POSE ESTIMATION FRAMEWORKS IN HEALTHCARE: A SYSTEMATIC REVIEW

Egga Asoka[1], Fathoni[2*], Hadipurnawan Satria[3], Indra Griha Tofik Isa[4]
Doctoral Program in Engineering Science, Sriwijaya University, Indonesia[1]
Computer Science, Sriwijaya University, Indonesia[2,3]
Prospective Technology of Electrical Engineering and Computer Science, National Chin-Yi University of Technology, Taiwan[4]
Informatics Management, Sriwijaya State Polytechnic, Indonesia[1,4]
fathoni@unsri.ac.id[*]

Received: 03 December 2025, Revised: 20 April 2026, Accepted: 22 April 2026
**Corresponding Author*

ABSTRACT
Human pose estimation has become increasingly important in healthcare applications such as fall detection, gait analysis, and rehabilitation monitoring. However, existing systematic reviews remain fragmented and largely descriptive, with limited comparative benchmarking and insufficient attention to clinical validation. This study addresses this gap by providing a structured comparison of major pose estimation frameworks in healthcare contexts. A systematic literature review was conducted using the PICOC framework and PRISMA guidelines. Studies published between 2020 and 2025 were retrieved from Scopus, Web of Science, IEEE Xplore, and PubMed based on predefined inclusion and exclusion criteria. Following screening and quality assessment, 41 studies were included in the final analysis. The results indicate that framework performance varies according to application requirements. OpenPose offers high anatomical precision but requires substantial computational resources, whereas MoveNet and MediaPipe enable real-time performance with lower latency, making them suitable for mobile and telehealth settings. Nevertheless, the evidence remains heterogeneous, with challenges related to occlusion, lighting variability, lack of standardized datasets, and limited real-world clinical validation. This study contributes by providing a theoretical synthesis and practical guidance for selecting appropriate pose estimation frameworks in healthcare applications.
Keywords : *Human Pose Estimation, Healthcare Applications, Computer Vision, Systematic Literature Review, Motion Analysis*

1. Introduction

Recent studies (Smith et al., 2023; Zhang et al., 2024; Zhang et al., 2023) have demonstrated the rapid advancement of pose estimation techniques for automated detection of key body parts in living organisms, supporting preventive healthcare strategies (Juraev et al., 2022) and treatment recommendation systems (Yu et al., 2024). This progress is largely driven by artificial intelligence, particularly in medical applications, where it enables automated posture analysis and anatomical feature identification (Munea et al., 2020). As a core computer vision technique, pose estimation facilitates the detection and tracking of human joint positions from visual data (Desai & Mewada, 2021; Wang et al., 2021), allowing non-invasive extraction of skeletal landmarks such as shoulders, elbows, and knees for quantitative motion analysis (Stenum et al., 2021).

The integration of deep learning with accessible hardware, such as smartphone cameras, has further accelerated the adoption of pose estimation in digital health systems. This technology now plays a critical role in tele-rehabilitation, remote patient monitoring, and elderly care, particularly in applications such as fall detection, gait analysis, and physical therapy evaluation. Several frameworks have emerged as dominant approaches in this domain. OpenPose provides highly detailed multi-person landmark detection but requires substantial computational resources (Cao et al., 2021). PoseNet has been successfully implemented in home-based rehabilitation systems to guide physiotherapy exercises (Hussein Abdullah et al., 2020), while MoveNet enables real-time performance for dynamic healthcare scenarios such as fall detection and mobility assessment (Xu et al., 2024). MediaPipe offers cross-platform deployment and stable multimodal tracking capabilities, making it suitable for scalable healthcare applications (Grishchenko & Bazarevsky, 2020; Grishchenko et al., 2021; Hii et al., 2024).

Despite these advancements, existing research remains fragmented and largely application-specific. Many studies focus on individual frameworks, isolated datasets, or controlled experimental environments, limiting the ability to generalize findings across diverse clinical settings (Chen et al., 2020; Chua et al., 2021; Ferrer-Mallol et al., 2022; Ge et al., 2025). Furthermore, real-world challenges such as occlusion, lighting variability, and patient heterogeneity significantly affect model performance but are insufficiently addressed in current studies (Yun et al., 2024). In addition, inconsistencies in dataset quality, evaluation metrics, and reporting standards hinder meaningful cross-study comparison (Gopal et al., 2023). While prior works have explored pose estimation within general computer vision or specific healthcare use-cases, they lack a structured and systematic comparative evaluation that links technical performance with clinical applicability.

Therefore, a critical research gap exists in the absence of a comprehensive and standardized comparative synthesis of pose estimation frameworks tailored to healthcare applications. Specifically, there is limited evidence integrating performance metrics, computational constraints, and clinical feasibility into a unified evaluation framework, resulting in uncertainty for clinicians and developers in selecting appropriate models for real-world deployment. To address this gap, this study conducts a Systematic Literature Review (SLR) of 41 peer-reviewed articles published between 2020 and 2025. This period is selected to capture the most recent advancements in deep learning-based pose estimation, particularly the emergence of lightweight and real-time models suitable for healthcare environments. The objectives of this study are: (i) to identify application trends and methodological developments in healthcare-based pose estimation, and (ii) to perform a comparative analysis of major frameworks OpenPose, PoseNet, MoveNet, and MediaPipe, based on performance, usability, and clinical suitability.

The contributions of this study are threefold. First, it provides a systematic and critical synthesis of pose estimation frameworks across diverse healthcare applications. Second, it bridges the gap between technical performance and clinical implementation by analyzing real-world constraints affecting model reliability. Third, it offers evidence-based guidance to support decision-making in selecting appropriate pose estimation frameworks for specific clinical use-cases.

