chen et al., 2021) applies ensemble methods (Random Forest and Gradient Boosting Decision Tree) and logistic regression (LR), and SVM. (Monje et al., 2021) utilizes LR, Naïve Bayes (NB), and RF, blending probabilistic classifiers and ensemble methods. (Wong et al., 2019) and (Vignoud et al., 2022) used NB and LR to classify their dataset. (Liu et al., 2019) tests various kernels in SVM, RF, and K-Nearest Neighbors (KNN).

A different approach by (Lu et al., 2021) employs a Temporal Convolutional Neural Network (TCNN) or OF DD-Net to manage temporal data. (Liu et al., 2023) introduces the Global Temporal-difference Shift Network (GTSN) to potentially address temporal shifts in data.

3.5.2 Validation Approaches

5-Fold Cross-Validation is frequently used in reviewed literature, seen in (Guo et al., 2022), (Li et al., 2022), (Chen et al., 2021), (Zhang et al., 2022), and (Liu et al., 2023). 10-Fold Cross-Validation is utilized by (Lin et al., 2020) and (Wang et al., 2021), providing detailed validation at a higher computational expense. (Li et al., 2021) and (Monje et al., 2021) opted for 4-Fold Cross-Validation.

(Chang et al., 2019), (Ali et al., 2020), and (Wong et al., 2019) employed Leave-One-Out Cross Validation (LOOCV), suitable for smaller datasets due to its computational intensity.

(Vignoud et al., 2022) used 100 randomly shuffled datasets for validation. (Yang et al., 2022) implemented a Train-Test Split without providing additional detail, while (Lu et al., 2021) and (Zhao and Li, 2022) did not specify their validation methodologies.

3.5.3 Metrics Selected

Accuracy is the most common chosen metric, used singularly or as combined with other metrics, utilized in (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 and Li, 2022), and (Liu et al., 2023).

Precision and Recall, often used with F1 Score, are selected in studies like (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 (also known as Recall or True Positive Rate) and Specificity (True Negative Rate) are utilized in (Zhang et al., 2022), (Monje et al., 2021), (Wong et al., 2019), and (Zhao and Li, 2022).

F1 score is employed in (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), while 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).

(Vignoud et al., 2022) is noted for using Coefficients of Determination with their predictions based on statistical learning regression algorithms.

3.5.4 Reported Performance

(Guo et al., 2022) and (Li et al., 2021) report general accuracies around 70% and notably high acceptable accuracies near 100%. (Li et al., 2022) and (Wang et al., 2021) yield stable performances with accuracies of 79.7% and 80.6% respectively. (Yang et al., 2022) achieves an F1-score of 88%, contrasted by (Lin et al., 2020)'s 77.78% F1-score. Challenges in multiclass classifications with varying accuracies are discussed by (Chang et al., 2019) and (Zhang et al., 2022). (Lu et al., 2021) and (Monje et al., 2021) emphasize AUC metrics, while high accuracies and precision in specific tasks are noted by (Ali et al., 2020), (Wong et al., 2019), and (Zhao and Li, 2022). (Vignoud et al., 2022) highlights the coefficient of determination for model predictability. (Liu et al., 2023) and (Liu et al., 2019) demonstrate strong accuracies in both binary and multiclass contexts.

4 DISCUSSION

4.1 Data Availability

One noticeable aspect from the literature is that many datasets used in these studies are not readily available to the public or other researchers. For instance, the data used in (Guo et al., 2022), (Li et al., 2021), (Lin et al., 2020), (Chang et al., 2019), (Chen et al., 2021), (Lu et al., 2021), (Ali et al., 2020), (Wong et al., 2019), (Zhao and Li, 2022), and (Vignoud et al., 2022) are either stated that are not available, or it is not mentioned at all.

While some datasets are available upon request, others implements safeguards in order to protect participants that can make it increasingly harder for researchers outside the medical field to have access. One example is (Liu et al., 2023) which imposes specific conditions, including providing proof of relevant medical studies and signing a contract.

It is also important to highlight that the TIM-Tremor dataset, used in (Wang et al., 2021) and (Zhang et al., 2022), has recently been removed from the internet due to privacy concerns. This issue underscores the critical and delicate balance between open-source data and maintaining the privacy of sensitive health-related information. Even with anonymization, healthcare datasets can sometimes be subject to potential re-identification risks or other ethical concerns, requiring vigilant management and ethical considerations.

In light of these challenges, data augmentation can be a potential solution to mitigate the scarcity of available videos in the datasets. Techniques such as video rotation and mirroring can be employed to generate new data instances from existing videos. This method, while not creating synthetic data, effectively increases the dataset size, offering a practical approach to enhance research outcomes in cases where data availability is limited.

