






Article

# Vision-Based Hand Function Evaluation with Soft Robotic Rehabilitation Glove

Mukun Tong <sup>1</sup>, Michael Cheung <sup>1</sup>, Yixing Lei <sup>1,2</sup>, Mauricio Villarroel <sup>1,2</sup> and Liang He <sup>1,2,\*</sup>

<sup>1</sup> Department of Engineering Science, University of Oxford, Oxford OX1 3PJ, UK; mukun.tong@eng.ox.ac.uk (M.T.); michael.cheung@hmc.ox.ac.uk (M.C.); chri6867@ox.ac.uk (Y.L.); mauricio.villarroel@eng.ox.ac.uk (M.V.)

<sup>2</sup> The Podium Institute for Sports Medicine and Technology, University of Oxford, Oxford OX1 3PJ, UK

\* Correspondence: liang.he@eng.ox.ac.uk

## Abstract

Advances in robotic technology for hand rehabilitation, particularly soft robotic gloves, have significant potential to improve patient outcomes. While vision-based algorithms pave the way for fast and convenient hand pose estimation, most current models struggle to accurately track hand movements when soft robotic gloves are used, primarily due to severe occlusion. This limitation reduces the applicability of soft robotic gloves in digital and remote rehabilitation assessment. Furthermore, traditional clinical assessments like the Fugl-Meyer Assessment (FMA) rely on manual measurements and subjective scoring scales, lacking the efficiency and quantitative accuracy needed to monitor hand function recovery in data-driven personalised rehabilitation. Consequently, few integrated evaluation systems provide reliable quantitative assessments. In this work, we propose an RGB-based evaluation system for soft robotic glove applications, which is aimed at bridging these gaps in assessing hand function. By incorporating the Hand Mesh Reconstruction (HaMeR) model fine-tuned with motion capture data, our hand estimation framework overcomes occlusion and enables accurate continuous tracking of hand movements with reduced errors. The resulting functional metrics include conventional clinical benchmarks such as the mean per joint angle error (MPJAE) and range of motion (ROM), providing quantitative, consistent measures of rehabilitation progress and achieving tracking errors lower than 10°. In addition, we introduce adapted benchmarks such as the angle percentage of correct keypoints (APCK), mean per joint angular velocity error (MPJAVE) and angular spectral arc length (SPARC) error to characterise movement stability and smoothness. This extensible and adaptable solution demonstrates the potential of vision-based systems for future clinical and home-based rehabilitation assessment.

**Keywords:** soft robotic glove; computer vision; hand pose estimation; hand rehabilitation; quantitative evaluation



Academic Editor: Yuanyuan Yang

Received: 2 December 2025

Revised: 18 December 2025

Accepted: 23 December 2025

Published: 25 December 2025

**Copyright:** © 2025 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons Attribution \(CC BY\) license](https://creativecommons.org/licenses/by/4.0/).

## 1. Introduction

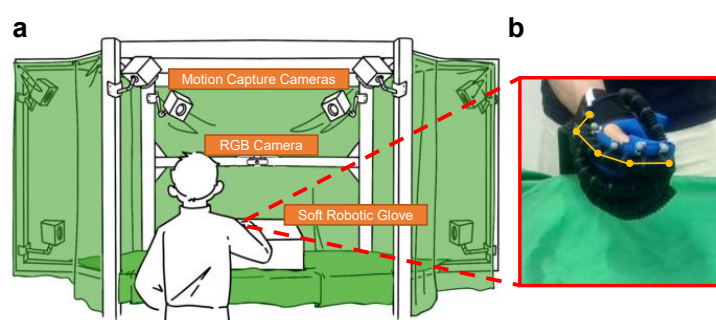
Hand function plays a crucial role in daily activities, and loss of hand motor abilities including the joint range of motion (ROM) can significantly impact a person's quality of life [1]. In hand function rehabilitation, soft robotic gloves and assistive technologies have gained emerging interest for aiding hand function recovery. Soft robotic gloves are wearable assistive devices with compliant, pneumatically or tendon-driven actuators that support finger flexion and extension during rehabilitation tasks, enabling safe, adaptive

and user-comfortable assistance for motor recovery [2–5]. While recent advances in computer vision have shown excellent results in bare-hand posture estimation, applying these models to evaluate hand function with soft robotic gloves remains challenging [6,7]. The primary difficulty lies in accurately tracking hand poses when the patient’s hand is fully or partially covered by soft robotic gloves. Issues like occlusion, finger slippage and glove size variability also contribute to significant tracking errors [8–10]. When parts of the hand or fingers are hidden from the camera’s view, current vision-based models struggle to maintain accuracy [11–14].