2. Literature Review

2.1 Development of Pose Estimation in Healthcare

Human pose estimation has evolved significantly as a core component of computer vision, enabling the automated detection and tracking of human body keypoints from visual data. Early developments focused on general object detection and motion tracking; however, recent advancements in deep learning architectures have enabled more accurate and scalable pose estimation models applicable to healthcare contexts (Munea et al., 2020; Wang et al., 2021). In the healthcare domain, pose estimation has been increasingly utilized for non-invasive monitoring and analysis of human movement, supporting applications such as rehabilitation, gait analysis, and fall detection(Stenum et al., 2021). Several studies demonstrate its effectiveness in preventive and diagnostic scenarios, including early fall detection in elderly populations (W. Chen et al., 2020; Juraev et al., 2022) and automated treatment or intervention support (Yu et al., 2024). Additionally, pose estimation has been integrated with emerging technologies such as thermal imaging (Smith et al., 2023) and lightweight deep learning architectures (Wenwen Zhang et al., 2023), further expanding its applicability in diverse healthcare environments. Despite these advancements, the literature reveals that the adoption of pose estimation in healthcare is still evolving, with variations in methodological approaches, datasets, and evaluation strategies across studies. This heterogeneity creates challenges in establishing standardized benchmarks and limits the comparability of results.

2.2 Pose Estimation Frameworks and Their Characteristics

Several pose estimation frameworks have emerged as dominant approaches in healthcare-related applications, each with distinct architectural designs and performance characteristics. Among them, OpenPose, PoseNet, MoveNet, and MediaPipe are the most widely adopted. OpenPose, based on a bottom-up approach, is recognized for its ability to extract dense skeletal keypoints with high anatomical accuracy (Cao et al., 2021). Its effectiveness has been demonstrated in biomechanical analysis and clinical gait assessment, where high-resolution skeletal tracking is essential (Ge et al., 2025; P. Lee et al., 2022). However, its computational complexity limits its use in real-time or resource-constrained environments (Zhu et al., 2022).

PoseNet, in contrast, represents a lightweight top-down approach designed for fast inference and ease of deployment. It has been successfully applied in home-based rehabilitation systems (Chua et al., 2021; Hussein Abdullah et al., 2020). However, several studies highlight its limitations in terms of keypoint precision and robustness, particularly for clinically sensitive applications requiring fine-grained motion analysis (Divya & Peter, 2022). MoveNet introduces a more recent lightweight architecture optimized for real-time performance, particularly in dynamic scenarios such as fall detection and activity monitoring (Chang et al., 2025; Hu et al., 2024). Its ability to maintain high frame rates makes it suitable for time-sensitive healthcare applications. Nevertheless, its reduced keypoint representation limits its applicability in detailed clinical assessments.

MediaPipe offers a hybrid approach that balances accuracy and computational efficiency while supporting multimodal landmark detection (body, hands, and face) (Bazarevsky & Grishchenko, 2020). Its versatility has been demonstrated across a wide range of applications, including rehabilitation monitoring, gait analysis, and fine-motor assessment (Latreche et al., 2023; Venerito et al., 2025). However, its performance may degrade under challenging conditions such as occlusion and low lighting (Olikkal et al., 2024). Overall, the literature suggests that each framework presents a trade-off between accuracy, computational efficiency, and applicability, indicating that no single model is universally optimal for all healthcare scenarios.

2.3 Applications of Pose Estimation in Healthcare

Pose estimation has been applied across a wide spectrum of healthcare domains, reflecting its flexibility as a tool for motion analysis and clinical assessment. One of the most prominent applications is fall detection and fall-risk analysis, where pose estimation enables the identification of abnormal movement patterns and sudden posture changes (Chen et al., 2020; Sirikongtham & Nimkoompai, 2025; Zhu et al., 2022). These systems are particularly relevant for elderly care and remote monitoring scenarios. In rehabilitation and physical therapy, pose estimation has been used to monitor patient exercises, assess movement quality, and provide automated feedback in home-based settings (Chua et al., 2021; Dudekula et al., 2024; Tong et al., 2021). These approaches reduce reliance on wearable sensors and enable scalable tele-rehabilitation solutions. Gait analysis and biomechanical assessment represent another important domain, where pose estimation is used to extract joint angles and movement patterns for clinical evaluation (Ge et al., 2025; Jaber et al., 2025). Additionally, pose estimation has been applied in neurological and fine-motor assessment, including the evaluation of hand function and detection of neuromuscular impairments(Gu et al., 2022; Venerito et al., 2025).Emerging applications include ergonomics and workplace safety monitoring (Hu et al., 2024; B. Lee & Kim, 2024), human activity recognition in healthcare environments (Dutt et al., 2024), and assistive robotics systems that rely on gesture recognition (Jadeja et al., 2025). Although these applications demonstrate the versatility of pose estimation, most studies focus on specific use cases, limiting the ability to generalize findings across different healthcare contexts.

2.4 Challenges and Limitations in Existing Studies

Despite significant progress, several critical challenges remain in the current literature. First, there is a lack of standardized evaluation metrics. Studies employ diverse performance indicators such as accuracy, precision/recall, and latency, making cross-study comparison difficult (Gopal et al., 2023). Second, dataset variability poses a major limitation. Many studies rely on small-scale or custom datasets collected under controlled laboratory conditions, which do not reflect real-world clinical environments (Yun et al., 2024). This limitation reduces the generalizability and robustness of pose estimation models. Third, real-world implementation challenges such as occlusion, lighting variation, and patient diversity significantly affect model performance but are not consistently addressed across studies (Yun et al., 2024). Fourth, there is limited integration between technical performance and clinical validation. While many studies report high accuracy, fewer provide clinically validated outcomes or comparisons with gold-standard measurement tools (Ferrer-Mallol et al., 2022). Finally, inconsistencies in reporting methodological details and experimental setups further complicate reproducibility and synthesis of findings (Gopal et al., 2023).

