Deep Learning for hand tracking in Parkinson's Disease video-based assessment: Current and future perspectives

Artificial Intelligence In Medicine 154 (2024) 102914

Contents lists available at ScienceDirect

Artificial Intelligence In Medicine

journal homepage: www.elsevier.com/locate/artmed

Deep Learning for hand tracking in Parkinson's Disease video-based assessment: Current and future perspectives

Gianluca Amprimo [a,b,*], Giulia Masi [a], Gabriella Olmo [a], Claudia Ferraris [b]

[a] Politecnico di Torino - Control and Computer Engineering Department, Corso Duca degli Abruzzi, 24, Turin, 10129, Italy
[b] National Research Council - Institute of Electronics, Information Engineering and Telecommunications, Corso Duca degli Abruzzi, 24, Turin, 10029, Italy

ARTICLE INFO

Keywords:
Hand tracking
Deep learning
Parkinson's disease
MDS-UPDRS
Computer vision
Bradykinesia
Tremor

ABSTRACT

Background: Parkinson's Disease (PD) demands early diagnosis and frequent assessment of symptoms. In particular, analysing hand movements is pivotal to understand disease progression. Advancements in hand tracking using Deep Learning (DL) allow for the automatic and objective disease evaluation from video recordings of standardised motor tasks, which are the foundation of neurological examinations. In view of this scenario, this narrative review aims to describe the state of the art and the future perspective of DL frameworks for hand tracking in video-based PD assessment.

Methods: A rigorous search of PubMed, Web of Science, IEEE Explorer, and Scopus until October 2023 using primary keywords such as *parkinson, hand tracking, and deep learning* was performed to select eligible by focusing on video-based PD assessment through DL-driven hand tracking frameworks

Results: After accurate screening, 23 publications met the selection criteria. These studies used various solutions, from well-established pose estimation frameworks, like OpenPose and MediaPipe, to custom deep architectures designed to accurately track hand and finger movements and extract relevant disease features. Estimated hand tracking data were then used to differentiate PD patients from healthy individuals, characterise symptoms such as tremors and bradykinesia, or regress the Movement Disorder Society-Unified Parkinson's Disease Rating Scale (MDS-UPDRS) by automatically assessing clinical tasks such as finger tapping, hand movements, and pronation-supination.

Conclusions: DL-driven hand tracking holds promise for PD assessment, offering precise, objective measurements for early diagnosis and monitoring, especially in a telemedicine scenario. However, to ensure clinical acceptance, standardisation and validation are crucial. Future research should prioritise large open datasets, rigorous validation on patients, and the investigation of new frontiers such as tracking hand-hand and hand-object interactions for daily-life tasks assessment.

Contents

Abbreviations: 2D, Two-dimensional; 3D, Three-dimensional; AR, Augmented Reality; CNN, Convolutional Neural Network; CV, Computer Vision; DBS, Deep Brain Stimulation; DL, Deep Learning; DSNT, Differentiable Spatial-to-Numerical Transform; DUC, Dense Upsampling Convolution; EMG, Electromiography; FPN, Feature Pyramid Network; GMH, Google MediaPipe Hand; HAR, Human Action Recognition; HPE, Human Pose Estimation; HLDM, Hand Landmarks Detection Module; IMU, Inertial Measurement Unit; MAK, Microsoft Azure Kinect; MBRS, Modified Bradykinesia Rating Scale; MDS-UPDRS, Movement Disorder Society- Unified Parkinson's disease Rating Scale; ML, Machine Learning; PAFs, Part Affinity Fields; PD, Parkinson's Disease; PDM, Palm Detection Module; TOF, Time of Flight; VR, Virtual Reality

* Corresponding author at: Politecnico di Torino - Control and Computer Engineering Department, Corso Duca degli Abruzzi, 24, Turin, 10129, Italy.
E-mail address: gianluca.amprimo@polito.it (G. Amprimo).
URLs: https://www.polito.it/personale?p=gianluca.amprimo (G. Amprimo), https://www.researchgate.net/profile/Giulia-Masi-2 (G. Masi), https://www.sysbio.polito.it/analytics-technologies-health/ (G. Olmo), https://www.ieiit.cnr.it/people/Ferraris-Claudia (C. Ferraris).

https://doi.org/10.1016/j.artmed.2024.102914
Received 31 October 2023; Received in revised form 19 May 2024; Accepted 21 May 2024
Available online 17 June 2024

1. Introduction

Parkinson's Disease (PD) is one of the most investigated neurodegenerative diseases in the literature, with more than 4404 scientific papers published in the last 20 years (*source*: PubMed, *keyword in title*: 'Parkinson'). With its etiology being unclear [1], and the current lack of a definitive medical treatment [2], various research works have focused both on identifying early markers of the disease [3,4], and novel methodologies for assessment and monitoring [5–7]. Indeed, PD consists in the progressive degeneration of the neural circuits devoted to motor control [8], leading to an irreversible escalation of symptoms such as tremor, bradykinesia, and akinesia [2]. These may arise in-between the infrequent outpatient visits, thus leading to a delayed identification of the disease aggravation [9]. Moreover, the current gold standard for the assessment of PD severity relies on subjective neurological examinations performed according to the standardised motor tasks coded in the Movement Disorder Society- Unified Parkinson's disease Rating Scale (MDS-UPDRS) [10]. Hence, the need for an objective assessment is a widely investigated topic in the literature, with numerous solutions encompassing wearable [7,11–20] and non-wearable [21–34] technologies to estimate the measurable features of motion. Wearable set-ups include Inertial Measurement Unit (IMU) [13,14,20], surface electromiography (EMG) sensors [15,18,19], and smart-textiles [16,17], embedding several types of integrated sensing devices. Mixed approaches also exist, leveraging the combination of computer vision and passive or active wearable devices [35–38]. Data collected through these devices are then analysed with data-driven models to automatically infer the motor condition of the subject [12,39]. While minimally invasive, these technologies have a limited perspective or require several sensors to reconstruct complex movements. On the other hand, non-wearable technologies span from radio sensors [23,24] to depth sensors [25,26,31,34], traditional RGB cameras [27,28,32,33], and mixed RGB-Depth cameras [29,30], either in monocular or multi-camera set-ups. Indeed, current advances in Deep Learning (DL) have paved the way for new computer vision methods for Human Pose Estimation (HPE). Microsoft Kinect, for example, showed a movement reconstruction accuracy comparable to gold standard motion capture systems [40,41]. A recent and prominent trend in DL research for computer vision focuses on deriving complex three-dimensional (3D) poses by exploiting a single RGB camera [42,43]. In the context of objective PD assessment, the reconstruction of the body pose through these methods constitutes a relevant asset, that may be employed in training automatic disease staging models, as previously done for wearable technologies [44,45].

A relevant aspect of PD progression involves hand movements: as PD progresses, the once-fluid and effortless movements of the hand and fingers, fundamental for carrying out several daily-life activities, become compromised. Hence, periodic monitoring of motor alterations involving the hands is pivotal for tailoring interventions to enhance the patient's quality of life. For instance, this would allow for more effective coping with symptoms such as bradykinesia and tremor, which mainly manifest in this body district [46,47]. However, while whole-body HPE has achieved a remarkable accuracy, the specific case of hand tracking presents with more complex challenges, still unsolved. Indeed, the human hand is an instrument of high dexterity and precision, thanks to its numerous degrees of freedom. Fingers can assume many configurations in the 3D space, and possibly rapidly shifting across them [48,49]. Complex poses and their evolution over time make hand tracking a cumbersome task, especially in real-time applications [49,50]. As for HPE, several approaches attempted hand pose estimation both through wearable set-ups (such as IMUs [51] or sensorised gloves [52]) and through traditional computer vision methods [53,54]. However, the former may imply interference with natural movements, discomfort, and system bulkiness. The latter approach suffers from several limitations, such as a drastic reduction in performance due to the self-occlusion of hand joints, computational cost, or domain-specific issues (e.g., skin colour segmentation). As for whole-body HPE, the introduction of DL frameworks for video-based hand tracking *in-the-wild* seeks to overcome these issues by allowing marker-less and occlusions-robust tracking [55–57]. However, while several DL architectures and frameworks exist in the literature, there is not a large consensus about the optimal technique for hand tracking *in-the-wild*, especially in the specific case of automatic assessment of PD.

This narrative review investigates the DL-driven hand tracking architectures and frameworks currently employed for video-based PD assessment. In further detail, it focuses on the solutions proposed since 2017, when the first DL approach for 3D hand joints regression from a single RGB video was published and became freely available [55]. Compared to other reviews broadly investigating marker-less pose estimation in medicine [21,58,59], this review favours a narrower scope to better delineate and discuss in detail the current architectures and trends for hand tracking in the clinical practice of PD. Specifically, the paper addresses the following research questions:

- What are the most popular hand tracking frameworks and architectures employed in automatic, video-based PD assessment?
- Were these architectures validated as a measurement system for objectively characterising hand and finger motion? Are the data employed in these studies openly available to the scientific community?

- Which are the most assessed symptoms, and through the tracking of which clinical tasks?

Therefore, this work aims to help future researchers select or construct innovative DL-driven frameworks for hand tracking in video-based PD assessment. Moreover, future assessment scenarios and promising research directions are also evaluated and discussed.

This review is organised as follows: Section 2 provides a background on the typical assessment tasks administered in the clinical practice of PD, objective assessment approaches through wearable sensors, and DL-driven hand tracking using video input. Section 3 reports the search method adopted to identify the works suitable for answering the main research questions of this review. Section 4 illustrates the results of the revision process on the selected articles. Finally, Section 5 provides a commentary on the identified trends and the future research directions, and Section 6 highlights the relevant, concluding remarks.

2. Background

This section provides an overview of how PD is assessed by probing hands and their functionalities during neurological examination. Moreover, for the sake of completeness, it reports a summary of established methods based on wearables. Finally, it presents a concise summary of current DL methods for video-based hand tracking. Indeed, the intersection between such DL frameworks and the assessment of PD-related hand impairment will be explored through the selected articles.

2.1. Hand impairment as a proxy for Parkinson's

During the neurological examination of PD patients, hand assessment is crucial to evaluate motor functions, thus monitoring the disease onset, severity, and progression. Part III of MDS-UPDRS [10], which is devoted to the *motor examination*, includes several components related to the assessment of the hands and the upper extremities. In particular, the following items are worth mentioning:

- **Finger Tapping** (MDS-UPDRS Part III - Section 4): This item assesses motor speed and coordination in repetitive finger tapping movements (i.e., index-to-thumb taps or consecutive thumb-to-any-finger taps). The examiner assigns a score based on the number of taps, amplitude, speed, and regularity (i.e., presence of hesitation, halts, or variations in task execution over time).
- **Hand Movements** (MDS-UPDRS Part III - Section 5): This item evaluates the amplitude and speed of repetitive opening and closing hand movements. The examiner assigns a score based on the number of repetitions, amplitude, speed, and regularity (i.e., presence of hesitation, halts, or variations in task execution over time).
- **Pronation-supination** (MDS-UPDRS Part III - Section 6): This section evaluates the amplitude and speed of repetitive turns up and down of the palm. The score is assigned based on the number of repetitions, amplitude, speed, and regularity (i.e., presence of hesitation, halts, or variations in task execution over time).
- **Postural Tremor** (MDS-UPDRS Part III - Section 15): This task evaluates the presence and severity of tremor in hands-outstretched holding by assigning a score based on tremor amplitude and frequency of tremor occurrences.
- **Kinetic Tremor** (MDS-UPDRS Part III - Section 16): This item assesses tremor during the finger-to-nose reaching action. The examiner considers the amplitude and frequency of tremor to score the task.
- **Rest Tremor** (MDS-UPDRS Part III - Section 17–18): These two items assess the tremor of limbs (including upper limbs and hands) at rest during the whole neurological administration of Part III. The examiner considers the number of occurrences, the amplitude, and the frequency of rest tremor to assess severity.

The MDS-UPDRS scores assigned to each task range from 0 (no impairment) to 4 (severe impairment). In unilateral tasks (e.g., finger tapping, hand movements), clinicians evaluate the left and the right sides independently to highlight asymmetries. The sum of all the MDS-UPDRS Part III scores guides the clinicians through treatment decisions and provides valuable insights into the patient's health condition. Alternative evaluation scales exist, such as the Modified Bradykinesia Rating Scale (MBRS) [60], that focuses specifically on highlighting the presence of bradykinesia using a 5-level scoring, or the Bain & Findley Tremor Clinical Rating Scale [61], that, instead, is designed to assess tremor, on a score from 0 to 10. However, a significant subjective bias exists in all these scoring methods, resulting in high inter- and intra-rater variability due to the lack of objective measurements in the assessment process [62]. This scenario justifies the growing interest in the research community for automatic, objective, and explainable tools for estimating clinical scores.

2.2. Objective hand assessment using wearables

Wearables represent an established approach for objective hand impairment evaluation in PD [63–67], often outperforming traditional computer vision approaches [53,54,68,69]. Most of the wearable solutions in the literature rely either on IMUs [20,63–65,70,71] to detect hand movements or surface EMG sensors [18,66,72] for muscular activity measurements. These sensors may be integrated into *smart* gloves or simple exoskeletons, alongside additional components such as pressure and force measurements systems and microcontrollers [67,73,74].

While finger tapping and hand movements, as further discussed in the following sections, are intrinsically suited to be investigated by video approaches, wearables still represent an accurate means to establish the occurrence of tremor [64,71,75,76] and to evaluate complex 3D movements such as pronation-supination [65,70,77]. Indeed, especially for tremor and bradykinesia, IMUs are often employed as the gold standard to validate purely video-based approaches [78–80]. Although wearables have proved accurate for these tasks, the advent of DL-based hand tracking has reinvigorated the interest in developing similarly accurate contactless approaches, exploiting solely video-based inputs. Such interest mainly stems from the low-cost hardware employed and the easiness of acquiring video recordings even in large clinical studies, such as that in [81].

2.3. Video-based hand tracking and deep learning

Video-based hand tracking, sometimes known as *hand pose estimation*, is commonly formulated in the literature as the problem of deriving a pre-defined set of hand joint positions from the frames constituting an RGB, depth, or RGB-Depth video. The most popular set of joints, used in frameworks such as OpenPose [82] and MediaPipe [83], is the COCO Hand model [84], which consists of 21 virtual points modelling the main skeletal joints of the hand, as reported in Fig. 1. This schematic representation is intrinsically a non-fully-connected, undirected graph, where the joints and their connections represent, respectively, the nodes and the sides of the graph. A representation based only on joint angles (i.e., roll, pitch, and yaw angles) exists, but is less frequently considered [50]. Recently, much attention has grown around the estimation of hand poses in terms of meshes or geometric primitives, due to its possible applications in Virtual Reality (VR) and Augmented Reality (AR) scenarios [50,85–87]. However, these types of representation are less straightforward and interpretable for clinical examinations, thus neglected in applications for PD assessment. Moreover, mesh reconstruction is still particularly challenging and often requires high computing performance for its calculation.

Hand tracking requires first solving a hand localisation problem. Several methods employ a *top-down* approach and start by identifying the bounding boxes containing the hands on the video scene, then

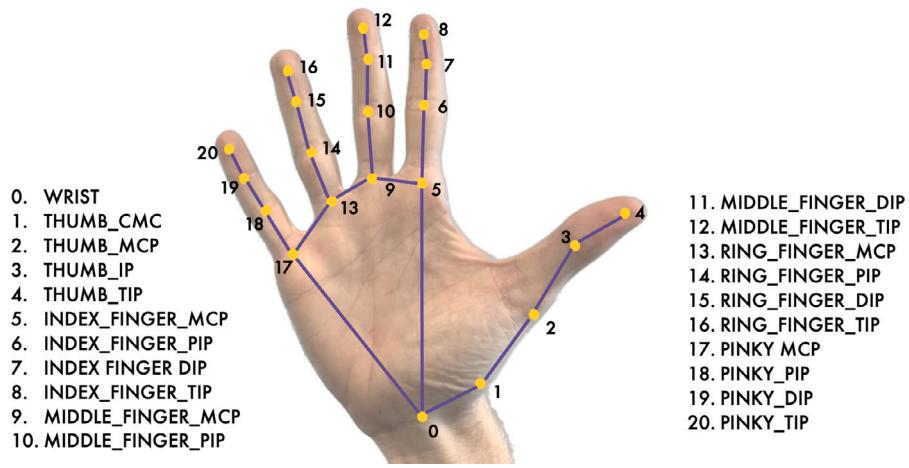

Fig. 1. The 21-joints COCO hand skeletal model, the typical joint configuration in frameworks such as OpenPose and MediaPipe. The evolution over time of joint positions and of their relative distances are used to assess motor symptoms in PD.

use this information to derive the associated hand graph models. Alternative solutions exploit a *bottom-up* strategy. Indeed, they directly estimate a set of heatmaps modelling the probability distribution of the position of each joint inside the video frame [50]. The model then infers the most likely connections between these joints to reconstruct the hands in the scene. The top-down approach is more straightforward and can achieve high accuracy given an accurate hand detector. Computational performance, however, may be lower when simultaneously tracking several hands. The bottom-up approach is theoretically more complex to formulate and may produce less accurate tracking results, but it can scale up to tracking several hands without a significant performance decrease [50].