Moreover, an effective approach to quantitatively evaluate hand function as well as the effectiveness of the latest technologies, including soft robotic gloves, remains an open challenge. Clinically, the ROM is often assessed through structured tests like the Fugl-Meyer Assessment (FMA) [15]. These standardised tests evaluate the hand motor function, balance and sensation of patients by scoring their performance in various movements across multiple joints and fingers as well as in both hands. However, they typically rely on manual measurements and subject scoring scales, which can yield inconsistent results and demand considerable staff time, thereby limiting their usability. Quantitative metrics like the joint ROM in degrees provide more precise information to help clinicians understand patient recovery stages and further develop personalised recovery plans.

Recent advances have explored embedding sensors in soft robotic gloves to quantitatively measure joint movements [16]. For example, flexible sensors and inertial measurement units (IMUs) can provide real-time data on finger joint angles, offering valuable information for feedback control of gloves in rehabilitation [17]. Nevertheless, incorporating embedded sensors would significantly increase the cost of soft robotic gloves, and many flexible sensors still face challenges with reliability and durability under repeated use [16,18]. In contrast, using camera-based systems to measure the posture of soft robotic gloves and evaluate rehabilitation performance presents a promising solution for fast, reliable and low-cost implementation in both clinical and home environments [19,20].

The motivation for this research lies in overcoming these limitations. We propose a hand motion evaluation system integrating camera-based hand tracking with minimal fine-tuning of the Hand Mesh Reconstruction (HaMeR) model [7] to manage occlusions caused by gloves or hand positioning, as shown in Figure 1. It compensates for occluded areas, ensuring continuous tracking even in difficult scenarios. Additionally, this method introduces quantitative evaluation metrics, including the range of motion (ROM) and mean per joint angle error (MPJAE), which provide standardised assessments in rehabilitation and clearer insights into movement dynamics. Unlike models that require frequent recalibration, this method offers an adaptable and scalable solution for various rehabilitation applications, achieving tracking errors below  $10^\circ$  even under challenging conditions. The system reduces the need for extensive retraining, making it suitable for long-term rehabilitation in both clinical and home-based settings.



**Figure 1.** Motion capture system for calibration and visualisation of finger motion with soft robotic gloves. (a) Motion capture system. (b) Estimated finger motion with the proposed RGB-based model.

## 2. Materials and Methods

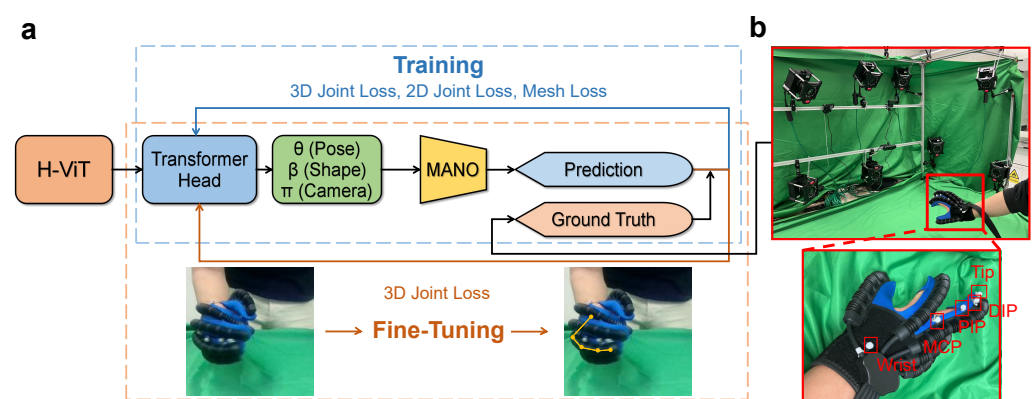
### 2.1. Overview of the Vision-Based Evaluation System

This computer vision-based system was constructed to quantitatively evaluate the hand recovery process with the assistance of soft robotic gloves. We aimed to fine-tune a powerful hand pose estimation model for our scenarios where hands are wearing rehabilitation gloves as well as define the metrics for evaluation.

We applied the HaMeR [7] model for fine-tuning, which is a fully transformer-based approach designed for reconstructing 3D hand meshes from monocular images or video frames. Its architecture leverages a large-scale vision transformer (ViT) as the backbone (also known as H-ViT), which processes the image patches and returns a series of output tokens. It also includes a transformer head to transfer the tokens into MANO and camera parameters, which can then be transformed to joint positions and hand meshes [21]. HaMeR excels in capturing complex hand configurations, thanks to its capacity to scale with larger datasets and utilise powerful deep learning architectures. It consistently outperforms previous state-of-the-art methods in hand pose benchmarks, particularly in challenging wild scenarios, such as hands interacting with objects or other hands, or hands captured from different viewpoints. However, this model still could not be applied to our situation due to the lack of training data captured from hands with rehabilitation gloves, as the testing results turn out to be unsatisfactory in Section 3. Thus, a fine-tuning method is proposed in our system to help bridge this gap in performance.