4.2 Preprocessing Approaches

In the reviewed research, most studies tend to lean towards minimal preprocessing of data, often limiting themselves to basic techniques like cropping. This is noteworthy since the context in which the data is recorded – especially in diverse and uncontrolled clinical environments – naturally presents various challenges, such as varied lighting and cluttered backgrounds, which could significantly impact the quality and reliability of the data.

Only a handful of works, like that of (Liu et al., 2023), employ more advanced preprocessing methods, for instance, using Eulerian Video Magnification (EVM). This approach amplifies subtle movements in the video data, potentially unveiling detailed information about hand tremors which might be missed with more straightforward approaches.

Additionally, methodologies like optical flow, used by (Wong et al., 2019) and (Liu et al., 2023), which prioritize the movement of the subject (hand tremors) and ignore irrelevant static backgrounds, offer another pathway to potentially enhance data quality. These strategies, concentrating on motion, directly target the core interest of the studies – the tremor – thereby possibly providing a more accurate representation of the condition.

In essence, despite the prevalent trend towards simpler preprocessing, there is a case to be made for the adoption of more sophisticated techniques. Enhanced preprocessing could feasibly unearth more nuanced data and, by extension, lead to more accurate and reliable machine-learning models in the diagnosis and analysis of hand tremors.

5 CONCLUSION

Exploring different ways to use machine learning to diagnose hand tremors has given us a wide look at many research methods and results. We've seen a wide range of approaches from Graph Convolutional Networks to Support Vector Machines being used, along with various validation approaches, all aimed at improving how accurately and reliably these tools can help during diagnosis.

However, there's a clear need for more shared datasets. Without them, it's hard to replicate studies or compare different approaches, so finding the best diagnostic model becomes a tricky task. Making datasets available for more researchers will help improve, compare, and validate models in a straightforward way.

Another potential solution to increase the amount of data available is for researchers to collect videos from studies where the data is made available upon request. By combining these videos of the same task, it is possible to create a larger, more diverse dataset. This approach could provide a valuable baseline for future research. However, it's crucial to first exam if publishing such a combined dataset is feasible, considering the various privacy concerns and consent agreements associated with the original sources. Compliance with privacy regulations and ethical standards is of utmost importance when dealing with medical data. However, this strategy could significantly contribute to advancing the field, if managed correctly, allowing for more comprehensive and comparative studies in hand tremor diagnosis.

Looking ahead, future research should also consider other forms of simpler and cheaper tests that do not rely on video images, like analyzing drawings and handwriting from participants. Combining advanced models with simple, low-cost tests could make them useful diagnostic tools, including in settings where resources are limited. So, balancing technological advancements with practical diagnostic methods will be key.

In summary, it's important for future studies to work together, to share and build upon available datasets, and highlight the benefits of combining accessible, yet high-performing technology with easy and low-cost tests, to help develop useful tools to better aid both practitioners and patients.

REFERENCES

Abboud, H., Ahmed, A., and Fernandez, H. H. (2011). Essential tremor: Choosing the right management plan for your patient. *Cleveland Clinic Journal of Medicine*, 78(12):821–828.

Ali, M. R., Hernandez, J., Dorsey, E. R., Hoque, E., and McDuff, D. (2020). 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.

Armstrong, M. J. and Okun, M. S. (2020). Diagnosis and treatment of parkinson disease. *JAMA*, 323(6):548.

Baumann, C. R. (2012). Epidemiology, diagnosis and differential diagnosis in parkinson's disease tremor. *Parkinsonism & Related Disorders*, 18:S90–S92.

Beck, C. A., Beran, D. B., Biglan, K. M., Boyd, C. M., Dorsey, E. R., Schmidt, P. N., Simone, R., Willis, A. W., Galifianakis, N. B., Katz, M., Tanner, C. M., Dodenhoff, K., Aldred, J., Carter, J., Fraser, A., Jimenez-Shahed, J., Hunter, C., Spindler, M., Reichwein, S., Mari, Z., Dunlop, B., Morgan, J. C., McLane, D., Hickey, P., Gauger, L., Richard, I. H., Mejia, N. I., Bwala, G., Nance, M., Shih, L. C., Singer, C., Vargas-Parra, S., Zadikoff, C., Okon, N., Feigin, A., Ayan, J., Vaughan, C., Pahwa, R., Dhall, R., Hassan, A., DeMello, S., Riggare, S. S., Wicks, P., Achey, M. A., Elson, M. J., Goldenthal, S., Keenan, H. T., Korn, R., Schwarz, H., Sharma, S., Stevenson, E. A., and Zhu, W. (2017). National randomized controlled trial of virtual house calls for parkinson disease. *Neurology*, 89(11):1152–1161.

Chang, C.-M., Huang, Y.-L., Chen, J.-C., and Lee, C.-C. (2019). 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.

Chen, Y., Ma, H., Wang, J., Wu, J., Wu, X., and Xie, X. (2021). 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.

Elble, R. J. (2013). What is essential tremor? *Current Neurology and Neuroscience Reports*, 13(6).