These considerations apply mainly to hand tracking in video recordings from *in-the-wild* contexts, while videos for clinical assessment are often standardised, easing the hand localisation. Moreover, they usually require tracking a maximum of two hands simultaneously. Therefore, the hand detection problem is generally trivial, while most of the effort is devoted to achieving high accuracy in the hand pose reconstruction.

Focusing on skeletal representation, two possible pose spaces can be considered. In the two-dimensional (2D) pose space, each joint K is expressed as a tuple $(x_K, y_K) \in N^2$ that corresponds to the location measured in image pixels. In the 3D pose space, each joint K is associated with a triplet $(x_K, y_K, z_K) \in R^3$ representing coordinates in metres with respect to a reference system which can be centred in the tracked hand itself or the centre of the recording camera. A less popular option derives 2.5D poses, in which the joint K corresponds to a triplet (x_K, y_K, z_K) with $(x_K, y_K) \in N^2$ that is the 2D position in pixels, plus $z_K \in R$ which represents a dimensionless parameter expressing a relative depth concept, as it happens, for instance, in MediaPipe [88].

Another possible taxonomy for DL frameworks for video-based hand tracking takes into account modality of the input video: RGB, depth map, or mixed RGB-Depth [49]. Researchers first investigated depth-based approaches to enable 3D hand tracking due to the increase in market availability of low-cost depth sensors. In the prevalent architecture, a Convolutional Neural Network (CNN) processes depth data and extracts hand tracking information [89–92], often incorporating kinematic-based rules to enhance estimation [93]. Despite their accuracy, depth-based methods bear limitations, such as high energy consumption, unfavourable form factors, limited near-distance coverage, and challenges in outdoor usage due to the interference between light and Time of Flight (TOF) technology [50].

In multi-modal methods (RGB-Depth), two approaches are possible. In the first one, the 2D hand localisation exploits the RGB stream only, while 3D hand pose estimation includes the associated depth stream [94]. In the second approach, both modalities are fused to achieve a single-shot estimation [95,96]. The mixed modality RGB-Depth is also frequently employed to enhance the training of DL models that perform inference from RGB-only data [49].

RGB-only methods have gradually gained interest in the Computer Vision (CV) research community due to the significant reduction in the complexity and the cost of the required acquisition system. Earlier solutions were limited to 2D hand landmark extraction, which is typically integrated into many state-of-the-art 2D-HPE methods, such as OpenPose [82], DeepLabCut [97], and AlphaPose [98]. However, only few approaches focus exclusively on tracking the 2D hand using specialised architectures [99,100]. Indeed, this pose space confines the analysis to simple movements that can be approximated by a planar projection. To address more complex hand motions with these methods, multiple-camera set-ups and geometric triangulation or camera calibration techniques are needed, to uplift 2D coordinates into the 3D space [101,102]. Nevertheless, this type of set-up increases the complexity and the cost of the overall acquisition system.

Many recent works on hand tracking focus on addressing the challenging task of directly estimating 3D coordinates of the skeletal model from depth cues in monocular RGB videos. Starting with the pioneering work of Zimmerman et al. [55], many architectures have been explored [103–105]. However, these works often lack detailed information about the efficiency [106,107] or claim real-time performance (i.e., processing more than 30 frames per second) without providing the source code for replicating their results [108]. Furthermore, achieving top-tier accuracy on benchmark datasets typically requires high-performance GPUs, making these solutions impractical for applications beyond research laboratories.

3. Method

3.1. Search strategy

One of the authors (GA) conducted a computerised literature search across several electronic databases, including PubMed, Web Of Science, Scopus, and IEEE Explorer, to answer the research questions highlighted in Section 1. The eligible articles were only those peer-reviewed, and published between January 2017 and October 2023. The starting date of the search interval corresponds to the release year of the first solution for 3D hand-tracking from monocular RGB videos, by Zimmerman et al. [55]. Moreover, two popular frameworks applied for automatic assessment in the clinical domain, namely OpenPose [82] and DeepLabCut [97], were openly released after that date. The query to identify eligible works was: "(HAND TRACKING OR HAND POSE ESTIMATION) AND (PARKINSON OR FINGER TAPPING

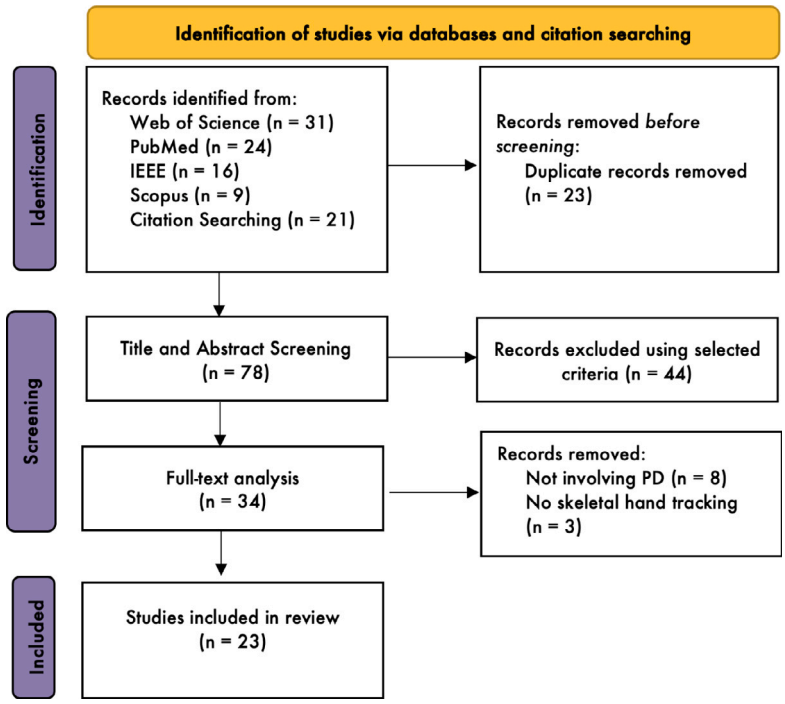

Fig. 2. PRISMA flowchart of the literature search: 101 studies were found from databases and citation searching; after duplicate removal (discarded=23), title and abstract screening (discarded=44) and full-text analysis (discarded=11), a final set of 23 research papers was obtained.

OR BRADYKINESIA OR UPDRS OR TREMOR OR PRONO SUPINATION OR AKINESIA) AND (DEEP LEARNING OR CONVOLUTIONAL NEURAL NETWORK OR NEURAL NETWORK OR ARTIFICIAL INTELLIGENCE OR MACHINE LEARNING)". In addition, a second-level reference screening from the retrieved publications highlighted other relevant studies, thus included.

3.2. Inclusion criteria

Additional, pre-defined inclusion criteria further filtered the identified works. Specifically, the included studies had to: (1) be peer-reviewed (i.e., no pre-prints); (2) apply a DL architecture for deriving the hand skeletal model from any video input modalities (i.e., RGB, RGB-Depth, depth), either by using COCO hand model or by tracking a specific subset of joints of interest; (3) propose a pipeline for assessing subjects with PD or specific symptoms of the disease from hand tracking data; and, (4) be written in English.

3.3. Exclusion criteria

The works were excluded if they: (1) focused on hand localisation, though without deriving the COCO skeletal model or a subset of its joints (e.g., fingertips) to perform assessment; (2) focused only on preliminary testing on healthy controls (HC); (3) assessed other diseases than PD; (4) exploited hand tracking for purposes different than the assessment of PD, its severity or its motor symptoms; (5) consisted in a review of other research works.

3.4. Data extraction

Relevant information retrieved during article screening included: year and type of publication; DL framework or architecture employed for hand tracking; whether the authors validated the accuracy of their DL-based hand tracking method with respect to some reference measurement system (e.g., motion capture systems, manual measurements or IMUs); type of assessment tasks tracked; the goal of the work (e.g., PD diagnosis, MDS-UPDRS score regression, specific symptoms assessment) and main findings; data availability.

4. Results

For the sake of clarity, the literature screening performed for this narrative review is summarised using the template defined by the PRISMA consortium [109], as reported in Fig. 2. A total of 101 articles were identified through database and citation searching, 34 of which were considered for full-text analysis after removing duplicates and irrelevancies, according to titles and abstracts. During full-text analysis, the subset was reduced to 23 selected articles. Indeed, eight works tested their methodology only on HC, while the other three excluded studies performed hand detection only, but not skeletal model estimation.

The automatic search of the most comprehensive and relevant scientific electronic databases produced a small subset of studies. This outcome demonstrates that DL approaches for hand tracking are highly innovative and, currently, still under-explored for applications in PD assessment. However, there is evidence of a rising trend in recent years, denoting a growing interest in the topic, as supported by the distribution of the found publications over time (Fig. 3) . In particular, 2023 was characterised by the issuance of several conference proceedings, with preliminary works to investigate the domain. Furthermore, as mentioned previously, other preliminary works, testing the feasibility of their solutions only on HC, had to be excluded [110,111] but still support the relevance of this research domain.

The pie chart displayed in Fig. 4 summarises the DL frameworks employed in the selected studies. As appreciable, the most popular approach is OpenPose (6 works), followed by DeepLabCut (5 works), MediaPipe (5 works), and MMPose (3 works). The label *Others* (6 works) includes DL architectures for which only one application was found (i.e., VitPose, HandGraphCNN) and works with a custom hand tracking architecture, specifically designed for the investigated task. For the sake of clarity, a work [112] that employed and compared 2D tracking of DeepLabCut with 2D and 3D tracking of HandGraphCNN was counted once for the *DeepLabCut* label and once for the *Others* label. The same happened for [113], which compared MediaPipe and MMPose, resulting in a total of 25 applications starting from 23 studies

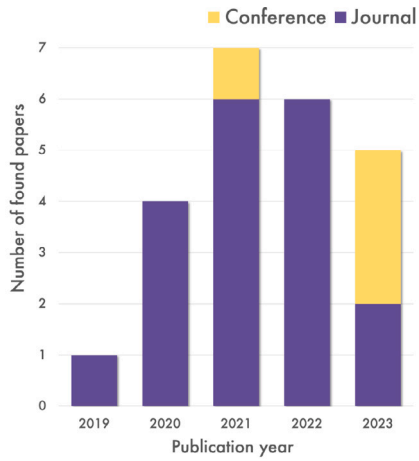

Fig. 3. Research works included in the review, divided by year and type of publication (conferences in yellow, journal papers in purple). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

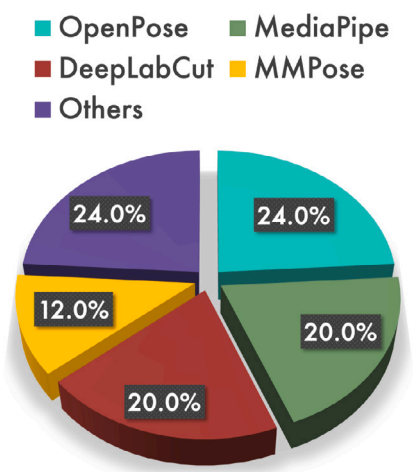

Fig. 4. DL-driven hand tracking frameworks employed for PD assessment from video, according to selected articles and applications.

analysed. A more detailed analysis of these approaches will be the focus of Section 4.1.

Regarding the input modality, only two works out of 23 employed a modality different than RGB (8.70%, 1 RGB-Depth, 1 depth). This reflects the growing focus on advancing the automatic assessment by leveraging the most accessible video format for data collection (i.e., RGB), to make this kind of applications low-cost and thus, widely deployable. This outcome is consistent with the results on pose space representation: only 8 works out of 23 (34.78%) derived 3D poses, while the remaining (65.22%) exploited 2D poses. Indeed, deriving accurate 3D poses from RGB-only is still a widely investigated and challenging task, as mentioned in Section 2. In addition, the prevalence of 2D methods likely relates with the predominant type of investigated assessment tasks and clinical goals, as further discussed in Section 4.2.

4.1. Deep learning frameworks and architectures

The 23 found works were grouped according to their hand tracking framework or architecture, following the categorisation reported in Fig. 4. Works applying, or comparing, two methods are counted and appear twice, in each corresponding group (i.e., 25 hand tracking applications are considered from 23 found studies, as previously described). In the following paragraphs each approach is described in detail; for each, the main finding is summarised in a table along with the following information: data, goals, assessment tasks, type of validation (if any), and type of tracking (*HT* column, either 2D or 3D).

4.1.1. OpenPose

OpenPose [82] is an open-source computer vision framework designed for multi-person 2D body keypoints detection, including face and hands, in RGB images and videos. The core DL architecture is based primarily on CNN modules for feature extraction, which are organised in a multi-stage architecture to enhance the accuracy of keypoint prediction. In fact, the model generates a coarse heatmap of body part locations, which gets iteratively refined in the subsequent estimation stages. The main novelty relies on the concept of Part Affinity Fields (PAFs), that is applied to establish the connections between body parts and enables tracking across frames through a bottom-up strategy. The hand tracking module of OpenPose, inspired by [102], can also be employed as a stand-alone module, provided that a hand detector is run first on the image. Indeed, when using whole-body pose estimation, the other body joints are used as a reference to localise the hands before COCO model regression. The library is optimised for real-time performance (22 fps), but only if a GPU acceleration is available (CPU only runs at 0.5 fps) [114]. Being open-source and compatible with various platforms, OpenPose is accessible to researchers and developers for diverse computer vision projects, reflecting its popularity among the identified solutions. Indeed, 6 out of the 25 applications selected in this review (24%) exploit OpenPose for hand tracking. Table 1 reports their summary.

Among these works, only *Pang-20* [115] exploits a multiple camera set-up (2 RGB cameras) to uplift coordinates from 2D to 3D. In particular, this study aims at developing a system able to characterise two cardinal symptoms in PD: bradykinesia and tremor at rest. Moreover, even though a rigorous validation with respect to gold standard systems is not provided, the architecture of OpenPose employed in this work is fine-tuned on a portion of the considered clinical assessment videos, improving the accuracy of the tracking during all the four investigated tasks (i.e., finger tapping, hand movements, pronation-supination, and rest tremor). All the remaining works exploit the native 2D pose representation of OpenPose, four to regress MDS-UPDRS scores using Machine Learning (ML) or DL [81,116–118] and one [78] to distinguish PD and HC by employing a set of derived features and statistical testing. The latter, *Monje-21* [78], is also the only work that validates the extracted metrics with respect to a standard approach, i.e., an IMU device, finding a significant correlation with its measures. On the contrary, the applications for MDS-UPDRS regression do not validate OpenPose tracking or derived kinetic features, but only the coherence between the regressed clinical value and the scores assigned by human evaluators, finding overall good agreement.