In evaluation, different from qualitative measurements in clinical tests, we naturally applied the joint angle accuracy from hand pose estimation tasks to our framework, including the angle error and the percentage of correctly predicted joints within different error thresholds. These are commonly used metrics for hand pose estimation, especially when joint angles are taken into consideration, from which we could easily calculate the ROM for each joint, which is a comparatively accurate standard of measurement.

Our training and fine-tuning framework is shown in Figure 2. Firstly, the model was trained following the process from the original HaMeR [7], computing the 3D joint loss, 2D joint loss and mesh loss. After training on large datasets, the transformer head was then fine-tuned on our small-scale dataset collected from gloved hands to fit to our environment, merely considering the 3D joint loss. The details will be explicated in Section 2.3.



**Figure 2.** Our training and fine-tuning framework, as well as the data collection system and markers on the glove. The HaMeR model was firstly trained on a large-scale dataset and then fine-tuned on our glove dataset. In data collection, the motion capture system had 8 motion capture cameras to record joint coordinates and 1 RGB camera to capture RGB images. (a) Training and fine-tuning framework. (b) Data collection system and markers on the glove.

## 2.2. Data Collection

As our dataset included RGB images and corresponding 3D joint positions for calibration, we used a motion capture system to capture the data we needed, as shown in Figure 2. Here, we abbreviate the joints of a finger as the metacarpophalangeal joint (MCP), proximal interphalangeal joint (PIP) and distal interphalangeal joint (DIP). Note that the joints of the thumb, from the wrist to the tip, are abbreviated as the trapeziometacarpal joint (TM), MCP and interphalangeal joint (IP). For clarity, the thumb joints (TM–MCP–IP) follow a similar notation scheme to the remaining fingers (MCP–PIP–DIP), and the same abbreviations are used interchangeably when no ambiguity arises. Covered by green cloth, the system used 8 NOKOV MARS4H motion capture cameras (NOKOV Science & Technology Co., Ltd., Beijing, China) to locate the joint coordinates and 1 RGB camera (720 p, 30 fps) facing towards the hand to capture images. To represent the positions of the joints, we used grey spherical markers stuck on the glove. Considering the mutual inference of the markers when all 21 joints of the hands were simultaneously present, in each image, only one finger and the wrist were focused on. Thus, there were five markers to capture.

Our dataset consisted of  $\sim 3000$  images for training,  $\sim 1000$  for validation and  $\sim 1000$  for testing, with corresponding 3D joint positions, captured from one subject. However, current automatic 2D hand pose annotation methods such as MMPose [22] have been proven to perform poorly, especially when dealing with hand poses with self- and object occlusions. That aside, manually annotating 2D keypoints for each view is fairly expensive for data collection. However, our experiments have shown that 3D data alone is enough for fine-tuning. In addition, mesh data requiring scanners to collect is also unnecessary in our framework.

## 2.3. Training and Fine-Tuning

In the training process, the dataset was contributed by the authors of HaMeR [7], which consists of 40,400 RGB images in total. The model computes the 3D joint position losses, 2D joint position losses and mesh losses. For the ground-truth 3D joint positions  $X^*$ , 2D joint positions  $x^*$  and MANO parameters  $\theta^*, \beta^*$  and the predicted 3D joint positions  $X$ , 2D joint positions  $x$  and MANO parameters  $\theta, \beta$ , the loss is calculated as follows [7]:

$$\begin{aligned}\mathcal{L}_{Training} &= w_{3D}\mathcal{L}_{3D} + w_{2D}\mathcal{L}_{2D} + w_{MANO}\mathcal{L}_{MANO} \\ &= w_{3D}\|X - X^*\| + w_{2D}\|x - x^*\| + \\ &\quad w_{MANO}(\|\theta - \theta^*\| + \|\beta - \beta^*\|)\end{aligned}\quad (1)$$

where  $w_{3D}, w_{2D}, w_{MANO}$  refer to the weights of the 3D (0.05), 2D (0.01) and MANO parameter (0.0005) losses, respectively. Different from the training process, we merely utilised the 3D joint losses for fine-tuning:

$$\mathcal{L}_{Fine-Tuning} = \mathcal{L}_{3D} = \|X - X^*\| \quad (2)$$

This simplified process gets rid of complex annotation of the 2D ground truth joint positions and supervises the level of the 3D joints, encouraging consistency in the 3D space.