2.5 Research Gap and Theoretical Perspective

From a theoretical perspective, pose estimation in healthcare can be viewed as an intersection of computer vision, machine learning, and clinical decision-support systems. However, existing literature lacks a unified framework that integrates these domains into a coherent evaluation model. Most prior studies emphasize either technical performance or application-specific implementation, without systematically linking algorithmic characteristics to clinical requirements. Additionally, existing reviews tend to focus on general computer vision developments rather than healthcare-specific comparative analysis. As a result, there is a clear research gap in the absence of a structured and comprehensive synthesis that compares major pose estimation frameworks using consistent evaluation criteria across diverse healthcare applications. This gap limits the ability of researchers and practitioners to make informed decisions regarding model selection and deployment in real-world clinical environments.

3. Methodology

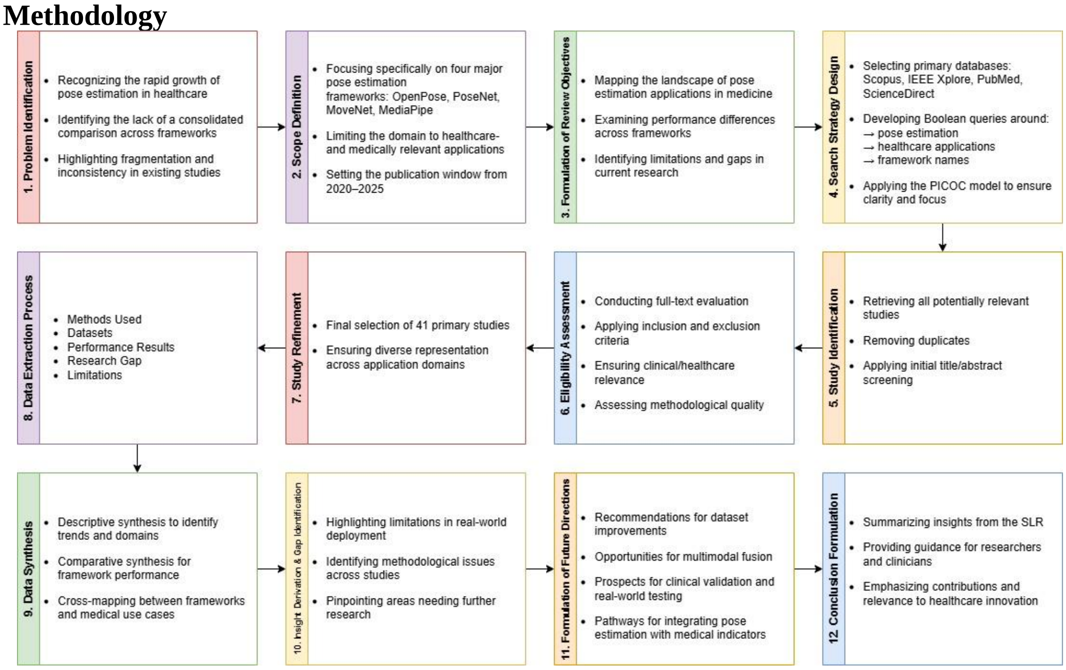

Fig. 1. Roadmap of The Systematic Review Structure

This study employed a SLR methodology structured according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Page et al., 2021) and refined using the PICOC framework (Kitchenham, 2007). The methodological process consisted of sequential stages beginning with problem identification, followed by scope definition, search strategy development, structured screening, and qualitative synthesis. Figure 1 illustrates the overall workflow of the SLR. An overview illustrating the sequential stages followed in conducting the SLR, including problem definition, scope refinement, search planning, study selection, data extraction, synthesis, and conclusion development.

3.1 Research Questions

The review was guided by two research questions designed to capture both the breadth and depth of pose estimation applications in healthcare:

RQ1: What are the trends and application domains of pose estimation in medical and healthcare settings?

RQ2: How do OpenPose, PoseNet, MoveNet, and MediaPipe compare in terms of performance, usability, and suitability for healthcare applications?

3.2 PICOC Framework

The PICOC framework (Population, Intervention, Comparison, Outcome, Context) was used to systematically define the scope and ensure alignment with the research objectives. The Population includes studies involving human motion, posture, or musculoskeletal analysis using computer vision. The Intervention focuses on pose estimation frameworks such as OpenPose, PoseNet, MoveNet, and MediaPipe. The Comparison involves cross-framework evaluation and benchmarking. The Outcome includes accuracy, latency, robustness, usability, and clinical relevance. The Context is limited to peer-reviewed studies published between 2020 and 2025 in healthcare-related domains. A formal summary of the PICOC criteria is presented in Table 1, which also served as the foundation for developing search strategies and inclusion–exclusion criteria.

Table 1 - PICOC Definition

Component	**Definition In This Review**
Population (P)	Studies involving human pose estimation for medical or healthcare purposes
Intervention (I)	Use of pose estimation frameworks (OpenPose, PoseNet, MoveNet, MediaPipe, or comparable CV-based methods)
Comparison (C)	Cross-framework comparison, performance evaluation, or benchmarking
Outcome (O)	Accuracy, robustness, latency, usability, application effectiveness in healthcare
Context (C)	Peer-reviewed studies in clinical, biomedical, rehabilitation, telehealth, or assistive-health domains

This PICOC structure forms the basis for the search keywords, inclusion/exclusion criteria, and data extraction categories.

3.3 Search Strategy

A structured literature search was conducted across four major academic databases: Scopus, Web of Science, IEEE Xplore, and PubMed. The search targeted peer-reviewed studies published between 2020 and 2025, representing the period when lightweight pose estimation frameworks became widely adopted in healthcare research. Keyword combinations related to computer vision, pose estimation, and medical applications were used, with Boolean operators applied to refine the search scope. The primary search query included variations of: ("pose estimation" OR "computer vision") AND ("healthcare" OR "medical") and ("OpenPose" OR "PoseNet" OR "MoveNet" OR "MediaPipe") AND ("clinical" OR "healthcare"). Reference lists of relevant studies were also manually screened to identify additional articles not captured in the database results.