4.1.2. DeepLabCut

DeepLabCut [97] is a versatile and open-source toolbox designed for markerless 2D pose estimation in images and videos, for tracking both animals and humans. Its main advantage relies on the user-defined marker configuration of the body parts of interest, making it a valuable asset for researchers in behavioural neuroscience, biomechanics, and related fields. Exploiting Transfer Learning [119], popular pre-trained CNN models, such as MobileNet [120], ResNet [121], EfficientNet [122], can be fine-tuned in the desired pose estimation task using a subset of frames from the video recordings to analyse, that have to be manually annotated by researchers. For instance, in the case of hand tracking, the whole COCO skeletal model or only a subpart of it (e.g., index and thumb fingertips for finger tapping) can be annotated and studied. The framework requires just a limited subset of annotated training data (50-200 frames on average) to achieve satisfactory accuracy. According to developers, the average duration of the complete fine-tuning procedure is around one hour [97]. After video annotation and fine-tuning, the tracking can run in real-time or

Table 1
Studies employing OpenPose for hand tracking. Studies marked with * in the *Study* column have open data or data available on request. Studies marked with ⋄ provide, additionally, a validation of their presented framework with respect to gold standard systems, such as accelerometers or motion capture.

Study	Data	Goal	Task	HT Type	Summary
Pang-20, [115]	216 videos (Sbj: 5 PD, 22 HC)	Bradykinesia and Tremor assessment	FT, HM, PS, TR	RGB, 3D	A system based on two RGB cameras is proposed. The system is able to characterise the two symptoms with respect to the benchmark of HC. A derived metric expressing the *Average Separability* between the two groups appears well correlated with MDS-UPDRS (0.9, *p*-value 0.0312).
Li-21, [116]	744 videos (Sbj: 154 PD)	MDS-UPDRS score regression	FT	RGB, 2D	Employing the skeletal data derived by OpenPose, a subsequent three-stream network exploiting spatio-temporal attention achieves an accuracy of 72.4% and an acceptable accuracy of 98.3% in the regression task.
Morinan-23, [81]	2312 videos (Sbj: 628 PD)	MDS-UPDRS score regression	FT, HM, PS	RGB, 2D	A large population, multi-centric study was carried out. Employing the skeletal data derived by OpenPose on collected clinical videos, a set of features is extracted to regress MDS-UPDRS scores, with a classifier achieving balanced accuracy of 45% and acceptable accuracy of 81%. A binary classification, low vs high severity ratings, results in 75% accuracy.
Lu-21, [117]	68 videos (Sbj: 34 PD)	MDS-UPDRS score regression	FT	RGB, 2D	A DL architecture modelling scorers' uncertainty in assessing gait, is tested on the FT task as well, achieving a macro-average AUC of 0.69, F1-score of 47%, precision of 47%, and balanced accuracy (average recall) of 48%.
Park-21, [118]	110 videos (Sbj: 55 PD)	MDS-UPDRS score regression	FT, TR	RGB, 2D	Features from OpenPose tracking combined with an SVM model to assess TR show a good to excellent reliability range (Cohen's K: 0.791; ICC 0.927) with respect to clinical scoring. Very good reliability range are found also for FT (Cohen's K: 0.700; ICC 0.793).
Monje-21*⋄, [78]	1146 videos (Sbj: 22 PD, 20 HC)	PD diagnosis	FT, HM, PS	RGB, 2D	A method combining webcam with OpenPose tracking is designed. Features from FT, HM, and PS correlate well with IMU validation and clinical scores. However, features combined with shallow learning achieve varying accuracy values in a 4-fold cross-validation, depending on the assessment task (from low to very good).

PD: Parkinson Disease; HC: Healthy Controls; FT: Finger Tapping, HM: Hand Movements; PS: Pronation-Supination; TR: Tremor; 2D: Two-Dimensional; 3D: Three-Dimensional; MDS-UPDRS: Motor Disorder Society-Unified Parkinson's Disease Rating Scale; AUC: Area Under the Curve; SVM: Support Vector Machine; ICC: Intra-class Correlation Coefficient; IMU: Inertial Measurement Unit.

even faster, given the GPU acceleration. On the CPU alone, real-time performance is solely available for extremely low-quality input (77x33 pixels videos) employing MobileNet [123].

DeeplabCut appears in 5 out of 25 (20%) of the reviewed applications. Its popularity stems from its high level of customisation and ease of use, since a graphical user interface simplifies the interaction with the framework. However, its accuracy highly depends on the quality of the data labelling performed by the user, which requires careful consideration.

The core information regarding the selected studies employing DeepLabCut is shown in Table 2. Two works do not describe the underlying DL architecture [112,124]. Other two works exploit a convolutional residual network as a backbone, *Baker-22* with 50 residual layers (ResNet-50) and *Nunes-21* with 152 layers (ResNet-152), the deepest configuration. These diverse choices are coherent with the size of the two datasets available (i.e., larger dataset, deeper structure trainable). Only *Shin-20* [79] employs MobileNetV2, a more lightweight inverted-residual architecture, with residual blocks operating as bottleneck layers [120]. This choice appears again coherent with the size of the dataset available for the study.

All methods exploit simple 2D tracking from RGB videos, even though DeepLabCut can also perform 3D estimation, provided that the user inputs two calibrated video streams. *Vignoud-22* [112] also compares its performance to an alternative, more complex architecture for 3D tracking (i.e., HandGraphCNN). Regarding the aim of methods employing DeepLabCut, *Williams-20* [124] focuses on bradykinesia assessment, whereas the others are mainly devoted to MDS-UPDRS regression [79,112,125,126], with *Nunes-21* [125] also performing PD detection with respect to subjects with Ataxia and HC. As concerns the validation approach, only *Shin-20* [79] performs validation of the estimated parameters from the finger tapping task with respect to a reference system (i.e., an accelerometer).

4.1.3. MediaPipe

MediaPipe [83] is the solution for lightweight and portable ML pipelines, developed by Google LLC. It contains a module, defined Google MediaPipe Hand (GMH) [88], which relies on a DL approach based on monocular RGB input for 2.5D or 3D hand tracking. The latter is obtained by fitting on the 2D estimations the GHUM mesh model [127]. This mesh model was used to generate the synthetic images employed together with real data, to train GMH (around 30k single images). The GMH framework is composed of two sub-modules: a Palm Detection Module (PDM), which performs hand localisation, and a Hand Landmarks Detection Module (HLDM). First, the PDM identifies the region of interest corresponding to the hand. Then, the HLDM detects the 21 key points of the COCO skeletal model. PDM employs an encoder–decoder convolutional structure similar to Feature Pyramid Network (FPN) [128], whereas HLDM is a *regression module*, whose internal architecture has not been disclosed in details. This

Table 2
Studies using DeepLabCut for hand tracking. Studies marked with * in *Study* column have open data or data available on request. Studies marked with ⋄ provide, additionally, a validation with respect to gold standard systems such as accelerometers or motion capture.

Study	Data	Goal	Task	HT type	Architecture	Summary
Williams-20, [124]	133 videos (Sbj: 39 PD, 30 HC)	Bradykinesia assessment	FT	RGB, 2D	–	Measures of bradykinesia from DeepLabCut tracking correlate well with the clinical ratings of bradykinesia (Spearman coefficients): −0.74 speed, 0.66 amplitude, −0.65 rhythm for MBRS; −0.56 speed, 0.61 amplitude, −0.50 rhythm for MDS-UPDRS; −0.69 combined for MDS-UPDRS, with all p-values < .001.
Shin-20⋄, [79]	54 videos (Sbj: 29 PD)	MDS-UPDRS score regression	FT	RGB, 2D	MobileNetV2	FT tracking of DeepLabCut is validated with an accelerometer. Moreover, several parameters (e.g., amplitude and inter-peak interval) appear correlated ($\|R\|$ ranging between 0.34 to 0.66 for different parameters) with the clinical scores.
Nunes-21*, [125]	305 videos (Sbj: 78 PD, 169 Ataxia, 58 HC)	MDS-UPDRS score regression, PD diagnosis	FT	RGB, 2D	ResNet-152	Employing the tracking data derived by DeepLabCut, a set of features is extracted to distinguish subjects with Ataxia, PD and HC and to regress MDS-UPDRS scores. The first task for PD achieves variable AUC, (PD vs HC: 0.68, PD vs ataxia: 0.91). In the regression task, low scores are obtained (R=0.21, $R^2 = 0.04$).
Baker-22, [126]	68 videos (Sbj: 5 PD)	HM assessment during DBS surgery	HM	RGB, 2D	ResNet-50	Using DeepLabCut, an automatic recognition of arm chain pulls and hand clenches is performed during DBS surgery. The derived features of motion, then input to the SVM, reach respectively 92.30% and 76.20% accuracy in the detection task.
Vignoud-22, [112]	272 videos (Sbj: 36 PD, 11 HC)	MDS-UPDRS score regression	FT,HM,PS	RGB, 2D	–	The work compares DeepLabCut 2D tracking with HandGraphCNN 2D and 3D tracking for estimating parameters relevant to MDS-UPDRS regression. A maximum $R^2 = 0.701$ is reached using a decision tree regressor with features derived from DeepLabCut 2D tracking of HM task.

PD: Parkinson Disease; HC: Healthy Controls; FT: Finger Tapping, HM: Hand Movements; PS: Pronation-Supination; MDS-UPDRS: Motor Disorder Society-Unified Parkinson's Disease Rating Scale; 2D: Two-Dimensional; 3D: Three-Dimensional; MBRS: Modified Bradykinesia Rating Scale; AUC: Area Under the Curve; SVM: Support Vector Machine; DBS: Deep Brain Stimulation.

framework represents an interesting approach, balancing accuracy and time efficiency. Indeed, it supports a frame rate over 50 fps on a Google Pixel 6 phone using CPU only, or even faster (> 80 fps) exploiting GPU acceleration, as reported by the official web page of the pipeline [129].

Moreover, enhancements of the basic framework were proposed in the literature, such as GMH-D [130], which exploits an RGB-Depth camera (i.e., Microsoft Azure Kinect (MAK)) as input. According to the authors, GMH-D has comparable computational performance to GMH but enhanced 3D tracking accuracy, by leveraging both the depth estimation performed by the DL network and the depth map provided by the RGB-Depth camera.

MediaPipe has seen an increase in popularity over time. Indeed, this review identified 5 applications exploiting this framework, out of 25 (20%): one study in 2021 [131], two studies in 2022 [132,133], and two studies in 2023 [113,134]. The time efficiency and the reduced computational complexity provided by this solution could be the reason for this finding. The works employing this framework are summarised in Table 3. Among them, two exploit MediaPipe for 3D tracking: *Li-22* [132] using native 3D coordinates on RGB input, *Amprimo-23* using GMH-D [134], therefore, leveraging an RGB-Depth modality. All the remaining applications limit their analysis to 2D estimation [113,131,133].

MediaPipe is transversely employed for different purposes, covering MDS-UPDRS regression [132] as well as PD recognition [134], and the assessment of tremor at rest [131,133]. Besides, it appears to be the most validated approach: *Li-22* [132] validates its measures using an accelerometer, whereas two works [113,134] perform manual validation on videos (*Amprimo-23* [134] reported validation from a previous work on HC).

4.1.4. MMPose

MMPose [135] is an open-source pose estimation toolkit developed in PyTorch as a part of the OpenMMLab Project [136]. This comprehensive framework encompasses an array of advanced algorithms tailored for different applications, including 2D and 3D multi-person HPE, hand tracking, face landmark detection, fashion landmark detection, and animal pose estimation. MMPose includes popular DL architectures such as HRNet [137], MobileNet and DeepPose [138], and several techniques to improve pose estimation results such as DarkPose [139] and Residual Log-Likelihood Estimation [140]. All models require GPU acceleration for real or almost real-time performance. In the reviewed works, 3 out of 25 applications (12%) exploit MMPose by combining different architectures. The selected works are concisely described in Table 4. Among them, *Yang-22* [141] and *Xie-23* [142] exploit architectures for 3D tracking on RGB videos, while one leverages only 2D on the same input modality. Regarding the underlying architectures, *Yang-22* [141] exploits DeepPose [138], a simple convolutional and fully connected deep network. *Xie-23* [142], instead, employs a combination of HRNetV2, a deep convolutional network designed to maintain high-resolution representation throughout the architecture [137], and DarkPose [139], a model-agnostic plugin, to improve pose estimation. Finally, *Trebbau-23* [113], as already mentioned in Section 4.1.3, compares MediaPipe to several architectures from MMPose, including HRNet, MobileNet, and ResNet. This work also validates tracking results with respect to manual evaluation from videos. MMPose is used both for MDS-UPDRS score regression [141,142] and bradykinesia assessment [113].

4.1.5. Others

Works classified as *Others* are described in Table 5, specifying their architectural details. Three of them [112,143,144] exploit three

Table 3
Selected studies employing MediaPipe. Studies marked with * in *Study* column have open data or data available on request. Studies marked with ⋄ provide, additionally, a validation with respect to gold standard systems (accelerometers, motion capture systems, manual evaluation).

Study	Data	Goal	Task	HT type	Summary
Li-22*, [132]	252 videos (Sbj: 93 PD, 30 HC)	MDS-UPDRS score regression	FT	RGB, 3D	The evolution over time of the distance between index finger and thumb tips in the FT task is obtained from MediaPipe tracking. This, along with its first and second derivatives are used to train a CNN model to regress MDS-UPDRS. Overall 79.2% accuracy in 5-fold cross-validation is achieved.
Güney-22⋄, [133]	11 videos (Sbj: 11 PD)	Tremor Assessment	TR	RGB, 2D	The performance of the video-tracking is in good agreement with the accelerometer-based tracking, resulting in a tremor frequency estimation with a small error rate (MAE: 0.229 ±0.174 Hz) and a high correlation between amplitude of movements detected. Moreover, a reduction in tremor before and after medication is found.
Amprimo-23⋄, [134]	130 videos (Sbj: 35 PD, 60 HC)	PD diagnosis	FT	RGB-Depth, 3D	The work employs GMH-D, a depth-enhanced version of MediaPipe, to characterise FT. The extracted fatures are used to train several shallow learning models. Results in a Leave-One-Subject-Out cross-validation achieve accuracy and F1-score above 95%.
Wang-21*, [131]	272 videos (Sbj: 55 PD)	Tremor Assessment	TR	RGB, 2D	Starting from the TIM-Tremor dataset, MediaPipe is used to extract features for automatic identification of PD tremor from videos using a binary classification. SVM, LSTM, and CNN-LSTM models trained on such features achieve respectively, 59%, 79%, and 80% F1-score in the task.
Trebbau-23⋄, [113]	88 videos (Sbj: 6 PD, 10 HC)	Bradykinesia Assessment	FT	RGB, 2D	A comparison between MediaPipe and MMPose is performed, considering the assessment of bradykinesia from videoconference recordings and high quality videos. MediaPipe achieves best tracking accuracy in both scenarios in terms of R^2. Moreover, good correlation is found between the extracted FT parameters and the clinical scores (ICC> 0.90).

HT: Hand Tracking; PD: Parkinson Disease; HC: Healthy Controls; MDS-UPDRS: Motor Disorder Society-Unified Parkinson's Disease Rating Scale; FT: Finger Tapping, TR: Tremor; 2D: Two-Dimensional; 3D: Three-Dimensional CNN: Convolutional Neural Network; MAE: Mean Absolute Error; SVM: Support Vector Machine; LSTM: Long–Short Term memory; ICC: Intra-class Correlation Coefficient.

Table 4
Selected studies employing MMPose. Studies marked with * in *Data* column have open data or data available on request. Studies marked with ⋄ provide, additionally, a validation with respect to gold standard systems (accelerometers, motion capture systems, manual evaluation).

Study	Data	Goal	Task	HT Type	Architecture	Summary
Yang-22*, [141]	611 videos (Sbj: - PD)	MDS-UPDRS score regression	FT	RGB, 3D	DeepPose	A dataset containing clinically-scored FT and Postural Stability videos is released. MMPose is used to track the hand and extract FT parameters. Classification using a fully-connected NN achieves an average accuracy and F1-score above 80% for both hands, separately assessed.
Xie-23, [142]	490 videos (Sbj: - PD)	MDS-UPDRS score regression	PS	RGB, 3D	HRNetv2 with DarkPose	Tracking data from MMPose of PS are fed to a multi-scale framework with two graph convolutional networks for score regression. An averaged-across-scores accuracy of 61.11%, an acceptable accuracy of 91.85%, and an F1-score of 56.31% are obtained in a 5-fold cross-validation.
Trebbau-23⋄, [113]	88 videos (Sbj: 6 PD, 10 HC)	Bradykinesia Assessment	FT	RGB, 2D	HRNet, MobileNet, ResNet	A comparison between MediaPipe and MMPose is performed, considering the assessment of bradykinesia from videoconference recordings and high quality videos. MediaPipe achieves the best tracking accuracy in both scenarios. MMPose models pre-trained on OneHand10K dataset appear to track better in both scenarios, especially with HRNet backbone network.