Guo, R., Li, H., Zhang, C., and Qian, X. (2022). A tree-structure-guided graph convolutional network with contrastive learning for the assessment of parkinsonian hand movements. *Medical Image Analysis*, 81:102560.

Jankovic, J. (1980). Physiologic and pathologic tremors. *Annals of Internal Medicine*, 93(3):460.

Li, H., Shao, X., Zhang, C., and Qian, X. (2021). Automated assessment of parkinsonian finger-tapping tests through a vision-based fine-grained classification model. *Neurocomputing*, 441:260–271.

Li, Z., Lu, K., Cai, M., Liu, X., Wang, Y., and Yang, J. (2022). An automatic evaluation method for parkinson's dyskinesia using finger tapping video for small samples. *Journal of Medical and Biological Engineering*, 42(3):351–363.

Lin, B., Luo, W., Luo, Z., Wang, B., Deng, S., Yin, J., and Zhou, M. (2020). Bradykinesia recognition in parkinson's disease via single RGB video. *ACM Transactions on Knowledge Discovery from Data*, 14(2):1–19.

Liu, W., Lin, X., Chen, X., Wang, Q., Wang, X., Yang, B., Cai, N., Chen, R., Chen, G., and Lin, Y. (2023). Vision-based estimation of MDS-UPDRS scores for quantifying parkinson's disease tremor severity. *Medical Image Analysis*, 85:102754.

Liu, Y., Chen, J., Hu, C., Ma, Y., Ge, D., Miao, S., Xue, Y., and Li, L. (2019). Vision-based method for automatic quantification of parkinsonian bradykinesia. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 27(10):1952–1961.

Locatelli, P., Alimonti, D., Traversi, G., and Re, V. (2020). Classification of essential tremor and parkinson's tremor based on a low-power wearable device. *Electronics*, 9(10):1695.

Louis, E. D. and Ferreira, J. J. (2010). How common is the most common adult movement disorder? update on the worldwide prevalence of essential tremor. *Movement Disorders*, 25(5):534–541.

Lu, M., Zhao, Q., Poston, K. L., Sullivan, E. V., Pfefferbaum, A., Shahid, M., Katz, M., Kouhsari, L. M., Schulman, K., Milstein, A., Niebles, J. C., Henderson, V. W., Fei-Fei, L., Pohl, K. M., and Adeli, E. (2021). Quantifying parkinson's disease motor severity under uncertainty using MDS-UPDRS videos. *Medical Image Analysis*, 73:102179.

Monje, M. H. G., Domínguez, S., Vera-Olmos, J., Antonini, A., Mestre, T. A., Malpica, N., and Sánchez-Ferro, Á. (2021). Remote evaluation of parkinson's disease using a conventional webcam and artificial intelligence. *Frontiers in Neurology*, 12.

Parkinson, J. (2002). An essay on the shaking palsy. *The Journal of Neuropsychiatry and Clinical Neurosciences*, 14(2):223–236.

Rana, A. Q. and Chou, K. L. (2015). *Essential Tremor in Clinical Practice*. Springer International Publishing.

Shen, J., Zhang, C. J. P., Jiang, B., Chen, J., Song, J., Liu, Z., He, Z., Wong, S. Y., Fang, P.-H., and Ming, W.-K. (2019). Artificial intelligence versus clinicians in disease diagnosis: Systematic review. *JMIR Medical Informatics*, 7(3):e10010.

Vignoud, G., Desjardins, C., Salardaine, Q., Mongin, M., Garcin, B., Venance, L., and Degos, B. (2022). Video-based automated assessment of movement parameters consistent with MDS-UPDRS III in parkinson's disease. *Journal of Parkinson's Disease*, 12(7):2211–2222.

Wang, X., Garg, S., Tran, S. N., Bai, Q., and Alty, J. (2021). Hand tremor detection in videos with cluttered background using neural network based approaches. *Health Information Science and Systems*, 9(1).

Wong, D. C., Relton, S. D., Fang, H., Qhawaji, R., Graham, C. D., Alty, J., and Williams, S. (2019). 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.

Yang, N., Liu, D.-F., Liu, T., Han, T., Zhang, P., Xu, X., Lou, S., Liu, H.-G., Yang, A.-C., Dong, C., Vai, M. I., Pun, S. H., and Zhang, J.-G. (2022). Automatic detection pipeline for accessing the motor severity of parkinson's disease in finger tapping and postural stability. *IEEE Access*, 10:66961–66973.

Zhang, H., Ho, E. S. L., Zhang, X., and Shum, H. P. H. (2022). Pose-based tremor classification for parkinson's disease diagnosis from video. In *Lecture Notes in Computer Science*, pages 489–499. Springer Nature Switzerland.

Zhao, A. and Li, J. (2022). Two-channel lstm for severity rating of parkinson's disease using 3d trajectory of hand motion. *Multimedia Tools and Applications*, 81(23):33851–33866.