3.4 Inclusion and Exclusion Criteria

To ensure that only studies directly relevant to the scope of this review were included, a set of predefined inclusion and exclusion criteria was established based on the PICOC framework. These criteria were applied consistently during full-text screening and guided the refinement of the final dataset. The inclusion criteria required that studies: (i) employed human pose estimation as a core component of the methodological pipeline, (ii) applied pose estimation within medical, clinical, rehabilitation, or broader health-monitoring contexts, (iii) used one or more of the following pose estimation frameworks, OpenPose, PoseNet, MoveNet, or MediaPipe or comparable computer-vision methods capable of skeletal keypoint extraction, and (iv) reported measurable outcomes such as accuracy, performance metrics, usability, or clinically relevant indicators. Furthermore, only peer-reviewed articles published in English between 2020 and 2025 were eligible. Studies were excluded if they lacked a healthcare focus, relied exclusively on wearable sensors without computer-vision components, provided theoretical descriptions without empirical experimentation, or were categorized as reviews, commentaries, or non-scientific reports. Articles were additionally excluded if the full text was inaccessible or if methodological details were insufficient to support reliable data extraction. These criteria ensured that the final selection consisted solely of methodologically rigorous studies with clear relevance to the evaluation of pose estimation frameworks in healthcare. The inclusion and exclusion criteria were defined based on the PICOC framework to ensure relevance and methodological rigor. These criteria were applied consistently during full-text screening. A summary of the inclusion and exclusion criteria is presented in Table 2.

Table 2 - Inclusion and Exclusion Criteria

Inclusion Criteria Studies were included if they:	**Exclusion Criteria** Studies were excluded if they:
1. Applied human pose estimation within medical, clinical, rehabilitation, ergonomics, or health-monitoring contexts.	1. Focused solely on sports, gaming, animation, or entertainment without any healthcare relevance.
2. Used at least one major framework (OpenPose, PoseNet, MoveNet, MediaPipe) or comparable computer-vision–based skeletal extraction methods.	2. Did not involve human subjects or human motion data.
3. Reported measurable outcomes such as accuracy, model performance, usability, or clinically relevant indicators.	3. Presented only conceptual or theoretical methods without empirical experiments.
4. Were peer-reviewed journal articles, peer-reviewed conference papers, or high-quality scholarly publications.	4. Were reviews, commentaries, opinion papers, or non-scientific reports.
5. Were published in English within 2020–2025 and accessible in full text.	5. Lacked sufficient methodological detail or did not provide extractable performance metrics.

3.5 Study Selection

A multi-stage study selection process was conducted following the PRISMA 2020 guidelines. The initial database search identified 73 records. After duplicate removal, 64 unique studies remained. These records underwent title and abstract screening to eliminate articles that lacked pose estimation, were unrelated to healthcare applications, or belonged to non-medical domains such as sports or entertainment, resulting in 60 articles eligible for full-text review.

During full-text screening, each study was evaluated using the predefined PICOC-based inclusion and exclusion criteria. Articles were included if they: (i) applied human pose estimation within a healthcare-related context, (ii) used or referenced OpenPose, PoseNet, MoveNet, or MediaPipe, (iii) reported measurable outcomes, and (iv) provided sufficient methodological detail for data extraction. Studies were excluded if the full text was unavailable, lacked human-motion data, or presented only conceptual methods without empirical validation. Following this screening process, a total of 41 studies met all criteria and were included in the final synthesis. Figure 2 presents the PRISMA flow diagram summarizing each stage of identification, screening, exclusion, and inclusion. The screening process was conducted systematically, and all inclusion decisions were re-evaluated to ensure consistency and minimize selection bias.

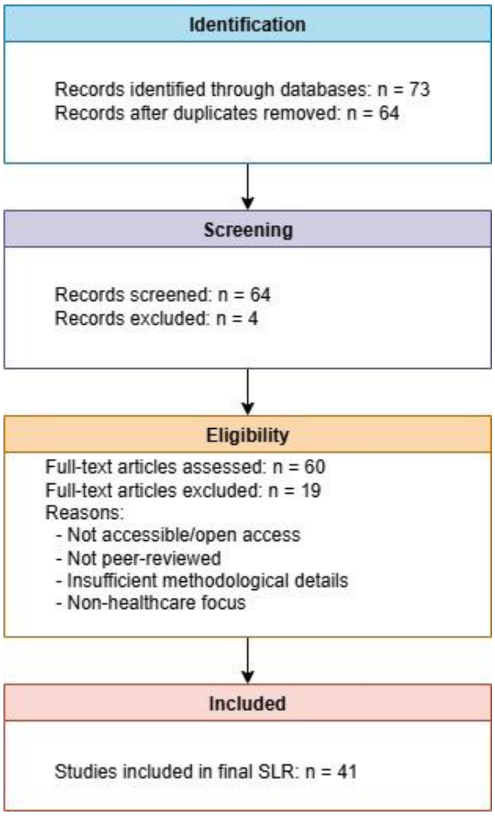

Fig. 2. PRISMA 2020 flow diagram of the study selection process

3.6 Data Extraction and Synthesis

Following the completion of the screening process, a structured data extraction procedure was conducted to ensure the systematic and reproducible collection of all relevant information from the 41 included studies. A predefined extraction matrix was developed based on the PICOC framework and refined iteratively during pilot testing. For each study, detailed information was recorded, including publication year, clinical application domain, study population, pose estimation framework used (OpenPose, PoseNet, MoveNet, or MediaPipe), dataset characteristics, camera configuration, evaluation metrics, computational setup, and key findings. Additional methodological attributes were also extracted, such as landmark accuracy, robustness under occlusion, temporal modeling techniques, and validation procedures against clinical or biomechanical references. Extracted data also captured the primary outcomes reported in each study, including diagnostic performance, usability, range-of-motion estimation accuracy, and fall-detection sensitivity.

To address the research questions, a hybrid data synthesis approach was applied, combining thematic analysis and comparative matrix analysis. For RQ1, a thematic categorization was performed to identify trends and classify application domains, such as fall detection, rehabilitation monitoring, gait analysis, and clinical assessment. For RQ2, a structured comparative analysis was conducted to evaluate pose estimation frameworks based on key criteria, including accuracy, computational efficiency, robustness, and clinical applicability. This approach enables both descriptive and critical synthesis, ensuring that cross-study comparisons are systematically derived and aligned with the research objectives.