HT: Hand Tracking; PD: Parkinson Disease; HC: Healthy Controls; MDS-UPDRS: Motor Disorder Society-Unified Parkinson's Disease Rating Scale; FT: Finger Tapping, PS: Pronation-Supination; 2D: Two-Dimensional; 3D: Three-Dimensional; NN:Neural Network.

Table 5
Selected studies employing VitPose, HandGraphCNN or custom architectures. Studies marked with * in the *Study* column have open data or data available on request. Studies marked with ⋄ provide, additionally, a validation with respect to gold standard systems, such as accelerometers or motion capture systems.

Study	Data	Goal	Task	HT type	Architecture	Summary
Zhang-23*, [143]	917 videos (Sbj: 55 PD)	Tremor Assessment	TR	RGB, 2D	VitPose	A pipeline for PD tremor detection in the TIM-Tremor dataset is designed by combining hand tracking by VitPose and a transformer network (*SimpleHandFomer*). 93% accuracy and 92.6% F1-score are achieved.
Vignoud-22, [112]	272 videos (Sbj: 36 PD, 11 HC)	MDS-UPDRS score regression	FT,HM,PS	RGB, 2D-3D	Hand-GraphCNN	The work compares DeepLabCut 2D tracking with HandGraphCNN 2D and 3D tracking for estimating parameters relevant to MDS-UPDRS regression. PS is evaluated using HandGraphCNN 3D only; however, according to the authors, the score regression for this task fails.
Liu-19, [146]	360 videos (Sbj: 60 PD)	MDS-UPDRS score regression	FT, HM, PS	RGB, 2D	MobileNetv2 with DUC and DSNT	A combination of MobileNet, Dense Upsampling Convolution, and Differentiable Spatial-to-Numerical Transform module is used to perform hand-tracking. The parameters extracted for each task allow an average accuracy across task and MDS-UPDRS scores of 89.7% using a SVM.
Lin-20, [144]	177 videos (Sbj: 121 PD)	Bradykinesia assessment	HM	RGB, 3D	Zimmer-man3D	Exploiting the architecture from *Zimmerman et al.* for 3D tracking from RGB video, the hand movements are analysed. An encode-decoder model called PM-Net achieves a 77.78% accuracy in binary detection of bradykinesia, using a single-split testing on the extracted kinetic features.
Guo-22, [105]	112 videos (Sbj: 48 PD, 11 HC)	MDS-UPDRS regression	FT	Depth, 3D	ST-A2J	An enhanced version of the A2J architecture for hand tracking on depth images is implemented and used to evaluate FT depth videos. The extracted kinematic features from tracking together with shallow learning achieve a 81.20% mean accuracy in a 5-fold cross-validation, and 76.79% in a Leave-One-Subject-Out cross-validation.
Chen-21, [147]	894 videos (Sbj: 149 PD)	MDS-UPDRS regression	FT, HM, PS	RGB, 2D	OpenPose with SHG model	After finding the ROIs using OpenPose, a SHG model infers 21-joints. The derived features of motion, combined with shallow learning, achieve an average accuracy across-score and across-fold of 87.62%.

HT: Hand Tracking; PD: Parkinson Disease; HC: Healthy Controls; MDS-UPDRS: Motor Disorder Society-Unified Parkinson's Disease Rating Scale; FT: Finger Tapping; HM: Hand Movements ;PS: Pronation-Supination; 2D: Two-Dimensional; 3D: Three-Dimensional; CNN: Convolutional Neural Network; SHG: StackedHourglass; ROI: Region Of Interest; SVM: Support Vector Machine.

state-of-the-art, general-purpose hand tracking architectures, namely VitPose, HandGraphCNN and Zimmerman3D.

VitPose [145] is a state-of-the-art transformer model for human and hand pose estimation. The model achieves remarkable tracking accuracy by leveraging plain and non-hierarchical vision transformers as backbones to extract features for a given person instance and a lightweight decoder for pose estimation. However, to exploit the complete capabilities of VitPose, potential users necessitate robust hardware set-ups, including GPUs and extensive memory capacity, to deal with the computational demand typical of vision transformers. This aspect, together with the novelty of the method, justifies why currently only a single work [143] was found exploiting this kind of architecture. In particular, *Zhang-23* [143] adopts VitPose to estimate 2D hand poses and assess PD tremor in RGB videos. The authors do not validate its tracking accuracy, but PD tremor was identified with good accuracy and F1-score values (above 90%).

HandGraphCNN [148] is a hand tracking architecture exploiting a combination of stacked hourglass, residual and graph-convolutional layers. The model was designed mainly for hand mesh recovery, but can also derive the corresponding COCO skeletal model. The complete network was trained using both supervised and unsupervised 3D data, thanks to a weakly supervised schema during fine-tuning on real-world datasets lacking depth information. This architecture is compared with DeepLabCut in *Vignoud-22* [112] to derive both 2D and 3D poses for MDS-UPDRS regression, especially in the case of the pronation-supination movements, which is tracked using HandGraphCNN-3D only. However, the authors do not validate its tracking accuracy in this specific task with respect to any standard reference for measurements.

Zimmerman 3D [55] is the first solution that was developed for 3D hand tracking in monocular RGB videos. The architecture is composed of three sub-components: the HandSegNet module, which performs the hand localisation task; the PoseNet module, which localises the hand joints using heatmaps; and the PosePrior network, which estimates the most likely 3D structure according to the PoseNet output. The first two modules are CNNs, whereas PosePrior is based on a mix of convolutional and feed-forward layers. In [144] this architecture is used for bradykinesia assessment from RGB videos, but no validation of the hand tracking quality for this specific task is reported.

Custom models The remaining three applications (12%) developed a custom hand tracking solution to address their needs. Overall, the rationale of all these solutions consists in combining popular architectures for hand or whole-body pose estimation, generating new hybrid approaches. For instance, the model in *Liu-19* [146] first performs preliminary hand detection based on whole-body key points, exploiting the same procedure of [102]. Then, a novel architecture is introduced to derive accurate 2D hand poses from RGB videos. This architecture

combines high-quality heatmaps for joint regression obtained through MobileNetV2 with two custom modules, a Dense Upsampling Convolution (DUC), and a Differentiable Spatial-to-Numerical Transform (DSNT). The model is trained on a non-PD-specific hand tracking dataset (i.e., MPII Human Pose Dataset [149]) and then used for MDS-UPDRS estimation on a private PD dataset. No validation of the quality of the tracking alone is provided in the paper. *Guo-22* [105] employs an enhanced version of their A2J architecture [91] for hand tracking on depth video streams, after hand detection using YOLO V3 [150]. In particular, a temporal encoding module is added to the original model to incorporate temporal contextual information, as well as a non-DL-based pose refinement procedure that applies physical constraints to hand movements. The approach, not validated against any gold standard system, is first trained on a non-PD-specific hand tracking dataset (i.e., HANDS-2017 [151]), and then applied on a PD-videos dataset for MDS-UPDRS regression. Finally, *Chen-21* [147] utilises OpenPose-predicted body key points to perform hand detection and then to infer 2D hand key points exploiting a Stacked Hourglass (SHG) model [152], with each stage containing a U-Net [153] alike architecture performing multi-scale feature fusion. The optimal model is trained on a mix of hand tracking data coming from publicly available datasets for hand pose estimation (i.e., Panoptic Hand [102], FreiHand [154]) as well as a sub-portion of the PD videos to analyse. The obtained hand poses were then used to regress MDS-UPDRS scores. Also, this custom hand tracking method was not compared to any other measurement system.

4.2 Assessment tasks and goals

In this section, the perspective moves to the type of assessment tasks and clinical goals that emerged from the analysis of the selected studies. The results of these observations are summarised in Figs. 5 and 6, respectively. In the former, the bar chart highlights finger tapping as the most investigated task (19 studies) [78,79,81,105,112,113,115–118,124,125,132,134,141,146,147], followed by hand movements (8 studies) [78,81,112,115,126,144,146,147], pronation-supination (7 studies) [78,81,112,115,142,146,147], and tremor (5 studies) [115,118,131,133,155]. The first three tasks are also often studied together [78,81,112,115,146,147], as a proxy for the whole MDS-UPDRS-section III assessment (6 works out of 23). The popularity of the finger tapping task is likely related to two aspects: firstly, the evident connection between this fine-motor task and the disease severity; secondly, given the correct positioning of the recording camera, the feasibility of accurately evaluating the movement using a simpler 2D framework (14 out of 19 cases). Instead, the scarcity of applications regarding tremor is likely justified by the complexity of identifying this symptom by employing a non-contact-based solution.

Regarding the clinical goals pursued by the reviewed works, the results have been summarised in four main groups:

1. Staging (i.e., MDS-UPDRS automatic regression);
2. Diagnosis (i.e., automatic recognition of PD vs HC or other pathological conditions);
3. Specific symptoms assessment;
4. Other.

As can be seen in Fig. 6, the first group is the largest (14 works) and includes works employing several different types of pipelines, ranging from those deriving first handcrafted features from hand tracking data and then inputting these into shallow [78,79,81,105,112,118,125,146,147] or deep [132,141] models, to solutions exploiting directly (or after minimal pre-processing) tracking data as input to DL regression networks [116,117,142]. The works in the second group (3 studies) all apply handcrafted feature extraction and shallow learning to perform PD detection [78,125,134]. This application usually serves as a preliminary investigation before advancing on a deeper analysis of the pathological subjects. The third group (7 works) involves mainly the

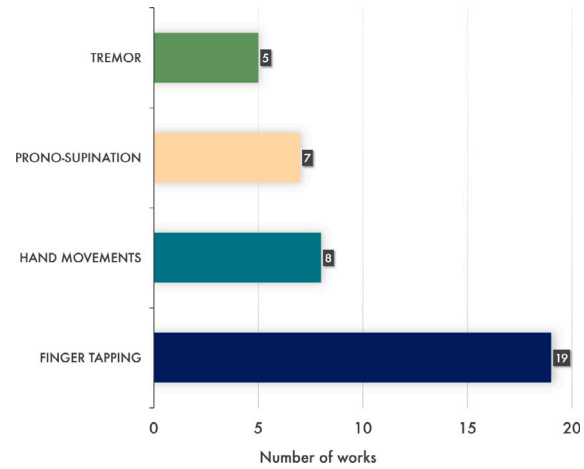

Fig. 5. Assessment tasks evaluated using DL-based hand tracking in the reviewed studies: most of the works involve the automatic characterisation of finger tapping, followed by hand movements, pronation-supination, and finally tremor.

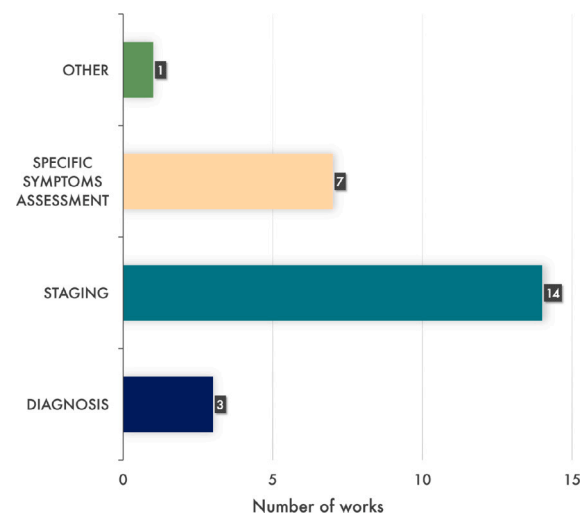

Fig. 6. Aims of the reviewed works: staging according to MDS-UPDRS score is the most common task (14), followed by the assessment of a specific symptom (i.e., bradykinesia, tremor) (7), PD diagnosis with respect to HC (3), and other goals (1).

assessment of tremor [131,133,143], bradykinesia [113,124,144] or both [115]. This goal is achieved in these studies either by investigating the correlation between relevant handcrafted features extracted from tracking data and the clinical scores [113,115,124,133], or by looking at how these features performed in the automatic detection of the symptom [131,143,144]. The group *Other* contains a single work [126] which sought to automatically assess the type of hand movements performed by PD patients during Deep Brain Stimulation (DBS) surgery.

Regarding the performances achieved by the works across all the four groups, the lack of publicly available benchmarks, except for tremor, and the lack of a systematic validation (see Section 4.3), as well as diverse strategies in reporting results (i.e., different evaluation metrics and testing or cross-validation strategies), do not allow a systematic comparison (more in Section 5). Overall, regarding diagnosis, all three identified works [78,125,134] reported very high accuracy, especially when assessing PD from finger tapping, whereas the pronation-supination task appeared more challenging [78]. When considering disease staging, instead, the finer distinction between the five levels of severity of the MDS-UPDRS appears complex, especially between adjacent scores. Indeed, many works reported very low averaged across-task and across-cross-validation-stage accuracy, but good to

the excellent value of acceptable accuracy (i.e., accuracy considering as correct the predictions within ±1 score) [81,116,147].

4.3 Validation and data availability

As regards the validation of the hand tracking methodology with respect to standard measurement systems, only 5 works out of 23 (21.73%) performed it or mentioned previous works addressing the issue [78,79,113,133,134]. Among these, three studies conducted the validation using an accelerometer [78,79,133], and two employed manual evaluation from video analysis [113,134].

With respect to data availability, just 6 works out of 23 (26%) were based on open data or reported a *data available on request* statement [78,125,131,132,141,143]. Out of these, two studies on tremor [131,143] exploited the publicly available TIM-TREMOR dataset [80], while *Yang-22* [141] proposed a new open database containing RGB videos of both the finger tapping and the whole-body postural stability tasks.

5 Discussion

The selected works provide a comprehensive analysis of the state of the art in DL-driven hand tracking frameworks and architectures applied to video-based assessment of PD. In particular, they allow answering the research questions in Section 1 and gathering insight into potential further investigations in the domain, as summarised in the following subsections.

5.1 The current perspective and its limitations

The inspection of the most popular architectures and methods reveals an evident unbalance towards easily deployable models. Indeed, most studies employ off the shelf hand tracking frameworks (i.e., OpenPose, DeepLabCut, MediaPipe), while focusing their efforts on the automatic assessment stage of their pipeline. Specifically, only three works proposed novel or custom-made architectures to solve the hand tracking problem and more complex frameworks such as HandGraphNet, Vitpose, and MMPose were less considered.

As concerns the computational complexity, most studies still heavily rely on GPU acceleration, whereas only those employing MediaPipe provided real-time computing through the use of the CPU alone. While the burden of computational complexity generally tends to be overlooked by the necessity of higher accuracy, in the perspective of integrating these solutions in real-world telemedicine scenarios for PD assessment, researchers will likely shift towards more computationally efficient approaches. Indeed, MediaPipe is the second most popular method, exhibiting a stable increase over time (1 paper in 2021, 2 in 2022, and 2 in 2023). Also, DeepLabCut represents an appealing and widely investigated approach. However, the need for two calibrated cameras to retrieve 3D poses, together with the initial manual labelling and fine-tuning stage, limits its applicability outside the theoretical research field.

The predominance of 2D over 3D tracking methods and the choice of RGB-videos as the most popular input modality confirm the interest in low-cost but accurate methods. This result is reasonable, considering that 3D tracking from an RGB input still exhibits limited accuracy and that RGB-Depth or depth modalities require dedicated and generally more expensive instrumentation. However, as depth sensors are gradually becoming more pervasive (e.g., in smartphones and VR headsets), it is likely that future solutions will begin to leverage anew this modality, as in earlier research studies.