## 2.4. Implementation Details

All the experiments were conducted on a single Tesla V100 GPU, with Python 3.10.15, PyTorch 2.6.0 + CUDA 12.6 on a Linux workstation. Numerical and scientific computations were performed with NumPy 1.26.1 and SciPy 1.14.1, while image processing and visualisation relied on OpenCV 4.11.0 and Matplotlib 3.9.2. The MANO hand model used in this work corresponds to version 1.2, obtained from the official release [21]. All software packages were installed in a controlled environment to ensure reproducibility.

In training, all the settings remained the same as those provided by the authors of HaMeR [7], with the number of epochs being 1000, the learning rate being  $1 \times 10^{-5}$  and the weight decay being  $1 \times 10^{-4}$ . In the fine-tuning process, the number of epochs was 40, with a learning rate of  $1 \times 10^{-5}$  and weight decay of  $1 \times 10^{-4}$ . Aside from that, as in our data collection, the markers were stuck on the glove rather than the skin, and there was an offset between the real joint positions and those of the markers. As the offset of the markers on the fingers was the same, the model could adapt to the offset during the fine-tuning process. However, the marker representing the wrist was comparatively far from the ground truth, and thus we added  $40^\circ$  to the output MCP angle (since the marker was a bit higher than the skin). Note that this did not influence the fine-tuning process, as the angles were calculated from the joint positions and not included in the loss computation. We compare the performance of our fine-tuned HaMeR model with the model without fine-tuning as the baseline to prove the effect of our framework.

### 3. Results

In addition to the original HaMeR model, we included Hamba as a representative interaction-aware RGB-based baseline [23]. Hamba has demonstrated robustness in challenging hand–object interaction scenarios and serves as a competitive comparison for evaluating performance under glove-induced occlusion.

#### 3.1. Joint Angle Accuracy

To evaluate the accuracy of three joint angles per finger, we chose the mean per joint angle error (MPJAE) [24] to measure the error:

$$MPJAE = \frac{1}{N} \sum_{i=1}^N |\theta_i - \theta_i^*| \quad (3)$$

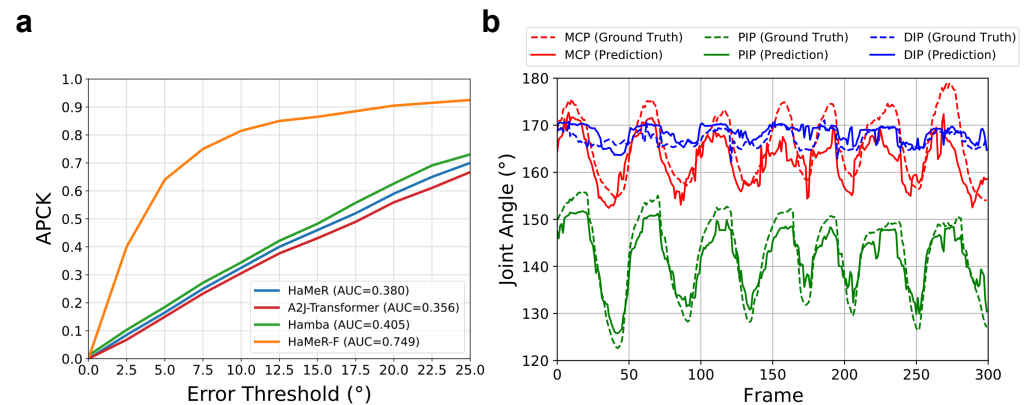
where  $N$  is the number of joints and  $\theta_i^*$  and  $\theta_i$  are the ground-truth and predicted angles, respectively. Based on the percentage of correct keypoints (PCK) score [25], we introduce the angle PCK (APCK) score, which represents the percentage of correctly predicted angles with an error threshold (unit:  $^\circ$ ). Subsequently, we could draw the curve of the APCK score with different thresholds and calculate the area under the curve (AUC).

We present the MPJAE and angle PCK scores with error thresholds of  $5^\circ$  and  $10^\circ$  in Table 1, where we abbreviate our fine-tuned HaMeR model as HaMeR-F. We observed that over 60% of the angle prediction reached an error lower than  $5^\circ$ , and more than 80% of the angles were accurately predicted with a threshold of  $10^\circ$ . The APCK score curves with different error thresholds are illustrated in Figure 3a, where the AUC of our fine-tuned model reached almost twice those of the Hamba model and HaMeR model without fine-tuning. However, it can also be seen that the model performed worst for the MCP angle. This phenomenon can be partly explained by the huge offset of the wrist, which is hard to concisely rectify in different environments and with different movements.

We also demonstrate the ground-truth and predicted ROMs of the joints in Table 2, each presented as the minimum and maximum angle of the joint. That aside, to observe the dynamic estimation of the joint angles, we tested our model on a 10 s (300 frames) video recording movements of the middle finger and compared the predicted angle curves with the ground truth in Figure 3b. In general, the output results manifested the finger movements in the temporal sequence well, further validating its feasibility in ROM monitoring.