3.7 Quality Assessment

To ensure that the synthesized evidence was derived from methodologically sound research, a quality assessment was performed on all included studies using criteria adapted from the Joanna Briggs Institute (JBI) checklist for analytical cross-sectional studies and SLR guidelines in medical computer vision. Each study was evaluated across seven dimensions: (1) clarity of research objectives, (2) methodological appropriateness, (3) dataset size and diversity,

(4) algorithmic transparency, (5) rigor of performance evaluation, (6) clinical relevance of outcomes, and (7) reporting completeness. Each criterion was scored on a three-point scale: 1 (low), 2 (moderate), and 3 (high). The total score ranged from 7 to 21 and was used to classify studies into three categories: high quality (≥16), moderate quality (11–15), and low quality (≤10). These quality scores were incorporated into the synthesis process to ensure that conclusions were primarily informed by studies with higher methodological rigor.

Overall, most studies demonstrated moderate to high quality, particularly in applications such as fall detection, gait analysis, and rehabilitation monitoring. However, several limitations were identified, including small sample sizes, reliance on controlled environments, and incomplete reporting of evaluation metrics. These limitations were explicitly considered during the interpretation of results.

4. Results

This section presents the findings derived from the qualitative synthesis of 41 peer-reviewed studies that met all inclusion criteria. To address the research questions, the results are organized into three major components: (i) an overview of the included studies, (ii) trends and application domains of pose estimation in healthcare (RQ1), and (iii) comparative insights across the four major pose estimation frameworks (RQ2).

4.1 Overview

The final set of 41 studies covered a diverse range of healthcare applications, including gait analysis, fall detection, rehabilitation monitoring, musculoskeletal assessment, and remote patient evaluation. Most studies used lightweight 2D pose estimation models, reflecting a shift toward real-time and low-resource clinical settings. OpenPose remained the most frequently used framework due to its multi-person tracking capability and well-established skeletal topology. PoseNet and MoveNet were increasingly adopted in studies emphasizing mobile or embedded deployment because of their speed and reduced computational overhead. MediaPipe also appeared in several mobile-health applications, but its detailed strengths are discussed in RQ1 to avoid redundancy. Across all studies, evaluation metrics commonly included landmark detection accuracy, precision/recall for clinical events (e.g., falls), temporal consistency in gait cycles, and computational performance indicators such as latency and frames per second (FPS). The diversity of study designs underscored the growing integration of pose estimation into both clinical and telehealth environments.

4.2 RQ1-Trends and Application Domains of Pose Estimation in Healthcare

The analysis of the 41 selected studies shows that pose estimation has been applied across a diverse set of healthcare-related domains. Rather than concentrating on a single clinical function, researchers have adopted pose estimation as a flexible tool that supports assessment, monitoring, and decision-making in multiple contexts. These applications range from movement evaluation to safety monitoring and functional motor analysis. The distribution across domains illustrates how pose estimation has transitioned from a laboratory concept into practical implementations within both clinical and non-clinical settings. Table 3 summarizes these application categories, the number of studies associated with each, and the representative frameworks commonly employed.

Table 3 - Application Domains of Pose Estimation in Healthcare (RQ1)

Application Domain	Number of Studies	Representative Frameworks	Description
Fall Detection & Fall Risk Analysis	14 studies	OpenPose, MoveNet, MediaPipe	Detecting falls, predicting fall risk, and analyzing fall-related posture dynamics in elderly or clinical populations
Rehabilitation & Physical Therapy Monitoring	10 studies	MediaPipe, OpenPose	Monitoring rehabilitation exercises, evaluating recovery progress, and assessing movement quality in clinical and home-based settings

Gait Analysis & Mobility Assessment	7 studies	OpenPose, MediaPipe, PoseNet	Assessing gait parameters, mobility impairments, and balance using video-based pose estimation
Exercise & Fitness Movement Tracking	6 studies	MediaPipe, OpenPose, PoseNet	Tracking exercise correctness, counting repetitions, evaluating posture accuracy, and estimating exercise difficulty
Neurological & Motor Function Assessment	5 studies	MediaPipe Hands, MediaPipe Pose	Evaluating fine-motor skills, hand function, nerve injury, dexterity progression, and neuromuscular performance
Human Activity Recognition (HAR) & Behavioral Monitoring	4 studies	MediaPipe, OpenPose	Recognizing human activities, abnormal behaviors, and occluded actions in healthcare or assisted-living environments
Ergonomics & Workplace Health/Safety	3 studies	MoveNet, MediaPipe	Assessing workplace risk, unsafe postures, and ergonomic hazards using pose-based risk scoring
Telehealth & Remote Patient Monitoring	3 studies	MediaPipe	Using smartphone or webcam-based CV for continuous remote monitoring of patient health and rehabilitation
Orthopedics & Biomechanical Angle Measurement	3 studies	OpenPose, MediaPipe	Measuring joint angles (e.g., HKA), range of motion, and orthopedic-related biomechanics
Robotics & Assistive Systems	2 studies	MediaPipe Hands	Gesture-based robotic control and imitation learning for healthcare assistance or rehabilitation robotics

As shown in Table 3, pose estimation has been applied across a wide range of healthcare domains, with fall detection and rehabilitation emerging as the most dominant areas due to their strong clinical relevance and compatibility with markerless motion analysis. Several studies utilized MoveNet and OpenPose for fall detection due to their handling of dynamic actions (Chen et al., 2020; Sirikongtham & Nimkoompai, 2025). Other significant applications include gait analysis, ergonomics, and neurological motor-function assessment, where skeletal keypoints provide valuable quantitative indicators for diagnosis, monitoring, and risk evaluation.