Another relevant aspect is the validation of such hand tracking frameworks. As reported in [156–158], the clinical acceptability of objective measurement tools for PD requires accuracy validation and explainability of the reported measures, to provide a trustworthy estimation of impairment. However, when observing the studies found, an evident lack of rigorous validation appears. Indeed, several works rely solely on the coherence of their prediction with ground-truth clinical scores to indirectly validate the hand tracking framework at the source of their application pipeline. However, this approach presents with two main drawbacks: first, the clinical scores themselves do not represent an objective evaluation and thus cannot be employed for assessing the quality of the tracking methodology. As remarked in [156], it is reasonable to expect that a more complex quantitative interaction exists between tracking measures and clinical scores, considering that a clinical rater combines many sources of information to assign a subjective score, including prior experience and expectations. Second, for the MDS-UPDRS regression pipelines without validation, it may be hard to determine whether the poor performance depends on a wrongly designed MDS-UPDRS scoring module or on the low-quality features obtained by an inaccurate hand tracking framework. For instance, two independent works observed that by fine-tuning the generic OpenPose architecture specifically on PD assessment data, the tracking quality was significantly improved, with inherent effects on the overall performances of their application [115,147].

Trebbau-23 [113] also reported a higher accuracy when training MMPose models on a dataset of hand movements instead of a generic whole-body pose dataset. Additional works performing validation on HC with motion capture systems were found for OpenPose [110], MediaPipe [159], and DeeplabCut [111]when further inspecting the excluded papers. However, while validation on HC represents a first step, performing the same procedure with PD subjects should be preferred. Indeed, an evident difference exists between these two populations, which may have relevant effects on tracking quality, thus reducing the applicability in real clinical scenarios.

The need for a robust validation inherently reconnects with the data availability issue. As highlighted in [160], the lack of large-population studies is among the limitations hampering the translation of most of the research outcomes in this field into deployed technologies. Especially for video-based solutions, due to the privacy constraints in sharing the patients' RGB videos, there is a lack of a unified benchmark over which the different hand tracking frameworks could be validated and compared. The largest datasets reported in the reviewed papers are not *open-access*. The only exceptions are the dataset published in *Yang-22* [141], containing only finger tapping RGB videos, and the TIM-TREMOR dataset [80], which focuses mainly on tremor. The lack of larger datasets also hinders the development of more complex and deeper automatic assessment models, which might better investigate the finer distinction between adjacent MDS-UPDRS severity levels. Moreover, most solutions still do not address the problem of the quality of the input data: indeed, only *Trebbau-23* [113] partially addressed the issue by comparing in-presence high quality recordings versus recordings collected during Zoom videoconferencing examinations.

Regarding the type of assessment, finger tapping is the most studied and the most promising task for video-based tracking. This outcome suggests that the alterations from its correct execution are quite evident to detect, even for simple 2D tracking architectures. At the same time, the lower frequency of pronation-supination and tremor video-based assessments is likely due to the complexity of evaluating these tasks using simple hand tracking methods, rather than their clinical significance. In particular, almost all the reviewed applications struggled to attain good results for the pronation-supination task [78,81,112,142]. This outcome suggests that wearable methods, such as those in [65,70,77] still represent the state of the art to track this task quantitatively.

Finally, considering the research goals of the reviewed works, the predominant target is the finer classification of impairment by regressing the MDS-UPDRS scores, which represents a challenging goal even for clinicians. Indeed, also the study of specific symptoms, such as tremor and bradykinesia, eventually aims at this outcome. Currently, the problem remains far from being solved, and even though several works claimed good-to-excellent accuracy on their custom, private datasets, the validity and generalisability of such approaches still need

to be proved, as discussed above. The outcomes of the multi-centric, large-population study in *Morinan-23* [81] support the claim that a simple but effective evaluation system based on consumer RGB cameras may be possible. However, its results in the MDS-UPDRS regression task (Table 4 in *Morinan-23* [81]) still offer wide margins for improvement.

5.2 The future perspective

The growing interest for PD video-based assessment using hand tracking promoted the advent of numerous applications and promising research directions, as shown through the identified works. However, several challenges persist.

The first, and most evident issue to overcome in future research trajectories is the need for open and large datasets, possibly encompassing several types of assessment tasks and including not only the video modality but also additional information for a rigorous validation (e.g., motion capture tracking, IMUs recordings, manually annotated measurements). Future and current hand tracking frameworks could benefit from an enlarged data availability, and enhance their accuracy, being fine-tuned on the peculiarities of Parkinsonian hand movements. Moreover, larger, open datasets could become benchmarks for a systematic comparison between hand tracking solutions for PD assessment, as commonly done in other computer vision tasks, such as object recognition. Achieving a standardised comparison among frameworks would also require a systematic investigation of the currently employed metrics (e.g., correlation, AUC, accuracy, acceptable accuracy) and ML validation methodologies (e.g., k-fold cross-validation, leave-one-subject-out cross-validation, single training/testing split). This specific aspect is left for future literature reviews.

One other main challenge is the translation of this technological approach to real applications, where the main hindrance is the current predominance of computationally-expensive methods (i.e., GPU-accelerated architectures). Indeed, from the selected studies, a need for straightforward and user-friendly approaches for hand tracking emerged. Forthcoming methodologies should strive to develop or exploit hand tracking techniques that balance complexity and accuracy. This aspect is deemed essential to facilitate the integration of such frameworks into the routine clinical examinations. Indeed, as pointed out in [158], among the facilitators for acceptance both by patients and neurologists, is the maintenance of the human factor within the solution. Most of the reviewed works were preliminary investigations, aiming at minimal clinical supervision in the prospective usage in *at-home* assessment scenarios. However, for those solutions aiming also at *in-clinic* assessment, the participation of neurologists should be attributed a more central role. For example, the frameworks could integrate user-friendly graphical interfaces to allow clinicians to perform manual correction during hand tracking in challenging scenarios

(e.g., pronation-supination) or to fine-tune models on their patient-specific data. Moreover, frameworks allowing clinicians to track only specific hand joints of interest for each clinical task could be a promising research direction. Indeed, the popularity of the DeepLabCut framework, which partially allows these options, with the drawback of an initial data labelling stage, suggests that simple interaction and flexibility in the tracking model may be significant aspects to consider along with accuracy in real-world, deployable applications.

Regarding the investigated clinical tasks, the lower number of studies focusing on tremor and the limited results on pronation-supination suggests that wearables may still represent the best tracking methodology for the quantitative assessment of complex 3D movements. Nevertheless, those methods combining DL video-based hand tracking and wearables, such as those in [161,162] could be used to improve the tracking accuracy, at least in the near future.

Moreover, all the identified studies focused on the automatic assessment of MDS-UPDRS-Section III tasks, usually involving only one hand at a time. While this outcome may be biased by the original search query, when observing the *in-the-wild* hand tracking domain, the simultaneous tracking of multiple hands interacting, and the tracking of the hand interacting with other objects remain complex tasks to solve technically [163–166]. This remark may give explanation for the yet scanty application of such methodologies in the assessment of PD. Nevertheless, the introduction of robust and easy-to-deploy *hand-hand* and *hand-object* tracking frameworks could guarantee a more comprehensive evaluation of the subject's motor impairment. For instance, the MDS-UPDRS-Section II [10] includes the assessment, based on the patient or caregiver's reports, of some issues in daily life, such as handling cutlery during meals (task 4), dressing (task 5), personal hygiene (task 6), and handwriting (task 7). Hand-object tracking could quantitatively assess impairment during real-life tasks, in contrast to the movements coded in the MDS-UPDRS Section 3. Indeed, such movements precisely probe symptoms such as bradykinesia, but fail to reflect other daily aspects of impairment. This scenario promotes encouraging research directions to provide an exhaustive picture of the actual motor conditions of patients, in ecological contexts.

6 Conclusions

This narrative review investigated the applications of Deep-Learning-driven hand tracking for the quantitative video-based assessment of Parkinson's Disease. Alterations in hand functionality due to symptoms such as bradykinesia and tremor are closely associated with the identification and staging of the disease. The automatic analysis of clinical videos involving hand tasks by marker-less tracking may be pivotal for shifting the disease diagnosis and severity staging to an objective perspective. Moreover, these approaches may be central for developing accurate and easy-to-use novel telemedicine applications.

In particular, this review focused on identifying and describing the most popular frameworks and architectures for video-based hand tracking currently employed in this domain. Validation with respect to gold standard measurement systems and the availability of the data on which the models were trained and tested were considered relevant information. The results were also discussed by highlighting and considering the type of assessment tasks and the clinical goals investigated by the examined works. The results reveal a clear preference towards user-friendly and well-established methods such as OpenPose, DeepLabCut, and MediaPipe, exploiting coded clinical tasks such as the finger tapping test, and mainly focusing on the use of hand tracking data to regress MDS-UPDRS scores. For this goal, high accuracy is reached on several assessment tasks, thus proving the efficacy of hand tracking through the examined methodologies.

Future research efforts should address the current limitations, such as the lack of open benchmarks for the systematic validation of measures generated by hand tracking frameworks in the context of Parkinson's Disease. Additionally, the creation of such benchmarks could allow for systematically comparing different assessment pipelines, to ensure their generalisability to larger cohorts of patients. Finally, new *hand-hand* and *hand-object* tracking architectures could pave the way for innovative applications, assessing the hand impairment throughout daily-life tasks, rather than during traditional clinical examination solely.

CRediT authorship contribution statement

Gianluca Amprimo: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Conceptualization. **Giulia Masi:** Writing – review & editing, Visualization, Formal analysis. **Gabriella Olmo:** Writing – review & editing, Supervision, Conceptualization. **Claudia Ferraris:** Writing – review & editing, Validation, Supervision, Resources, Methodology, Funding acquisition, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Schapira Anthony H, Jenner Peter. Etiology and pathogenesis of Parkinson's disease. Mov Disord 2011;26(6):1049–55. http://dx.doi.org/10.15586/codonpublications.parkinsonsdisease.2018.ch1.

[2] Armstrong Melissa J, Okun Michael S. Diagnosis and treatment of Parkinson disease: A review. Jama 2020;323(6):548–60. http://dx.doi.org/10.1001/jama.2019.22360.

[3] Rechichi Irene, Iadarola Antonella, Zibetti Maurizio, Cicolin Alessandro, Olmo Gabriella. Assessing rem sleep behaviour disorder: From machine learning classification to the definition of a continuous dissociation index. Int J Environ Res Public Health 2021;19(1):248. http://dx.doi.org/10.3390/ijerph19010248.

[4] Amato Federica, Borzì Luigi, Olmo Gabriella, Orozco-Arroyave Juan Rafael. An algorithm for Parkinson's disease speech classification based on isolated words analysis. Health Inform Sci Syst 2021;9:1–15. http://dx.doi.org/10.1007/s13755-021-00162-8.

[5] Giannakopoulou Konstantina-Maria, Roussaki Ioanna, Demestichas Konstantinos. Internet of Things technologies and machine learning methods for Parkinson's disease diagnosis, monitoring and management: A systematic review. Sensors 2022;22(5):1799. http://dx.doi.org/10.3390/s22051799.

[6] Lipsmeier Florian, Taylor Kirsten I, Postuma Ronald B, Volkova-Volkmar Ekaterina, Kilchenmann Timothy, Mollenhauer Brit, et al. Reliability and validity of the roche PD mobile application for remote monitoring of early Parkinson's disease. Sci Rep 2022;12(1):12081. http://dx.doi.org/10.1038/s41598-022-15874-4.

[7] Omberg Larsson, Chaibub Neto Elias, Perumal Thanneer M, Pratap Abhishek, Tediarjo Aryton, Adams Jamie, et al. Remote smartphone monitoring of Parkinson's disease and individual response to therapy. Nat Biotechnol 2022;40(4):480–7. http://dx.doi.org/10.1038/s41587-021-00974-9.

[8] Blesa Javier, Foffani Guglielmo, Dehay Benjamin, Bezard Erwan, Obeso Jose A. Motor and non-motor circuit disturbances in early Parkinson disease: which happens first? Nat Rev Neurosci 2022;23(2):115–28. http://dx.doi.org/10.1038/s41583-021-00542-9.

[9] Bloem Bastiaan R, Okun Michael S, Klein Christine. Parkinson's disease. Lancet 2021;397(10291):2284–303. http://dx.doi.org/10.1016/S0140-6736(21)00218-X.

[10] Goetz Christopher G, Tilley Barbara C, Shaftman Stephanie R, Stebbins Glenn T, Fahn Stanley, Martinez-Martin Pablo, et al. Movement disorder society-sponsored revision of the unified Parkinson's disease rating scale (MDS-UPDRS): Scale presentation and clinimetric testing results. Mov Disord: Off J Mov Disord Soc 2008;23(15):2129–70. http://dx.doi.org/10.1002/mds.22340.

[11] Demrozi Florenc, Borzì Luigi, Olmo Gabriella. Wearable sensors for supporting diagnosis, prognosis, and monitoring of neurodegenerative diseases. Electronics 2023;12(6):1269. http://dx.doi.org/10.1177/20552076231173569.

[12] Sigcha Luis, Borzì Luigi, Amato Federica, Rechichi Irene, Ramos-Romero Carlos, Cárdenas Andrés, et al. Deep learning and wearable sensors for the diagnosis and monitoring of Parkinson's disease: A systematic review. Expert Syst Appl 2023;120541. http://dx.doi.org/10.1016/j.eswa.2023.120541.

[13] Borzì Luigi, Olmo Gabriella, Artusi Carlo Alberto, Fabbri Margherita, Rizzone Mario Giorgio, Romagnolo Alberto, et al. A new index to assess turning quality and postural stability in patients with Parkinson's disease. Biomed Signal Process Control 2020;62:102059. http://dx.doi.org/10.1016/j.bspc.2020.102059.

[14] Daneault Jean-Francois, Vergara-Diaz Gloria, Parisi Federico, Admati Chen, Alfonso Christina, Bertoli Matilde, et al. Accelerometer data collected with a minimum set of wearable sensors from subjects with Parkinson's disease. Sci Data 2021;8(1):48. http://dx.doi.org/10.1038/s41597-021-00830-0.

[15] Fadhlannisa Nisrina Firyal, Basari Basari. Design of wireless electromyography (EMG) monitoring system for muscle activity detection on Parkinson disease. In: 2020 international conference on smart technology and applications. IEEE; 2020, p. 1–4. http://dx.doi.org/10.1109/ICoSTA48221.2020.1570609982.

[16] Ravichandran Vignesh, Sadhu Shehjar, Convey Daniel, Guerrier Sebastien, Chomal Shubham, Dupre Anne-Marie, et al. Itex gloves: Design and in-home evaluation of an E-textile glove system for tele-assessment of Parkinson's disease. Sensors 2023;23(6):2877. http://dx.doi.org/10.3390/s23062877.

[17] Parati Monica, Gallotta Matteo, Muletti Manuel, Pirola Annalisa, Bellafà Alice, De Maria Beatrice, et al. Validation of pressure-sensing insoles in patients with Parkinson's disease during overground walking in single and cognitive dual-task conditions. Sensors 2022;22(17):6392. http://dx.doi.org/10.3390/s22176392.

[18] Kleinholdermann Urs, Wullstein Max, Pedrosa David. Prediction of motor unified Parkinson's disease rating scale scores in patients with Parkinson's disease using surface electromyography. Clin Neurophysiol 2021;132(7):1708–13. http://dx.doi.org/10.1016/j.clinph.2021.01.031.

[19] Hill Emily J, Mangleburg C Grant, Alfradique-Dunham Isabel, Ripperger Brittany, Stillwell Amanda, Saade Hiba, et al. Quantitative mobility measures complement the MDS-UPDRS for characterization of Parkinson's disease heterogeneity. Parkinsonism Relat Disord 2021;84:105–11. http://dx.doi.org/10.1016/j.parkreldis.2021.02.006.

[20] Anbalagan Brindha, Karnam Anantha Sunitha, Kalpana R. Novel approach to prognosis Parkinson's disease with wireless technology using resting tremors. Wirel Pers Commun 2022;125(4):2985–99. http://dx.doi.org/10.1007/s11277-022-09694-y.

[21] Scott Bradley, Seyres Martin, Philp Fraser, Chadwick Edward K, Blana Dimitra. Healthcare applications of single camera markerless motion capture: A scoping review. PeerJ 2022;10:e13517. http://dx.doi.org/10.7717/peerj.13517.