**Table 1.** 3D joint angle estimation.

Joint	Method	MPJAE (°)	APCK@5	APCK@10
MCP	HaMeR	29.76 ± 17.78	0.071	0.126
	Hamba	25.62 ± 15.61	0.082	0.137
	HaMeR-F	9.39 ± 8.58	0.288	0.514
PIP	HaMeR	15.44 ± 10.30	0.176	0.361
	Hamba	12.28 ± 11.76	0.183	0.376
	HaMeR-F	3.32 ± 2.89	0.783	0.957
DIP	HaMeR	12.85 ± 9.97	0.263	0.493
	Hamba	10.67 ± 13.58	0.290	0.519
	HaMeR-F	2.74 ± 2.83	0.861	0.969
Overall	HaMeR	19.35 ± 15.14	0.170	0.326
	Hamba	16.19 ± 15.98	0.185	0.344
	HaMeR-F	5.48 ± 6.17	0.644	0.814



**Figure 3.** (a) APCK curves of HaMeR model and our fine-tuned model. The score of our model was apparently higher than that of the original HaMeR model. (b) Joint angle curves for the middle finger. Our model successfully achieved the temporal robustness for predicting joint angles, with errors lower than 10°.

**Table 2.** 3D joint ROM evaluation (°).

Finger	Result	MCP	PIP	DIP
Thumb	GT	100.6~108.7	130.0~174.5	143.8~179.4
	HaMeR-F	100.0~109.6	131.2~170.2	151.9~176.5
Index	GT	155.9~176.4	120.4~152.9	154.8~172.4
	HaMeR-F	150.3~169.4	126.8~151.0	151.8~169.0
Middle	GT	148.0~179.3	123.7~155.7	164.4~172.1
	HaMeR-F	140.0~169.3	129.7~148.1	161.1~169.0
Ring	GT	126.3~157.5	143.2~150.9	158.8~174.1
	HaMeR-F	120.8~159.6	145.0~153.9	158.0~165.3
Pinky	GT	150.0~163.4	129.8~151.5	159.2~172.7
	HaMeR-F	140.0~159.3	137.6~152.0	158.0~167.0

### 3.2. Qualitative Results

We demonstrate here the visualisation of 3D hand joints for the thumb, middle finger and pinky finger in Figure 4. After fine-tuning, the model became consistently robust and accurate in predicting the joint angles as well as the finger poses, even with the glove.



**Figure 4.** Qualitative results of predictions. Our model accurately estimated finger poses at different angles. (a) Thumb motions. (b) Middle finger motions. (c) Pinky finger motions.

### 3.3. Ablation Study: 3D Joint Accuracy and Kinematic Smoothness

In this ablation study, we further analysed both the 3D joint accuracy and the kinematic quality of the reconstructed motion. The 3D accuracy was quantified with the mean per joint position error MPJPE, which measures the Euclidean distance between the predicted and ground-truth joint positions in millimetres. Beyond static positional accuracy, we introduce two kinematic-level metrics to assess temporal fidelity:

- **Mean per Joint Angular Velocity Error MPJAVE** This was computed as the mean absolute difference between the predicted and ground-truth angular velocities (in  $^{\circ}/s$ ) of each joint. This metric reflects how well the model captured the motion dynamics, i.e., the speed consistency of each finger joint during flexion and extension. A smaller MPJAVE indicates more temporally stable motion estimation.
- **Angular SPARC Error:** This is derived from the spectral arc length SPARC metric, a frequency-domain measure of motion smoothness [26]. The SPARC quantifies how smoothly a joint angle trajectory evolves over time by integrating the curvature of its amplitude spectrum. Here, we report the absolute SPARC difference between the estimated and ground-truth trajectories, where lower values denote higher smoothness consistency.

Table 3 summarises the per-joint results of the baseline models (HaMeR and Hamba) and the fine-tuned model (HaMeR-F). While fine-tuning significantly improved the spatial accuracy in terms of joint position and angle estimation, the gains in temporal metrics such as the MPJAVE and SPARC were more limited. This is expected, as the fine-tuning process optimises the per-frame spatial alignment using 3D joint supervision without explicitly modeling temporal dynamics or motion smoothness priors. Consequently, improvements in velocity consistency and spectral smoothness arise primarily as indirect effects of reduced spatial noise rather than from dedicated temporal constraints. Incorporating sequence-level supervision or explicit temporal regularisation is therefore a promising direction for further improving kinematic smoothness.

**Table 3.** 3D joint position and kinematic accuracy evaluation, where the MPJAVE and Angular SPARC Error for the wrist and tip cannot be meaningfully computed due to the lack of well-defined flexion–extension trajectories at these joints, and are therefore marked as N/A.