The systematic analysis reveals that fall detection and rehabilitation constitute the most dominant application domains, driven by the clinical need for non-invasive, continuous monitoring systems. Research by (Chen et al., 2020; Zhu et al., 2022) demonstrates that pose estimation frameworks, particularly OpenPose, provide the necessary skeletal fidelity to distinguish falls from daily activities in elderly populations. Furthermore, the integration of lightweight models like MoveNet has advanced real-time fall alert systems, as evidenced by (Chang et al., 2025) who highlighted its efficacy in dynamic action recognition.

In the domain of rehabilitation, studies have increasingly leveraged computer vision for home-based recovery monitoring(Chua et al., 2021; Tong et al., 2021) validated the use of PoseNet and fusion models to track stroke rehabilitation exercises, offering a cost-effective alternative to wearable sensors. Beyond these primary areas, significant applications have emerged in gait analysis and neurological assessment. For instance, (Ge et al., 2025) utilized OpenPose to measure Hip-Knee-Ankle (HKA) angles with radiological-level precision, while (Ferrer-Mallol et al., 2022) applied similar techniques for motor-function assessment in rare diseases. Emerging trends also indicate a shift towards specialized diagnostic support, such as the use of MediaPipe Hand landmarks as digital biomarkers for rheumatoid arthritis (Venerito et al., 2025). This expansion from general movement tracking to fine-motor analysis highlights the increasing maturity of pose estimation technologies in supporting clinical assessment, monitoring, and functional evaluation. Fall detection and rehabilitation appear to be the most mature application domains, whereas robotics assistance and emerging neurological biomarker applications remain at relatively early developmental stages.

4.3 RQ2-Comparison of OpenPose, PoseNet, MoveNet, and MediaPipe

The comparative analysis of the included studies reveals distinct performance trade-offs among the four frameworks, necessitating a strategic selection based on the specific clinical application's requirements for accuracy, latency, and hardware constraints. OpenPose retains its status as the gold standard for high-fidelity anatomical tracking. As evidenced in laboratory-based settings, its multi-person detection capabilities and detailed skeletal extraction—including facial and hand keypoints make it indispensable for precision-critical tasks such as clinical gait analysis and biomechanical measurements (Ge et al., 2025; P. Lee et al., 2022). However, this precision comes at a high computational cost. The literature consistently notes that OpenPose's heavy resource demands render it less viable for real-time mobile health (mHealth) or low-bandwidth telehealth deployments, restricting its primary utility to offline processing or high-performance clinical workstations (Zhu et al., 2022).

In contrast, PoseNet represents an earlier generation of lightweight models. While it offers faster inference speeds suitable for browser-based implementation, its utility in modern healthcare applications is increasingly limited by lower keypoint precision. Studies employing PoseNet primarily focus on general activity recognition or simple posture classification rather than fine-grained diagnostics (Chua et al., 2021; Divya & Peter, 2022). Consequently, for applications requiring detailed rehabilitation tracking or neurological assessment, researchers are shifting away from PoseNet in favor of more robust architectures. MoveNet has emerged as a superior alternative for scenarios demanding rapid temporal resolution. Its architecture, particularly the Thunder and Lightning variants, demonstrates exceptional robustness in dynamic environments, making it the preferred choice for fall detection systems and continuous physical activity monitoring (Chang et al., 2025; Hu et al., 2024). Although MoveNet excels in processing speed and motion handling, it exhibits limitations in scenarios involving significant occlusion or extreme body angles. Furthermore, its restriction to a 17-keypoint body model limits its applicability in studies requiring intricate hand or facial analysis.

Finally, MediaPipe is identified as the most versatile and widely adopted framework within the reviewed period. It achieves an optimal equilibrium between accuracy and computational efficiency, supporting multimodal landmarks (body, hands, and face) on edge devices. Numerous studies have validated MediaPipe against motion-capture baselines, confirming its reliability for diverse applications ranging from hand-function evaluation to remote rehabilitation monitoring (Ferrer-Mallol et al., 2022; Venerito et al., 2025). Despite reported challenges in low-light conditions and complex occlusions, MediaPipe's cross-platform compatibility positions it as the most practical solution for scalable, patient-centric digital health interventions.

4.4 Summary of Comparative Insights

The comparative evaluation of OpenPose, PoseNet, MoveNet, and MediaPipe across the 41 included studies demonstrates clear performance distinctions that reflect the inherent design of each framework. Although all four models provide skeletal keypoint extraction suitable for healthcare applications, they differ substantially in accuracy, computational efficiency, latency, robustness, and anatomical coverage. These variations directly shape their suitability for specific clinical tasks, including fall detection, gait and mobility assessment, rehabilitation monitoring, and fine-motor analysis. To consolidate these findings, Table 4 offers a structured comparison summarizing the key strengths, limitations, and optimal clinical use-cases of each framework, providing a comprehensive overview of how pose-estimation technologies are currently implemented in healthcare settings.

Table 4 - Summary of Comparative Insights Across Pose Estimation Frameworks

Aspect Compared	OpenPose	PoseNet	MoveNet	MediaPipe	Comparative Insight
Accuracy (General Clinical Tasks)	Frequently reported as high	Lower relative performance	High	High	OpenPose is frequently associated with higher precision, while MediaPipe offers competitive accuracy