[22] Mehta D, Asif U, Hao T, Bilal E, von Cavallar S, Harrer S, et al. Towards automated and marker-less Parkinson disease assessment: Predicting UPDRS scores using sit-stand videos. In: 2021 IEEE/CVF conference on computer vision and pattern recognition workshops. Los Alamitos, CA, USA: IEEE Computer Society; 2021, p. 3836–44. http://dx.doi.org/10.1109/CVPRW53098.2021.00425.

[23] Liu Yingcheng, Zhang Guo, Tarolli Christopher G, Hristov Rumen, Jensen-Roberts Stella, Waddell Emma M, et al. Monitoring gait at home with radio waves in Parkinson's disease: A marker of severity, progression, and medication response. Sci Transl Med 2022;14(663):eadc9669. http://dx.doi.org/10.1126/scitranslmed.adc9669.

[24] Kabelac Zachary, Tarolli Christopher G, Snyder Christopher, Feldman Blake, Glidden Alistair, Hsu Chen-Yu, et al. Passive monitoring at home: A pilot study in Parkinson disease. Digit Biomark 2019;3(1):22–30. http://dx.doi.org/10.1159/000498922.

[25] Belmonte-Hernández Alberto, Theodoridis Thomas, González Marta Burgos, Hernández-Peñaloza Gustavo, Solachidis Vassilios, Ramos Jennifer Jiménez, et al. A novel framework for physical therapy rehabilitation monitoring and assessment in parkinson disease patients using depth information. In: Proceedings of the 12th ACM international conference on pErvasive technologies related to assistive environments. 2019, p. 535–9. http://dx.doi.org/10.1145/3316782.3322759.

[26] Paraschiv Elena, Petrache Cristian, Bica Ovidiu, Vasilevschi Ana. Fall detection system: Continuous in-home monitoring of Parkinson's patients. In: 2022 e-health and bioengineering conference. IEEE; 2022, p. 1–4. http://dx.doi.org/10.1109/EHB55594.2022.9991493.

[27] Rojas-Arce Jorge L, Marmolejo-Saucedo José Antonio, Jimenez-Angeles Luis. A camera-based remote sensor for physical therapy in Parkinson's disease. In: International conference on computer science and health engineering. Springer; 2022, p. 203–16. http://dx.doi.org/10.1007/978-3-031-34750-4_12.

[28] Sarapata Grzegorz, Dushin Yuriy, Morinan Gareth, Ong Joshua, Budhdeo Sanjay, Kainz Bernhard, et al. Video-based activity recognition for automated motor assessment of Parkinson's disease. IEEE J Biomed Health Inf 2023. http://dx.doi.org/10.1109/JBHI.2023.3298530.

[29] Rudå Ditte, Einarsson Gudmundur, Andersen Anne Sofie Schott, Matthiassen Jannik Boll, Correll Christoph U, Winge Kristian, et al. Exploring movement impairments in patients with Parkinson's disease using the microsoft kinect sensor: A feasibility study. Front Neurol 2021;11:610614. http://dx.doi.org/10.3389/fneur.2020.610614.

[30] Ferraris Claudia, Nerino Roberto, Chimienti Antonio, Pettiti Giuseppe, Azzaro Corrado, Albani Giovanni, et al. Assessment of Parkinson's disease at-home using a natural interface based system. In: Ambient assisted living: Italian forum 2018 9. Springer; 2019, p. 417–27. http://dx.doi.org/10.1007/978-3-030-05921-7_34.

[31] Muñoz-Ospina Beatriz, Alvarez-Garcia Daniela, Clavijo-Moran Hugo Juan Camilo, Valderrama-Chaparro Jaime Andrés, García-Peña Melisa, Herrán Carlos Alfonso, et al. Machine learning classifiers to evaluate data from gait analysis with depth cameras in patients with Parkinson's disease. Front Hum Neurosci 2022;16:826376. http://dx.doi.org/10.3389/fnhum.2022.826376.

[32] Rupprechter Samuel, Morinan Gareth, Peng Yuwei, Foltynie Thomas, Sibley Krista, Weil Rimona S, et al. A clinically interpretable computer-vision based method for quantifying gait in Parkinson's disease. Sensors 2021;21(16). http://dx.doi.org/10.3390/s21165437.

[33] Khan Taha, Zeeshan Ali, Dougherty Mark. A novel method for automatic classification of Parkinson gait severity using front-view video analysis. Technol Health Care 2021;29(4):643–53. http://dx.doi.org/10.3233/THC-191960.

[34] van Kersbergen Jannis, Otte Karen, de Vries Nienke M, Bloem Bastiaan R, Röhling Hanna M, Mansow-Model Sebastian, et al. Camera-based objective measures of Parkinson's disease gait features. BMC Res Notes 2021;14(1):1–6. http://dx.doi.org/10.1186/s13104-021-05744-z.

[35] Spasojević Sofija, Ilić Tihomir V, Milanović Slađan, Potkonjak Veljko, Rodić Aleksandar, Santos-Victor José. Combined vision and wearable sensors-based system for movement analysis in rehabilitation. Methods Inform Med 2017;56(02):95–111. http://dx.doi.org/10.3414/ME16-02-0013.

[36] Bijalwan Vishwanath, Semwal Vijay Bhaskar, Mandal TK. Fusion of multi-sensor-based biomechanical gait analysis using vision and wearable sensor. IEEE Sens J 2021;21(13):14213–20. http://dx.doi.org/10.1109/JSEN.2021.3066473.

[37] Franzo' Michela, Pascucci Simona, Serrao Mariano, Marinozzi Franco, Bini Fabiano. Kinect-based wearable prototype system for ataxic patients neurorehabilitation: software update for exergaming and rehabilitation. In: 2021 IEEE international symposium on medical measurements and applications. 2021, p. 1–6. http://dx.doi.org/10.1109/MeMeA52024.2021.9478668.

[38] Albani Giovanni, Ferraris Claudia, Nerino Roberto, Chimienti Antonio, Pettiti Giuseppe, Parisi Federico, et al. An integrated multi-sensor approach for the remote monitoring of Parkinson's disease. Sensors 2019;19(21). http://dx.doi.org/10.3390/s19214764.

[39] Landolfi Annamaria, Ricciardi Carlo, Donisi Leandro, Cesarelli Giuseppe, Troisi Jacopo, Vitale Carmine, et al. Machine learning approaches in Parkinson's disease. Curr Med Chem 2021;28(32):6548–68. http://dx.doi.org/10.2174/0929867328999210111211420.

[40] Cimolin Veronica, Vismara Luca, Ferraris Claudia, Amprimo Gianluca, Pettiti Giuseppe, Lopez Roberto, et al. Computation of gait parameters in post stroke and Parkinson's disease: A comparative study using RGB-D sensors and optoelectronic systems. Sensors 2022;22(3). http://dx.doi.org/10.3390/s22030824.

[41] Bertram Johannes, Krüger Theresa, Röhling Hanna Marie, Jelusic Ante, Mansow-Model Sebastian, Schniepp Roman, Wuehr Max, et al. Accuracy and repeatability of the microsoft Azure kinect for clinical measurement of motor function. Plos one 2023;18(1):e0279697. http://dx.doi.org/10.1371/journal.pone.0279697.

[42] Zheng Ce, Wu Wenhan, Chen Chen, Yang Taojiannan, Zhu Sijie, Shen Ju, et al. Deep learning-based human pose estimation: A survey. ACM Comput Surv 2023;56(1):1–37. http://dx.doi.org/10.1145/3603618.

[43] Choi Sangbum, Choi Seokeon, Kim Changick. MobileHumanPose: Toward real-time 3D human pose estimation in mobile devices. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021, p. 2328–38. http://dx.doi.org/10.1109/CVPRW53098.2021.00265.

[44] D'Antonio Erika, Taborri Juri, Mileti Ilaria, Rossi Stefano, Patané Fabrizio. Validation of a 3D markerless system for gait analysis based on OpenPose and two RGB webcams. IEEE Sens J 2021;21(15):17064–75. http://dx.doi.org/10.1109/JSEN.2021.3081188.

[45] Amsaprabhaa M, Jane Y Nancy, Nehemiah H Khanna. A survey on spatio-temporal framework for kinematic gait analysis in RGB videos. J Vis Commun Image Represent 2021;79:103218. http://dx.doi.org/10.1016/j.jvcir.2021.103218.

[46] Zhou Yue, Jenkins Mary E, Naish Michael D, Trejos Ana Luisa. The measurement and analysis of parkinsonian hand tremor. In: 2016 IEEE-eMBS international conference on biomedical and health informatics. IEEE; 2016, p. 414–7. http://dx.doi.org/10.1109/BHI.2016.7455922.

[47] Pal Gian, Goetz Christopher G. Assessing bradykinesia in Parkinsonian disorders. Front Neurol 2013;4:54. http://dx.doi.org/10.1016/j.artmed.2020.101966.

[48] Mohr Daniel, Zachmann Gabriel. A survey of vision-based markerless hand tracking approaches. 2013, Preprint submitted to Computer Vision and Image Understanding, URL https://api.semanticscholar.org/CorpusID:216051158.

[49] Chatzis Theocharis, Stergioulas Andreas, Konstantinidis Dimitrios, Dimitropoulos Kosmas, Daras Petros. A comprehensive study on deep learning-based 3D hand pose estimation methods. Appl Sci (Basel) 2020;10(19):6850. http://dx.doi.org/10.3390/app10196850.

[50] Li Rui, Liu Zhenyu, Tan Jianrong. A survey on 3D hand pose estimation: Cameras, methods, and datasets. Pattern Recognit 2019;93:251–72. http://dx.doi.org/10.1016/j.patcog.2019.04.026.

[51] Chan Ting Kwok, Yu Ying Kin, Kam Ho Chuen, Wong Kin Hong. Robust hand gesture input using computer vision, inertial measurement unit (IMU) and flex sensors. In: 2018 IEEE international conference on mechatronics, robotics and automation. IEEE; 2018, p. 95–9. http://dx.doi.org/10.1109/ICMRA.2018.8490559.

[52] Baldi Tommaso Lisini, Scheggi Stefano, Meli Leonardo, Mohammadi Mostafa, Prattichizzo Domenico. GESTO: A glove for enhanced sensing and touching based on inertial and magnetic sensors for hand tracking and cutaneous feedback. IEEE Trans Hum-Mach Syst 2017;47(6):1066–76. http://dx.doi.org/10.1109/THMS.2017.2720667.

[53] Ferraris C, Pianu D, Chimienti A, Pettiti G, Cimolin V, Cau N, et al. Evaluation of finger tapping test accuracy using the LeapMotion and the intel RealSense sensors. In: Proceedings of the 37th international conference of the IEEE engineering in medicine and biology society. 2015, p. 25–9.

[54] Oudah Munir, Al-Naji Ali, Chahl Javaan. Hand gesture recognition based on computer vision: A review of techniques. J Imaging 2020;6(8):73. http://dx.doi.org/10.3390/jimaging6080073.

[55] Zimmermann Christian, Brox Thomas. Learning to estimate 3D hand pose from single RGB images. In: 2017 IEEE international conference on computer vision. IEEE; 2017, http://dx.doi.org/10.1109/ICCV.2017.525.

[56] Chen Xinghao, Wang Guijin, Guo Hengkai, Zhang Cairong. Pose guided structured region ensemble network for cascaded hand pose estimation. Neurocomputing 2020;395:138–49. http://dx.doi.org/10.1016/j.neucom.2018.06.097.

[57] Boukhayma A, de Bem R, Torr PS. 3D hand shape and pose from images in the wild. In: 2019 IEEE/CVF conference on computer vision and pattern recognition. Los Alamitos, CA, USA: IEEE Computer Society; 2019, p. 10835–44. http://dx.doi.org/10.1109/CVPR.2019.01110.

[58] Sibley Krista G, Girges Christine, Hoque Ehsan, Foltynie Thomas. Video-based analyses of Parkinson's disease severity: A brief review. J Parkinson's Dis 2021;11(s1):S83–93. http://dx.doi.org/10.3233/JPD-202402.

[59] Lam Winnie WT, Tang Yuk Ming, Fong Kenneth NK. A systematic review of the applications of markerless motion capture (MMC) technology for clinical measurement in rehabilitation. J NeuroEng Rehabil 2023;20(1):1–26. http://dx.doi.org/10.1186/s12984-023-01186-9.

[60] Kishore Asha, Espay Alberto, Marras Connie, Al-Khairalla Thamer, Arenovich Tamara, Asante Abena, et al. Unilateral versus bilateral tasks in early asymmetric Parkinson's disease: Differential effects on bradykinesia. Mov Disord : Off J Mov Disord Soc 2007;22:328–33. http://dx.doi.org/10.1002/mds.21238.

[61] Bain PG, Findley LJ, Atchison P, Behari M, Vidailhet M, Gresty M, et al. Assessing tremor severity. J Neurol Neurosurg Psychiat 1993;56(8):868–73. http://dx.doi.org/10.1136/jnnp.56.8.868.

[62] Post Bart, Merkus Maruschka P, de Bie Rob MA, de Haan Rob J, Speelman Johannes D. Unified Parkinson's disease rating scale motor examination: Are ratings of nurses, residents in neurology, and movement disorders specialists interchangeable? Mov Disord: Off J Mov Disord Soc 2005;20(12):1577–84. http://dx.doi.org/10.1002/mds.20640.

[63] Okuno Ryuhei, Yokoe Masaru, Akazawa Kenzo, Abe Kazuo, Sakoda Saburo. Finger taps movement acceleration measurement system for quantitative diagnosis of Parkinson's disease. In: 2006 international conference of the IEEE engineering in medicine and biology society, vol. Supplement, 2006, p. 6623–6. http://dx.doi.org/10.1109/IEMBS.2006.260904.

[64] Lukšys Donatas, Jonaitis Gintaras, Griškevičius Julius. Quantitative analysis of Parkinsonian tremor in a clinical setting using inertial measurement units. Parkinson's Dis 2018;2018. http://dx.doi.org/10.1155/2018/1683831.

[65] Garza-Rodríguez Alejandro, Sánchez-Fernández Luis Pastor, Sánchez-Pérez Luis Alejandro, Ornelas-Vences Christopher, Ehrenberg-Inzunza Mariane. Pronation and supination analysis based on biomechanical signals from Parkinson's disease patients. Artif Intell Med 2018;84:7–22. http://dx.doi.org/10.1016/j.artmed.2017.10.001.

[66] Askari Sina, Zhang Mo, Won Deborah S. An EMG-based system for continuous monitoring of clinical efficacy of Parkinson's disease treatments. In: 2010 annual international conference of the IEEE engineering in medicine and biology. 2010, p. 98–101. http://dx.doi.org/10.1109/IEMBS.2010.5626133.

[67] Niazmand K, Tonn K, Kalaras A, Fietzek UM, Mehrkens JH, Lueth TC. Quantitative evaluation of Parkinson's disease using sensor based smart glove. In: 2011 24th international symposium on computer-based medical systems. 2011, p. 1–8. http://dx.doi.org/10.1109/CBMS.2011.5999113.

[68] Butt AH, Rovini E, Dolciotti C, Bongioanni P, De Petris G, Cavallo F. Leap motion evaluation for assessment of upper limb motor skills in Parkinson's disease. In: 2017 international conference on rehabilitation robotics. 2017, p. 116–21. http://dx.doi.org/10.1109/ICORR.2017.8009232.

[69] Butt Abdul H, Rovini E, Dolciotti C, De Petris G, Bongioanni P, Carboncini MC, et al. Objective and automatic classification of Parkinson disease with leap motion controller. Biomed Eng Online 2018;17(1):1–21. http://dx.doi.org/10.1186/s12938-018-0600-7.

[70] Sánchez-Fernández Luis Pastor, Garza-Rodríguez Alejandro, Sánchez-Pérez Luis Alejandro, Martínez-Hernández Juan Manuel. A computer method for pronation-supination assessment in Parkinson's disease based on latent space representations of biomechanical indicators. Bioengineering 2023;10(5). http://dx.doi.org/10.3390/bioengineering10050588.