Joint	Method	MPJPE (mm)	MPJAVE ( $^{\circ}/s$ )	Angular SPARC Error
Wrist	HaMeR	$24.02 \pm 5.68$	N/A	N/A
	Hamba	$23.09 \pm 4.77$	N/A	N/A
	HaMeR-F	$23.06 \pm 2.00$	N/A	N/A
MCP	HaMeR	$15.85 \pm 3.71$	$40.49 \pm 8.94$	$2.32 \pm 2.29$
	Hamba	$14.37 \pm 3.68$	$38.95 \pm 8.68$	$2.20 \pm 2.27$
	HaMeR-F	$12.15 \pm 1.57$	$30.47 \pm 4.99$	$1.73 \pm 2.25$
PIP	HaMeR	$15.52 \pm 3.45$	$78.98 \pm 8.70$	$3.55 \pm 1.22$
	Hamba	$15.20 \pm 3.38$	$75.23 \pm 8.56$	$3.39 \pm 1.20$
	HaMeR-F	$12.23 \pm 1.63$	$63.71 \pm 7.52$	$2.90 \pm 0.96$

Table 3. Cont.

Joint	Method	MPJPE (mm)	MPJAVE (°/s)	Angular SPARC Error
DIP	HaMeR	14.88 ± 3.35	62.12 ± 9.04	2.04 ± 2.05
	Hamba	14.22 ± 2.99	58.98 ± 8.87	1.92 ± 1.95
	HaMeR-F	12.32 ± 1.74	46.78 ± 5.38	1.28 ± 1.73
Tip	HaMeR	14.28 ± 3.31	N/A	N/A
	Hamba	13.96 ± 3.20	N/A	N/A
	HaMeR-F	12.40 ± 1.86	N/A	N/A
Overall	HaMeR	16.91 ± 4.39	60.53 ± 8.89	2.64 ± 1.91
	Hamba	16.17 ± 3.74	57.72 ± 8.27	2.50 ± 1.85
	HaMeR-F	14.43 ± 2.24	46.99 ± 6.06	1.97 ± 1.73

#### 4. Discussion

The paper propels the use of soft-robotic-glove-based hand and finger tracking into more accessible rehabilitation environments, particularly by improving its integration into clinical settings. With the fine-tuned vision-based model and the use of an RGB-based camera, this approach offers a promising solution for estimating the ROMs for hands accurately. This method also offers standardised assessments for rehabilitation by building quantitative evaluation metrics, such as the ROM, MPJAE, MPJPE, MPJAVE and SPARC errors, which provide clearer insights into movement dynamics.

However, it should be noted that the current study focused on a subject-specific calibration scenario, with all training, validation and testing data collected from a single participant wearing one soft robotic glove under controlled conditions. This setting is aligned with practical rehabilitation use cases, where soft robotic gloves are typically calibrated to individual users due to variations in hand size, glove fit and motor ability. With calibration, the proposed framework already enables accurate and continuous estimation of joint angles and functional metrics (e.g., ROM and kinematic consistency) using only an RGB camera, providing a low-cost and easily deployable evaluation tool that can be readily adopted for subject-level assessment, system benchmarking and algorithm validation.

At the same time, cross-subject generalisation remains a well-recognised challenge in vision-based hand pose estimation. Prior studies have shown that models trained on limited subject populations often struggle to transfer reliably to unseen individuals with different hand geometries, appearance characteristics and motion patterns, particularly under occlusion and domain shifts [27]. Recent systematic reviews further indicate that despite rapid progress in deep learning-based methods, robust generalisation across diverse real-world conditions and subject populations is still an open problem due to anatomical variability, occlusions and dataset bias [28]. Thus, the present study is positioned as an initial feasibility investigation rather than a cross-subject evaluation.

The evaluation was therefore conducted using isolated single-finger motions, fixed camera viewpoints and uniform backgrounds to intentionally isolate the effect of visual occlusion in a controlled setting. In addition, the wrist angle correction applied in this study was used only during post-processing for evaluation purposes and did not affect model training or optimisation. This subject-specific adjustment further motivates the need for systematic calibration strategies. Building on the current framework, future work will expand the dataset to include multiple subjects and more diverse motion patterns, investigate cross-subject generalisation strategies and explore more comprehensive functional and kinematic metrics that better capture movement coordination, temporal consistency and rehabilitation-relevant performance beyond joint-level accuracy.