					with better deployment efficiency.
Speed / Real-Time Capability	Lower real-time efficiency	High	Frequently reported as high	Frequently reported as high	MoveNet and MediaPipe are generally preferred for latency-sensitive real-time healthcare applications.
Computational Cost	High (GPU-intensive)	Very Low	Low	Low–Moderate	PoseNet and MoveNet are lightweight options, whereas OpenPose requires greater computational resources.
Keypoint Resolution	135+ full-body landmarks	17 keypoints	17 keypoints	33 body + 21 hand + 468 face	MediaPipe provides the richest multimodal landmark set, while MoveNet and PoseNet prioritize efficiency.
Robustness to Occlusion	High	Medium–Low	Medium	Medium	OpenPose is often reported as more robust under partial occlusion, although performance depends on scene conditions.
Suitability for Fall Detection	High	Medium	Strong suitability	High	MoveNet is frequently preferred for fall detection due to speed and stable real-time inference.
Suitability for Rehabilitation & ROM Tracking	High	Medium	High	Strong suitability	MediaPipe is widely used for rehabilitation and range-of-motion tracking because of stable joint-angle estimation and hand landmarks.
Suitability for Gait Analysis	Strong suitability	Low	Medium	High	OpenPose is commonly favored for biomechanical gait analysis due to finer skeletal mapping.
Suitability for Fine-Motor / Neurology (Hand/Finger)	Medium	Low	Low	Strong suitability	MediaPipe Hands is frequently adopted for hand, finger, and fine-motor assessment tasks.
Ease of Integration / Deployment	Medium	Very Easy	Easy	Easy	PoseNet is simple for lightweight deployment, while MediaPipe provides the strongest balance between usability and functionality.

Table 4 elucidates the comparative architectural priorities of the four frameworks, highlighting a fundamental dichotomy between computational complexity and clinical utility. The synthesis of the included studies reveals that no single framework is universally optimal; rather, selection is strictly dictated by the specific requirements of the healthcare application. OpenPose distinguishes itself as the benchmark for high-fidelity anatomical tracking. Its capability to generate dense skeletal representations, including detailed facial and hand keypoints, renders it indispensable for precision-oriented tasks such as biomechanical gait analysis and fine-motor assessment (Ge et al., 2025; P. Lee et al., 2022). Furthermore, its robustness against occlusion makes it ideal for controlled laboratory environments where accuracy is paramount. However, as noted by (Zhu et al., 2022), the substantial computational cost associated with its bottom-up approach restricts its feasibility for deployment on mobile devices or in resource-constrained telehealth scenarios.

Conversely, PoseNet prioritizes inference speed and deployment simplicity through its lightweight architecture. While this facilitates rapid prototyping on mobile and web platforms, the trade-off is a notable reduction in keypoint precision. Consequently, (Chua et al., 2021; Divya & Peter, 2022) suggest that PoseNet is often preferred for reserved for basic posture tracking or ambient monitoring where medical-grade accuracy is not a prerequisite. Its utility is increasingly confined to legacy systems or extreme low-power environments. In the domain of real-time event detection, MoveNet particularly its Thunder and Lightning variants, demonstrates superior temporal consistency. Its ability to maintain high frame rates during rapid movements has made it a preferred choice for fall detection systems and dynamic ergonomic evaluations (Chang et al., 2025; Hu et al., 2024). However, the limitation of its 17-keypoint model becomes apparent in complex diagnostic tasks; it lacks the granular hand and face tracking required for neurological assessments, thereby limiting its scope to gross motor analysis.

MediaPipe emerges as the most balanced and versatile solution within the current landscape. Its multimodal capabilities (integrating pose, face, and hand landmarks) combined with efficient cross-platform inference have driven its widespread adoption in rehabilitation monitoring and telehealth (Ferrer-Mallol et al., 2022; Venerito et al., 2025). Although (Olikkal et al., 2024) noted performance degradation under severe occlusion or poor lighting, MediaPipe is often preferred for scalable, patient-centric applications. Conclusion: Ultimately, the review indicates a clear transition in the field from experimental validation to practical implementation. The choice of framework is task-dependent: OpenPose for offline precision, MoveNet for real-time safety monitoring, MediaPipe for multimodal interaction, and PoseNet for legacy lightweight tasks. Future developments will likely gravitate towards hybrid systems or adaptive model selection strategies to bridge these architectural gaps.

4.5 Quantitative Synthesis of Framework Performance

The quantitative synthesis of the selected studies reveals distinct patterns in the adoption of pose estimation frameworks across healthcare applications. Among the 41 analyzed studies, OpenPose and MediaPipe emerge as the most frequently utilized frameworks, each appearing in approximately 41% of the studies, Percentages are based on frequency of framework appearance across included studies, and some studies employed multiple frameworks. This indicates that both frameworks represent dominant approaches, although they are applied in different clinical contexts. OpenPose is predominantly used in studies requiring high anatomical precision, such as gait analysis and biomechanical assessment, where detailed skeletal representation is critical. In contrast, MediaPipe is more commonly employed in applications emphasizing real-time performance and scalability, including rehabilitation monitoring and activity recognition. This distribution reflects a growing preference for lightweight and deployable solutions in modern healthcare systems.

MoveNet and PoseNet are less frequently adopted, each appearing in a smaller proportion of studies (approximately 7%). MoveNet is primarily associated with real-time monitoring scenarios, particularly fall detection, while PoseNet is generally used in simpler or low-resource applications. The relatively limited usage of these frameworks suggests a shift toward more advanced or balanced models. Overall, the findings demonstrate that framework selection is strongly influenced by application requirements rather than absolute performance superiority. High-precision clinical tasks tend to favor OpenPose, whereas real-time and scalable healthcare applications increasingly rely on MediaPipe and, to a lesser extent, MoveNet. This trend highlights the importance of balancing accuracy, computational efficiency, and deployment feasibility in healthcare pose estimation. Detailed study characteristics are provided in Appendix A.

5. Discussion

The findings of this systematic review reveal a rapidly evolving landscape in which computer vision-based pose estimation is increasingly integrated into healthcare research and clinical innovation. Across the 41 included studies, two central patterns emerge: the expanding diversity of healthcare application domains and the task-dependent performance characteristics of major pose estimation frameworks. Together, these findings demonstrate both the growing technical maturity of pose estimation systems and the practical considerations that will shape their broader clinical adoption.