[71] Dai Houde, Cai Guoen, Lin Zhirong, Wang Zengwei, Ye Qinyong. Validation of inertial sensing-based wearable device for tremor and bradykinesia quantification. IEEE J Biomed Health Inf 2021;25(4):997–1005. http://dx.doi.org/10.1109/JBHI.2020.3009319.

[72] Boroojerdi Babak, Ghaffari Roozbeh, Mahadevan Nikhil, Markowitz Michael, Melton Katie, Morey Briana, et al. Clinical feasibility of a wearable, conformable sensor patch to monitor motor symptoms in Parkinson's disease. Parkinsonism Rel Disord 2019;61:70–6. http://dx.doi.org/10.1016/j.parkreldis.2018.11.024.

[73] Li Yu, Yin Junyi, Liu Shuoyan, Xue Bing, Shokoohi Cyrus, Ge Gang, et al. Learning hand kinematics for Parkinson's disease assessment using a multimodal sensor glove. Adv Sci 2023;2206982. http://dx.doi.org/10.1002/advs.202206982.

[74] Ravichandran Vignesh, Sadhu Shehjar, Convey Daniel, Guerrier Sebastien, Chomal Shubham, Dupre Anne-Marie, et al. Itex gloves: Design and in-home evaluation of an E-textile glove system for tele-assessment of Parkinson's disease. Sensors 2023;23(6). http://dx.doi.org/10.3390/s23062877.

[75] Legaria-Santiago Valeria Karina, Sánchez-Fernández Luis Pastor, Sánchez-Pérez Luis Alejandro, Garza-Rodríguez Alejandro. Computer models evaluating hand tremors in Parkinson's disease patients. Comput Biol Med 2022;140:105059. http://dx.doi.org/10.1016/j.compbiomed.2021.105059.

[76] Sánchez-Fernández Luis Pastor, Sánchez-Pérez Luis Alejandro, Concha-Gómez Paula Denisse, Shaout Adnan. Kinetic tremor analysis using wearable sensors and fuzzy inference systems in Parkinson's disease. Biomed Signal Process Control 2023;84:104748. http://dx.doi.org/10.1016/j.bspc.2023.104748.

[77] Cavallo Filippo, Moschetti Alessandra, Esposito Dario, Maremmani Carlo, Rovini Erika. Upper limb motor pre-clinical assessment in Parkinson's disease using machine learning. Parkinsonism Rel Disord 2019;63:111–6. http://dx.doi.org/10.1016/j.parkreldis.2019.02.028.

[78] Monje Mariana HG, Domínguez Sergio, Vera-Olmos Javier, Antonini Angelo, Mestre Tiago A, Malpica Norberto, et al. Remote evaluation of Parkinson's disease using a conventional webcam and artificial intelligence. Front Neurol 2021;12:742654. http://dx.doi.org/10.3389/fneur.2021.742654.

[79] Shin Jung Hwan, Ong Jed Noel, Kim Ryul, Park Sang-Min, Choi Jihyun, Kim Han-Joon, et al. Objective measurement of limb bradykinesia using a marker-less tracking algorithm with 2D-video in PD patients. Parkinsonism Relat Disord 2020;81:129–35. http://dx.doi.org/10.1016/j.parkreldis.2020.09.007.

[80] Pintea Silvia L, Zheng Jian, Li Xilin, Bank Paulina JM, van Hilten Jacobus J, van Gemert Jan C. Hand-tremor frequency estimation in videos. In: Proceedings of the European conference on computer vision (ECCV) workshops. 2018, http://dx.doi.org/10.1007/978-3-030-11024-6_14.

[81] Morinan Gareth, Dushin Yuriy, Sarapata Grzegorz, Rupprechter Samuel, Peng Yuwei, Girges Christine, et al. Computer vision quantification of whole-body Parkinsonian bradykinesia using a large multi-site population. NPJ Parkinsons Dis 2023;9(1):10. http://dx.doi.org/10.1038/s41531-023-00454-8.

[82] Cao Z, Hidalgo G, Simon T, Wei S, Sheikh Y. OpenPose: Realtime multi-person 2D pose estimation using part affinity fields. IEEE Trans Pattern Anal Mach Intell 2021;43(01):172–86. http://dx.doi.org/10.1109/TPAMI.2019.2929257.

[83] Lugaresi Camillo, Tang Jiuqiang, Nash Hadon, McClanahan Chris, Uboweja Esha, Hays Michael, et al. MediaPipe: A framework for building perception pipelines. 2019, http://dx.doi.org/10.48550/arXiv.1906.08172, arXiv preprint arXiv:1906.08172.

[84] COCO - common objects in context. 2023, http://cocodataset.org/. (Accessed: 30 Oct 2023).

[85] Kulon Dominik, Guler Riza Alp, Kokkinos Iasonas, Bronstein Michael M, Zafeiriou Stefanos. Weakly-supervised mesh-convolutional hand reconstruction in the wild. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020, p. 4990–5000. http://dx.doi.org/10.1109/CVPR42600.2020.00504.

[86] Zhang Xiong, Li Qiang, Mo Hong, Zhang Wenbo, Zheng Wen. End-to-end hand mesh recovery from a monocular rgb image. In: Proceedings of the IEEE/CVF international conference on computer vision. 2019, p. 2354–64. http://dx.doi.org/10.1109/TPAMI.2022.3222784.

[87] Moon Gyeongsik, Choi Hongsuk, Lee Kyoung Mu. Accurate 3D hand pose estimation for whole-body 3D human mesh estimation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022, p. 2308–17. http://dx.doi.org/10.1109/CVPRW56347.2022.00257.

[88] Zhang Fan, Bazarevsky Valentin, Vakunov Andrey, Tkachenka Andrei, Sung George, Chang Chuo-Ling, et al. MediaPipe hands: On-device real-time hand tracking. 2020, http://dx.doi.org/10.48550/arXiv.2006.10214, arXiv preprint.

[89] Tompson Jonathan, Stein Murphy, Lecun Yann, Perlin Ken. Real-time continuous pose recovery of human hands using convolutional networks. ACM Trans Graph 2014;33(5):1–10. http://dx.doi.org/10.1145/2629500.

[90] Otberdout Naima, Ballihi Lahoucine, Aboutajdine Driss. Hand pose estimation based on deep learning depth map for hand gesture recognition. In: 2017 intelligent systems and computer vision. IEEE; 2017, http://dx.doi.org/10.1109/ISACV.2017.8054904.

[91] Xiong Fu, Zhang Boshen, Xiao Yang, Cao Zhiguo, Yu Taidong, Zhou Joey Tianyi, et al. A2j: Anchor-to-joint regression network for 3d articulated pose estimation from a single depth image. In: Proceedings of the IEEE/CVF international conference on computer vision. 2019, p. 793–802. http://dx.doi.org/10.1109/ICCV.2019.00088.

[92] Madadi Meysam, Escalera Sergio, Baró Xavier, Gonzàlez Jordi. End-to-end global to local convolutional neural network learning for hand pose recovery in depth data. IET Comput Vis 2022;16(1):50–66. http://dx.doi.org/10.1049/cvi2.12064.

[93] Zhou Xingyi, Wan Qingfu, Zhang Wei, Xue Xiangyang, Wei Yichen. Model-based deep hand pose estimation. 2016, http://dx.doi.org/10.48550/arXiv.1606.06854, arXiv preprint.

[94] Mueller Franziska, Mehta Dushyant, Sotnychenko Oleksandr, Sridhar Srinath, Casas Dan, Theobalt Christian. Real-time hand tracking under occlusion from an egocentric RGB-D sensor. In: 2017 IEEE international conference on computer vision. IEEE; 2017, http://dx.doi.org/10.1109/ICCV.2017.131.

[95] Kazakos Evangelos, Nikou Christophoros, Kakadiaris Ioannis A. On the fusion of RGB and depth information for hand pose estimation. In: 2018 25th IEEE international conference on image processing. IEEE; 2018, http://dx.doi.org/10.1109/ICIP.2018.8451022.

[96] Sanchez-Riera Jordi, Srinivasan Kathiravan, Hua Kai-Lung, Cheng Wen-Huang, Hossain M Anwar, Alhamid Mohammed F. Robust RGB-D hand tracking using deep learning priors. IEEE Trans Circuits Syst Video Technol 2018;28(9):2289–301. http://dx.doi.org/10.1109/TCSVT.2017.2718622.

[97] Mathis Alexander, Mamidanna Pranav, Cury Kevin M, Abe Taiga, Murthy Venkatesh N, Mathis Mackenzie Weygandt, et al. DeepLabCut: Markerless pose estimation of user-defined body parts with deep learning. Nat Neurosci 2018;21(9):1281–9. http://dx.doi.org/10.1038/s41593-018-0209-y.

[98] Fang Hao-Shu, Li Jiefeng, Tang Hongyang, Xu Chao, Zhu Haoyi, Xiu Yuliang, et al. AlphaPose: Whole-body regional multi-person pose estimation and tracking in real-time. IEEE Trans Pattern Anal Mach Intell 2023;45(6):7157–73. http://dx.doi.org/10.1109/TPAMI.2022.3222784.

[99] Santavas Nicholas, Kansizoglou Ioannis, Bampis Loukas, Karakasis Evangelos, Gasteratos Antonios. Attention! a lightweight 2D hand pose estimation approach. IEEE Sens J 2021;21(10):11488–96. http://dx.doi.org/10.1109/JSEN.2020.3018172.

[100] Gouidis Filippos, Panteleris Paschalis, Oikonomidis Iason, Argyros Antonis. Accurate hand keypoint localization on mobile devices. In: 2019 16th international conference on machine vision applications. 2019, p. 1–6. http://dx.doi.org/10.23919/MVA.2019.8758059.

[101] Lim Guan Ming, Jatesiktat Prayook, Keong Kuah Christopher Wee, Tech Ang Wei. Camera-based hand tracking using a mirror-based multi-view setup. Annu Int Conf IEEE Eng Med Biol Soc 2020;2020:5789–93. http://dx.doi.org/10.1109/EMBC44109.2020.9176728.

[102] Simon Tomas, Joo Hanbyul, Matthews Iain, Sheikh Yaser. Hand keypoint detection in single images using multiview bootstrapping. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2017, p. 1145–53. http://dx.doi.org/10.1109/CVPR.2017.494.

[103] Mueller Franziska, Bernard Florian, Sotnychenko Oleksandr, Mehta Dushyant, Sridhar Srinath, Casas Dan, et al. GANerated hands for real-time 3D hand tracking from monocular RGB. In: 2018 IEEE/CVF conference on computer vision and pattern recognition. IEEE; 2018, http://dx.doi.org/10.1109/CVPR.2018.00013.

[104] He Yiming, Hu Wei. 3D hand pose estimation via regularized graph representation learning. In: Artificial intelligence. Cham: Springer International Publishing; 2021, p. 540–52. http://dx.doi.org/10.1007/978-3-030-93046-2_46.

[105] Guo Shaoxiang, Rigall Eric, Ju Yakun, Dong Junyu. 3D hand pose estimation from monocular RGB with feature interaction module. IEEE Trans Circuits Syst Video Technol 2022;32(8):5293–306. http://dx.doi.org/10.1109/TCSVT.2022.3142787.

[106] Spurr Adrian, Song Jie, Park Seonwook, Hilliges Otmar. Cross-Modal deep variational hand pose estimation. In: 2018 IEEE/CVF conference on computer vision and pattern recognition. IEEE; 2018, http://dx.doi.org/10.1109/CVPR.2018.00017.

[107] Sharma Sanjeev, Huang Shaoli, Tao Dacheng. An end-to-end framework for unconstrained monocular 3D hand pose estimation. 2019, http://dx.doi.org/10.48550/arXiv.1911.12501, arXiv preprint.

[108] Cai Yujun, Ge Liuhao, Cai Jianfei, Thalmann Nadia Magnenat, Yuan Junsong. 3D hand pose estimation using synthetic data and weakly labeled RGB images. IEEE Trans Pattern Anal Mach Intell 2021;43(11):3739–53. http://dx.doi.org/10.1109/TPAMI.2020.2993627.

[109] PRISMA contributors. PRISMA. 2023, http://prisma-statement.org. (Accessed: 30 Oct 2023).

[110] Gionfrida Letizia, Rusli Wan M R, Bharath Anil A, Kedgley Angela E. Validation of two-dimensional video-based inference of finger kinematics with pose estimation. PLoS One 2022;17(11):e0276799. http://dx.doi.org/10.1371/journal.pone.0276799.

[111] Li Renjie, St George Rebecca J, Wang Xinyi, Lawler Katherine, Hill Edward, Garg Saurabh, et al. Moving towards intelligent telemedicine: Computer vision measurement of human movement. Comput Biol Med 2022;147(105776):105776. http://dx.doi.org/10.1016/j.compbiomed.2022.105776.

[112] Vignoud Gaëtan, Desjardins Clément, Salardaine Quentin, Mongin Marie, Garcin Béatrice, Venance Laurent, et al. Video-based automated assessment of movement parameters consistent with MDS-UPDRS III in Parkinson's disease. J Parkinson's Dis 2022;(Preprint):1–12. http://dx.doi.org/10.3233/JPD-223445.

[113] Acevedo Trebbau Gabriela T, Bandini Andrea, Guarin Diego L. Video-based hand pose estimation for remote assessment of bradykinesia in Parkinson's disease. In: International workshop on pRedictive intelligence in mEdicine. Springer; 2023, p. 241–52. http://dx.doi.org/10.1007/978-3-031-46005-0_21.

[114] Developers, Openpose. OpenPose 1.1.0 benchmark. 2023, https://docs.google.com/spreadsheets/d/1-DynFGvoScvfWDA1P4jDInCkbD4lg0IKOYbXgEq0sK0/edit. (Accessed: 30 Oct 2023).

[115] Pang Yan, Christenson Jake, Jiang Feng, Lei Tim, Rhoades Remy, Kern Drew, et al. Automatic detection and quantification of hand movements toward development of an objective assessment of tremor and bradykinesia in Parkinson's disease. J Neurosci Methods 2020;333(108576):108576. http://dx.doi.org/10.1016/j.jneumeth.2019.108576.

[116] Li Hao, Shao Xiangxin, Zhang Chencheng, Qian Xiaohua. Automated assessment of Parkinsonian finger-tapping tests through a vision-based fine-grained classification model. Neurocomputing 2021;441:260–71. http://dx.doi.org/10.1016/j.neucom.2021.02.011.

[117] Lu Mandy, Zhao Qingyu, Poston Kathleen L, Sullivan Edith V, Pfefferbaum Adolf, Shahid Marian, et al. Quantifying Parkinson's disease motor severity under uncertainty using MDS-UPDRS videos. Med Image Anal 2021;73(102179):102179. http://dx.doi.org/10.1016/j.media.2021.102179.

[118] Park Kye Won, Lee Eun-Jae, Lee Jun Seong, Jeong Jinhoon, Choi Nari, Jo Sungyang, et al. Machine learning-based automatic rating for cardinal symptoms of Parkinson disease. Neurology 2021;96(13):e1761–9. http://dx.doi.org/10.1212/WNL.0000000000011654.

[119] Torrey Lisa, Shavlik Jude. Transfer learning. In: Handbook of research on machine learning applications and trends: algorithms, methods, and techniques. IGI global; 2010, p. 242–64. http://dx.doi.org/10.4018/978-1-60566-766-9.ch011.

[120] Howard Andrew G, Zhu Menglong, Chen Bo, Kalenichenko Dmitry, Wang Weijun, Weyand Tobias, et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications. 2017, http://dx.doi.org/10.48550/arXiv.1704.04861, arXiv preprint.

[121] He Kaiming, Zhang Xiangyu, Ren Shaoqing, Sun Jian. Deep residual learning for image recognition. In: 2016 IEEE conference on computer vision and pattern recognition. 2016, p. 770–8. http://dx.doi.org/10.1109/CVPR.2016.90.

[122] Tan Mingxing, Le Quoc. Efficientnet: Rethinking model scaling for convolutional neural networks. In: Proceedings of the 36th international conference on machine learning. PMLR; 2019, p. 6105–14. http://dx.doi.org/10.48550/arXiv.1905.11946, URL https://proceedings.mlr.press/v97/tan19a.html.