Although clinical relevance is a central motivation of this work, therapist involvement in the present study was limited and did not include qualitative clinical assessments. Con-

sequently, the study did not include validation of patient populations or direct alignment with established clinical assessment scales, such as the Fugl-Meyer Assessment (FMA) scale [15]. The quantitative metrics reported here (ROM, MPJAE, MPJPE, MPJAVE and angular SPARC error) are therefore intended as complementary, objective descriptors rather than replacements for clinician-rated scales. From a clinical workflow perspective, the proposed system is best regarded as a supportive tool for assessments with subject-specific calibration, enabling quantitative monitoring of rehabilitation progress alongside established protocols. There will also be limitations when extending the approach to pathological movement patterns such as spasticity, abnormal synergies or compensatory motions, which will require pathology-specific data collection, temporal modeling and therapist-guided validation. Future work will focus on clinical studies to relate the proposed quantitative metrics to established clinical assessments and to facilitate translation from laboratory evaluation to real-world rehabilitation practice.

**Author Contributions:** Conceptualisation, M.T., M.C., Y.L. and L.H.; methodology, M.T.; software, M.T.; validation, M.T.; formal analysis, M.T. and M.C.; investigation, M.T. and Y.L.; resources, L.H.; data curation, M.T.; writing—original draft preparation, M.T. and M.C.; writing—review and editing, M.T., M.C., Y.L. and L.H.; visualization, M.T.; supervision, M.V. and L.H.; project administration, M.V. and L.H. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by The Podium Institute for Sports Medicine and Technology.

**Institutional Review Board Statement:** Ethical review and approval were waived for this study as the experiments involved only the authors and did not include human subjects or identifiable personal data.

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** Data are available upon request from the corresponding author.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

HaMeR	Hand mesh reconstruction
ROM	Range of motion
MPJAE	Mean per joint angle error
MPJAVE	Mean per joint angular velocity error
FMA	Fugl-Meyer Assessment
MMIB	Mesh-Mano interaction block
CNN	Convolutional neural network
GCN	Graph convolutional network
MCP	Metacarpophalangeal joint
PIP	Proximal interphalangeal joint
DIP	Distal interphalangeal joint
TM	Trapeziometacarpal joint
IP	Interphalangeal joint
PCK	Percentage of correct keypoints
APCK	Angle percentage of correct keypoints
AUC	Area under the curve
SPARC	Spectral arc length

## References

1. Meng, F.; Liu, C.; Li, Y.; Hao, H.; Li, Q.; Lyu, C.; Wang, Z.; Ge, G.; Yin, J.; Ji, X.; et al. Personalized and Safe Soft Glove for Rehabilitation Training. *Electronics* **2023**, *12*, 2531. [[CrossRef](#)]
2. Zhang, Y.; Orban, M.; Wu, Y.; Liu, C.; Wang, J.; Elsamanty, M.; Yang, H.; Guo, K. A review of soft robotics and soft rehabilitation gloves: Exploring alternative soft robots actuation techniques. *Int. J. Intell. Robot. Appl.* **2025**, *9*, 1368–1393. [[CrossRef](#)]
3. Zhang, T.; Zheng, K.; Tao, H.; Liu, J. A Soft Wearable Modular Assistive Glove Based on Novel Miniature Foldable Pouch Motor Unit. *Adv. Intell. Syst.* **2025**, *7*, 2500274. [[CrossRef](#)]
4. Proulx, C.E.; Beaulac, M.; David, M.; Deguire, C.; Haché, C.; Klug, F.; Kupnik, M.; Higgins, J.; Gagnon, D.H. Review of the effects of soft robotic gloves for activity-based rehabilitation in individuals with reduced hand function and manual dexterity following a neurological event. *J. Rehabil. Assist. Technol. Eng.* **2020**, *7*, 2055668320918130. [[CrossRef](#)] [[PubMed](#)]
5. Kottink, A.I.; Nikamp, C.D.; Bos, F.P.; Sluis, C.K.v.d.; Broek, M.v.d.; Onneweer, B.; Stolwijk-Swüste, J.M.; Brink, S.M.; Voet, N.B.; Rietman, J.S.; et al. Therapy effect on hand function after home use of a wearable assistive soft-robotic glove supporting grip strength. *PLoS ONE* **2024**, *19*, e0306713. [[CrossRef](#)] [[PubMed](#)]
6. Jiang, C.; Xiao, Y.; Wu, C.; Zhang, M.; Zheng, J.; Cao, Z.; Zhou, J.T. A2j-transformer: Anchor-to-joint transformer network for 3d interacting hand pose estimation from a single rgb image. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 18–22 June 2023; pp. 8846–8855.
7. Pavlakos, G.; Shan, D.; Radosavovic, I.; Kanazawa, A.; Fouhey, D.; Malik, J. Reconstructing hands in 3D with transformers. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–20 June 2024; pp. 9826–9836.
8. Hampali, S.; Rad, M.; Oberweger, M.; Lepetit, V. Honnotate: A method for 3D annotation of hand and object poses. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 3196–3206.
9. Zimmermann, C.; Ceylan, D.; Yang, J.; Russell, B.; Argus, M.; Brox, T. Freihand: A dataset for markerless capture of hand pose and shape from single rgb images. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 813–822.
10. Hampali, S.; Sarkar, S.D.; Rad, M.; Lepetit, V. Keypoint transformer: Solving joint identification in challenging hands and object interactions for accurate 3D pose estimation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 19–24 June 2022; pp. 11090–11100.
11. Ge, L.; Liang, H.; Yuan, J.; Thalmann, D. Robust 3D hand pose estimation in single depth images: From single-view cnn to multi-view cnns. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 3593–3601.
12. Oberweger, M.; Wohlhart, P.; Lepetit, V. Hands deep in deep learning for hand pose estimation. *arXiv* **2015**, arXiv:1502.06807.
13. Romero, J.; Kjellström, H.; Kragic, D. Hands in action: Real-time 3D reconstruction of hands in interaction with objects. In Proceedings of the 2010 IEEE International Conference on Robotics and Automation, Anchorage, AK, USA, 3–8 May 2010; pp. 458–463.
14. Zhang, B.; Wang, Y.; Deng, X.; Zhang, Y.; Tan, P.; Ma, C.; Wang, H. Interacting two-hand 3D pose and shape reconstruction from single color image. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 11–17 October 2021; pp. 11354–11363.
15. Gladstone, D.J.; Danells, C.J.; Black, S.E. The Fugl-Meyer assessment of motor recovery after stroke: A critical review of its measurement properties. *Neurorehabilit. Neural Repair* **2002**, *16*, 232–240. [[CrossRef](#)] [[PubMed](#)]
16. Tiboni, M.; Amici, C. Soft gloves: A review on recent developments in actuation, sensing, control and applications. *Actuators* **2022**, *11*, 232. [[CrossRef](#)]
17. Hazman, M.A.W.; Nordin, I.; Noh, F.H.M.; Khamis, N.; Razif, M.; Faudzi, A.A.; Hanif, A.S.M. IMU sensor-based data glove for finger joint measurement. *Indones. J. Electr. Eng. Comput. Sci.* **2020**, *20*, 82–88. [[CrossRef](#)]
18. Li, F.; Chen, J.; Ye, G.; Dong, S.; Gao, Z.; Zhou, Y. Soft robotic glove with sensing and force feedback for rehabilitation in virtual reality. *Biomimetics* **2023**, *8*, 425. [[CrossRef](#)] [[PubMed](#)]
19. Zimmermann, C.; Brox, T. Learning to estimate 3D hand pose from single rgb images. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 4903–4911.
20. Cai, Y.; Ge, L.; Cai, J.; Yuan, J. Weakly-supervised 3D hand pose estimation from monocular rgb images. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 666–682.
21. Romero, J.; Tzionas, D.; Black, M.J. Embodied hands: Modeling and capturing hands and bodies together. *arXiv* **2022**, arXiv:2201.02610. [[CrossRef](#)]
22. OpenMMLab. OpenMMLab Pose Estimation Toolbox and Benchmark, 2020. Available online: <https://github.com/open-mmlab/mmpose> (accessed on 22 December 2025).
23. Dong, H.; Chharia, A.; Gou, W.; Vicente Carrasco, F.; De la Torre, F.D. Hamba: Single-view 3D hand reconstruction with graph-guided bi-scanning mamba. *Adv. Neural Inf. Process. Syst.* **2024**, *37*, 2127–2160.

24. Martinez, J.; Hossain, R.; Romero, J.; Little, J.J. A simple yet effective baseline for 3D human pose estimation. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2640–2649.
25. Yang, Y.; Ramanan, D. Articulated human detection with flexible mixtures of parts. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *35*, 2878–2890. [[CrossRef](#)] [[PubMed](#)]
26. Balasubramanian, S.; Melendez-Calderon, A.; Burdet, E. A robust and sensitive metric for quantifying movement smoothness. *IEEE Trans. Biomed. Eng.* **2011**, *59*, 2126–2136. [[CrossRef](#)] [[PubMed](#)]
27. Zhao, Z.; Yang, L.; Sun, P.; Hui, P.; Yao, A. Analyzing the Synthetic-to-Real Domain Gap in 3D Hand Pose Estimation. In Proceedings of the Computer Vision and Pattern Recognition Conference, Seattle, WA, USA, 16–20 June 2025; pp. 12255–12265.
28. Roumaissa, B.; Mohamed Chaouki, B. Deep learning based on hand pose estimation methods: A systematic literature review. *Multimed. Tools Appl.* **2025**, *84*, 38121–38158. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.