[123] Developers, DeepLabCut. DLC inference speed benchmark. 2023, https://deeplabcut.github.io/DLC-inferencespeed-benchmark/. (Accessed: 30 Oct 2023).

[124] Williams Stefan, Zhao Zhibin, Hafeez Awais, Wong David C, Relton Samuel D, Fang Hui, et al. The discerning eye of computer vision: Can it measure Parkinson's finger tap bradykinesia? J Neurol Sci 2020;416(117003):117003. http://dx.doi.org/10.1016/j.jns.2020.117003.

[125] Nunes Adonay S, Kozhemiako Nataliia, Stephen Christopher D, Schmahmann Jeremy D, Khan Sheraz, Gupta Anoopum S. Automatic classification and severity estimation of ataxia from finger tapping videos. Front Neurol 2021;12:795258. http://dx.doi.org/10.3389/fneur.2021.795258.

[126] Baker Sunderland, Tekriwal Anand, Felsen Gidon, Christensen Elijah, Hirt Lisa, Ojemann Steven G, et al. Automatic extraction of upper-limb kinematic activity using deep learning-based markerless tracking during deep brain stimulation implantation for Parkinson's disease: A proof of concept study. Plos one 2022;17(10):e0275490. http://dx.doi.org/10.1371/journal.pone.0275490.

[127] Xu Hongyi, Bazavan Eduard Gabriel, Zanfir Andrei, Freeman William T, Sukthankar Rahul, Sminchisescu Cristian. Ghum & ghuml: Generative 3d human shape and articulated pose models. In: 2020 IEEE/CVF conference on computer vision and pattern recognition. Los Alamitos, CA, USA: IEEE Computer Society; 2020, p. 6183–92. http://dx.doi.org/10.1109/CVPR42600.2020.00622.

[128] Lin Tsung-Yi, Dollár Piotr, Girshick Ross, He Kaiming, Hariharan Bharath, Belongie Serge. Feature pyramid networks for object detection. In: 2017 IEEE conference on computer vision and pattern recognition. Los Alamitos, CA, USA: IEEE Computer Society; 2017, p. 936–44. http://dx.doi.org/10.1109/CVPR.2017.106.

[129] Mediapipe hand landmarks detection guide. 2023, https://developers.google.com/mediapipe/solutions/vision/hand_landmarker. (Accessed: 21 July 2023).

[130] Amprimo Gianluca, Ferraris Claudia, Masi Giulia, Pettiti Giuseppe, Priano Lorenzo. GMH-D: Combining Google MediaPipe and RGB-depth cameras for hand motor skills remote assessment. In: 2022 IEEE international conference on digital health. 2022, p. 132–41. http://dx.doi.org/10.1109/ICDH55609.2022.00029.

[131] Wang Xinyi, Garg Saurabh, Tran Son N, Bai Quan, Alty Jane. Hand tremor detection in videos with cluttered background using neural network based approaches. Health Inf Sci Syst 2021;9(1):30. http://dx.doi.org/10.1007/s13755-021-00159-3.

[132] Li Zhu, Lu Kang, Cai Miao, Liu Xiaoli, Wang Yanwen, Yang Jiayu. An automatic evaluation method for Parkinson's dyskinesia using finger tapping video for small samples. J Med Biol Eng 2022;42(3):351–63. http://dx.doi.org/10.1007/s40846-022-00701-y.

[133] Güney Gökhan, Jansen Talisa S, Dill Sebastian, Schulz Jörg B, Dafotakis Manuel, Hoog Antink Christoph, et al. Video-based hand movement analysis of Parkinson patients before and after medication using high-frame-rate videos and MediaPipe. Sensors (Basel) 2022;22(20):7992. http://dx.doi.org/10.3390/s22207992.

[134] Amprimo Gianluca, Rechichi Irene, Ferraris Claudia, Olmo Gabriella. Objective assessment of the finger tapping task in Parkinson's disease and control subjects using Azure kinect and machine learning. In: 2023 IEEE 36th international symposium on computer-based medical systems. 2023, p. 640–5. http://dx.doi.org/10.1109/CBMS58004.2023.00293.

[135] MMPose Contributors. OpenMMLab pose estimation toolbox and benchmark. 2023, https://github.com/open-mmlab/mmpose. (Accessed: 24 Oct 2023).

[136] OpenMMLab contributors. OpenMMLab. 2023, https://openmmlab.com. (Accessed: 24 Oct 2023).

[137] Sun Ke, Xiao Bin, Liu Dong, Wang Jingdong. Deep high-resolution representation learning for human pose estimation. In: 2019 IEEE conference on computer vision and pattern recognition. 2019, http://dx.doi.org/10.1109/CVPR.2019.00584.

[138] Toshev Alexander, Szegedy Christian. DeepPose: Human pose estimation via deep neural networks. In: 2014 IEEE conference on computer vision and pattern recognition. 2014, p. 1653–60. http://dx.doi.org/10.1109/CVPR.2014.214.

[139] Zhang Feng, Zhu Xiatian, Dai Hanbin, Ye Mao, Zhu Ce. Distribution-aware coordinate representation for human pose estimation. In: 2020 IEEE/CVF conference on computer vision and pattern recognition. 2020, p. 7091–100. http://dx.doi.org/10.1109/CVPR42600.2020.00712.

[140] Li J, Bian S, Zeng A, Wang C, Pang B, Liu W, et al. Human pose regression with residual log-likelihood estimation. In: 2021 IEEE/CVF international conference on computer vision. Los Alamitos, CA, USA: IEEE Computer Society; 2021, p. 11005–14. http://dx.doi.org/10.1109/ICCV48922.2021.01084.

[141] Yang Ning, Liu De-Feng, Liu Tao, Han Tianyuan, Zhang Pingyue, Xu Xuenan, et al. Automatic detection pipeline for accessing the motor severity of Parkinson's disease in finger tapping and postural stability. IEEE Access 2022;10:66961–73. http://dx.doi.org/10.1109/ACCESS.2022.3183232.

[142] Xie Zheng, Guo Rui, Zhang Chencheng, Qian Xiaohua. A clinically guided graph convolutional network for assessment of Parkinsonian pronation-supination movements of hands. IEEE Trans Circuits Syst Video Technol 2023;1. http://dx.doi.org/10.1109/TCSVT.2023.3318243.

[143] Zhang MingYa, Zhao Na, Yu Yue, Zhuang YuQian, Zhu QuanQiu, Huang TianYuan, et al. A simple yet effective hand pose tremor classification algorithm to diagnosis Parkinsons disease. In: 2022 IEEE international conference on bioinformatics and biomedicine. 2022, p. 887–90. http://dx.doi.org/10.1109/BIBM55620.2022.9995709.

[144] Lin Bo, Luo Wei, Luo Zhiling, Wang Bo, Deng Shuiguang, Yin Jianwei, et al. Bradykinesia recognition in Parkinson's disease via single RGB video. ACM Trans Knowl Discov Data 2020;14(2):1–19. http://dx.doi.org/10.1145/3369438.

[145] Xu Yufei, Zhang Jing, Zhang Qiming, Tao Dacheng. ViTPose: Simple vision transformer baselines for human pose estimation. In: Advances in neural information processing systems, vol. 35, Curran Associates, Inc; 2022, p. 38571–84, URL https://proceedings.neurips.cc/paper_files/paper/2022/file/fbb10d319d44f8c3b4720873e4177c65-Paper-Conference.pdf.

[146] Liu Yu, Chen Jiansheng, Hu Chunhua, Ma Yu, Ge Dongyun, Miao Suhua, et al. Vision-based method for automatic quantification of Parkinsonian bradykinesia. IEEE Trans Neural Syst Rehabil Eng 2019;27(10):1952–61. http://dx.doi.org/10.1109/TNSRE.2019.2939596.

[147] Chen Yifei, Ma Haoyu, Wang Jiangyuan, Wu Jianbao, Wu Xian, Xie Xiaohui. 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. 2021, p. 2683–91. http://dx.doi.org/10.1145/3447548.3467130.

[148] Ge L, Ren Z, Li Y, Xue Z, Wang Y, Cai J, et al. 3D hand shape and pose estimation from a single RGB image. In: 2019 IEEE/CVF conference on computer vision and pattern recognition. Los Alamitos, CA, USA: IEEE Computer Society; 2019, p. 10825–34. http://dx.doi.org/10.1109/CVPR.2019.01109.

[149] Andriluka Mykhaylo, Pishchulin Leonid, Gehler Peter, Schiele Bernt. 2D human pose estimation: New benchmark and state of the art analysis. In: 2014 IEEE conference on computer vision and pattern recognition. 2014, p. 3686–93. http://dx.doi.org/10.1109/CVPR.2014.471.

[150] Redmon Joseph, Farhadi Ali. Yolov3: An incremental improvement. 2018, http://dx.doi.org/10.48550/arXiv.1804.02767, arXiv preprint.

[151] Yuan Shanxin, Ye Qi, Garcia-Hernando Guillermo, Kim Tae-Kyun. The 2017 hands in the million challenge on 3d hand pose estimation. 2017, http://dx.doi.org/10.48550/arXiv.1707.02237, arXiv preprint.

[152] Newell Alejandro, Yang Kaiyu, Deng Jia. Stacked hourglass networks for human pose estimation. In: 2016 computer vision–ECCV: 14th European conference. Springer; 2016, p. 483–99. http://dx.doi.org/10.1007/978-3-319-46484-8_29.

[153] Ronneberger Olaf, Fischer Philipp, Brox Thomas. U-net: Convolutional networks for biomedical image segmentation. In: 2015 medical image computing and computer-assisted intervention–MICCAI: 18th international conference. Springer; 2015, p. 234–41. http://dx.doi.org/10.1007/978-3-319-24574-4_28.

[154] Zimmermann Christian, Ceylan Duygu, Yang Jimei, Russell Bryan, Argus Max, Brox Thomas. Freihand: A dataset for markerless capture of hand pose and shape from single rgb images. In: 2019 IEEE/CVF international conference on computer vision. 2019, p. 813–22. http://dx.doi.org/10.1109/ICCV.2019.00090.

[155] Zhu Chenfei, Hu Boce, Chen Jiawei, Ai Xupeng, Agrawal Sunil K. SARN: Shifted attention regression network for 3D hand pose estimation. Bioengineering 2023;10(2):126. http://dx.doi.org/10.3390/bioengineering10020126.

[156] Maetzler Walter, Klucken Jochen, Horne Malcolm. A clinical view on the development of technology-based tools in managing Parkinson's disease. Mov Disord 2016;31(9):1263–71. http://dx.doi.org/10.1002/mds.26673.

[157] Espay Alberto J, Bonato Paolo, Nahab Fatta B, Maetzler Walter, Dean John M, Klucken Jochen, et al. Technology in Parkinson's disease: Challenges and opportunities. Mov Disord 2016;31(9):1272–82. http://dx.doi.org/10.1002/mds.26642.

[158] Godoy Junior Carlos Antonio, Miele Francesco, Mäkitie Laura, Fiorenzato Eleonora, Koivu Maija, Bakker Lytske Jantien, et al. Attitudes toward the adoption of remote patient monitoring and artificial intelligence in Parkinson's disease management: Perspectives of patients and neurologists. Patient-Patient-Centered Outcomes Res. 2024;1–11. http://dx.doi.org/10.1007/s40271-023-00669-0.

[159] Amprimo Gianluca, Masi Giulia, Pettiti Giuseppe, Olmo Gabriella, Priano Lorenzo, Ferraris Claudia. Hand tracking for clinical applications: Validation of the Google MediaPipe hand (GMH) and the depth-enhanced GMH-D frameworks. 2023, http://dx.doi.org/10.48550/arXiv.2308.01088, arXiv preprint.

[160] Laar Arjonne, de Lima Ana Ligia Silva, Maas Bart R, Bloem Bastiaan R, de Vries Nienke M. Successful implementation of technology in the management of Parkinson's disease: Barriers and facilitators. Clin Parkinsonism Rel Disord 2023;8:100188. http://dx.doi.org/10.1016/j.prdoa.2023.100188.

[161] Fiorini Laura, Loizzo Federica G Cornacchia, Sorrentino Alessandra, Kim Jaeseok, Rovini Erika, Di Nuovo Alessandro, et al. Daily gesture recognition during human-robot interaction combining vision and wearable systems. IEEE Sens J 2021;21(20):23568–77. http://dx.doi.org/10.1109/JSEN.2021.3108011.

[162] Gosala Nikhil, Wang Fangjinhua, Cui Zhaopeng, Liang Hanxue, Glauser Oliver, Wu Shihao, et al. Self-calibrated multi-sensor wearable for hand tracking and modeling. IEEE Trans Vis Comput Graphics 2023;29(3):1769–84. http://dx.doi.org/10.1109/TVCG.2021.3131230.

[163] Wang Jiayi, Mueller Franziska, Bernard Florian, Sorli Suzanne, Sotnychenko Oleksandr, Qian Neng, et al. Rgb2hands: Real-time tracking of 3d hand interactions from monocular rgb video. ACM Trans Graph (ToG) 2020;39(6):1–16. http://dx.doi.org/10.1145/3414685.3417852.

[164] Cao Zhe, Radosavovic Ilija, Kanazawa Angjoo, Malik Jitendra. Reconstructing hand-object interactions in the wild. In: 2021 IEEE/CVF international conference on computer vision. 2021, p. 12397–406. http://dx.doi.org/10.1109/ICCV48922.2021.01219.

[165] Huang Lin, Tan Jianchao, Meng Jingjing, Liu Ji, Yuan Junsong. Hot-net: Non-autoregressive transformer for 3d hand-object pose estimation. In: Proceedings of the 28th ACM international conference on multimedia. 2020, p. 3136–45. http://dx.doi.org/10.1145/3394171.3413775.

[166] Chen Jiayi, Yan Mi, Zhang Jiazhao, Xu Yinzhen, Li Xiaolong, Weng Yijia, et al. Tracking and reconstructing hand object interactions from point cloud sequences in the wild. In: Proceedings of the AAAI conference on artificial intelligence, vol. 37, (no. 1):2023, p. 304–12. http://dx.doi.org/10.1609/aaai.v37i1.25103.

Gianluca Amprimo received his Master's Degree in Computer Engineering from the Politecnico di Torino in 2020. He is currently a PhD student at the Control and Computer Engineering Department of Politecnico di Torino and a Research Fellow at the National Research Council (CNR-IEIIT) of Italy. His main research interests include human pose estimation from video, innovative technologies for telemonitoring and rehabilitation, and AI for medical applications.

Giulia Masi recevied her Master's Degree in Biomedical Engineering at Politecnico di Torino in 2021, specialising in biosignal processing. She then followed her research interests in neuroscience working as a research fellow for the neuroscience department of the University of Turin. She is currently a PhD student at the Control and Computer Engineering department of Politecnico di Torino. In particular, she is interested in the study of sleep and motion in the neurodegenerative diseases, as well as stress, using objective and quantitative measurements, such neurophysiological signals and motion trajectories.

Gabriella Olmo (IEEE Senior Member) received her M.E. and Ph.D. degrees in Electronic Engineering from Politecnico di Torino, Italy, in 1986 and 1992, respectively. In 2016, she received her Master's degree in Medicine and Surgery from Università di Torino, Italy. She is currently a full professor in the Department of Control and Computer Engineering, Politecnico di Torino, Italy. Her main research interests are in the fields of wearable sensors, signal processing, and machine learning techniques for medical applications. She is coauthor of more than 250 publications in international journals and proceedings in international conferences.

Claudia Ferraris received her degree in Computer Science from the University of Turin (Italy) in 1997, then joined CNR working on image/video coding, compression techniques, and motion estimation algorithms. After a long work experience in the industrial field, she joined again CNR-IEIIT, to work on non-invasive technologies for motion analysis, remote monitoring, and rehabilitation in elderly and pathological conditions. Since 2020, she has been a permanent researcher at the same organisation. She received her PhD in Neuroscience from the University of Turin (Italy) in 2021. She is the author and co-author of numerous research papers published in national and international journals, books, and conference proceedings.