(i) Expanding Clinical Application Domains

The reviewed studies demonstrate substantial growth in healthcare applications utilizing pose estimation. Fall detection and fall-risk prediction remain the most established domains, driven by the increasing need for continuous monitoring in aging populations and remote-care settings. Lightweight frameworks such as MoveNet and MediaPipe are frequently preferred in these contexts due to their real-time responsiveness and stable motion tracking. Rehabilitation and range-of-motion assessment also represent major areas of adoption. Accurate extraction of joint angles, movement trajectories, and posture alignment enables objective monitoring of therapeutic exercises and recovery progress. In these settings, OpenPose and MediaPipe are widely applied because of their detailed skeletal mapping and stable landmark detection. Beyond large-scale movement analysis, several studies highlight the growing role of pose estimation in neuromuscular and fine-motor assessment. MediaPipe Hands, for example, has enabled detailed hand-motion analysis and the exploration of digital biomarkers for neurological disorders. This expansion from general movement tracking to fine-motor analysis reflects the increasing maturity of pose estimation technologies in clinical assessment and patient monitoring.

(ii) Framework Suitability Is Strongly Task-Dependent

Framework is universally optimal for all healthcare applications. Instead, performance varies according to clinical objectives, environmental constraints, and computational requirements. OpenPose is frequently associated with high anatomical precision and stronger robustness under occlusion, making it suitable for biomechanics, gait analysis, and controlled clinical evaluations. However, its higher computational demands reduce practicality for mobile or low-resource environments. PoseNet remains attractive for lightweight deployment because of its simplicity and efficiency, although its lower landmark precision limits adoption in clinically sensitive tasks. MoveNet provides strong real-time speed and stable inference, making it highly suitable for fall detection, dynamic mobility monitoring, and continuous patient observation. However, its reduced keypoint topology may constrain detailed hand or facial assessment. MediaPipe demonstrates one of the most balanced performance profiles, combining speed, multimodal landmark coverage, and cross-platform usability. This versatility explains its broad adoption across rehabilitation, gait evaluation, ergonomic assessment, and fine-motor analysis.

(iii) Implementation Barriers and Practical Challenges

Despite promising progress, several barriers continue to limit seamless clinical integration. Environmental variability**,** including lighting conditions, camera quality, occlusion, and patient posture**,** can significantly affect model reliability, particularly in home-based or low-resource settings. In addition, heterogeneity in datasets, evaluation protocols, and movement definitions complicates cross-study comparison and prevents standardized benchmarking. Although many studies report favorable technical accuracy, fewer validate their outcomes against gold-standard clinical tools such as motion capture systems, EMG, or clinician-rated assessments. The predominance of positive pilot-scale studies may also suggest potential publication bias, whereby unsuccessful implementations or null findings are less frequently reported. This may lead to overestimation of current real-world readiness.

(iv) Ethical, Privacy, and Regulatory Considerations

The increasing use of camera-based systems in healthcare introduces important ethical and governance challenges. Because most frameworks rely on video or image data, concerns related to patient privacy, informed consent, secure data storage, and potential misuse must be carefully addressed. These concerns are especially relevant in telehealth and home-monitoring settings, where continuous observation may be perceived as intrusive. Future implementations should adopt privacy-by-design principles, anonymized skeletal representations where possible, and compliance with healthcare data-protection regulations. In addition, regulatory approval pathways for AI-enabled clinical tools remain an important requirement for large-scale deployment.

(v) Implications for Clinical Decision-Making

The findings suggest that framework selection should be guided by specific clinical requirements rather than a one-size-fits-all approach. High-precision applications, such as biomechanical gait analysis or detailed rehabilitation assessment, are more aligned with OpenPose. Real-time monitoring scenarios, including fall detection and home-based rehabilitation, are better supported by MoveNet and MediaPipe. Lightweight deployments may still benefit from PoseNet in selected low-complexity environments. These results reinforce the importance of context-aware decision-making that balances accuracy, computational efficiency, usability, and implementation feasibility.

(vi) Overall Outlook and Future Integration

Overall, pose estimation is transitioning from an experimental research tool into a practical component of intelligent healthcare systems. Its potential to enhance rehabilitation outcomes, enable continuous remote monitoring, improve patient engagement, and support data-driven assessment is increasingly evident. Future research should prioritize standardized benchmarking datasets, clinically annotated data, domain adaptation for real-world robustness, multimodal sensor fusion, and longitudinal validation studies. With continued collaboration between clinicians and AI researchers, pose estimation is well-positioned to become an increasingly reliable component of modern digital health ecosystems.

6. Limitations

Although this systematic review provides a structured and comprehensive synthesis of pose-estimation frameworks in healthcare, several limitations should be acknowledged. First, the review includes only open-access publications, which may exclude high-quality studies published in subscription-based journals or proprietary conference proceedings. This constraint introduces potential publication bias, as open-access articles may differ in methodological rigor, reporting quality, or experimental depth compared with closed-access sources. Second, substantial heterogeneity exists across the included studies, particularly in evaluation protocols, dataset characteristics, participant demographics, and performance metrics. Many studies rely on small samples, simulated environments, or custom datasets that do not reflect real clinical variability. The absence of standardized medical pose-estimation benchmarks further complicates cross-study comparison and may influence the perceived reliability of individual frameworks.

Third, the review focuses primarily on 2D pose estimation, reflecting the current dominance of RGB video in clinical and telehealth applications. As a result, studies employing 3D models, depth sensors, or multi-camera systems were excluded unless they reported comparable 2D outputs. This restriction may overlook emerging advancements in 3D biomechanical analysis that could enhance clinical accuracy for complex, multi-plane movements. Fourth, inconsistencies in reporting quantitative performance metrics limit the feasibility of conducting a formal meta-analysis. While some studies report keypoint detection accuracy, others rely on clinical indicators, qualitative assessments, or task-specific evaluation schemes. This variation reduces comparability and prevents the establishment of unified performance rankings across frameworks.

Furthermore, many studies were conducted under controlled laboratory conditions with ideal lighting, fixed camera positions, and minimal occlusion. These environments differ significantly from real-world settings, such as homes, hospitals, and assisted-living facilities, where patient behavior and environmental variability introduce greater uncertainty. Consequently, reported performance may overestimate robustness in practical deployments. Finally, the review excludes gray literature, preprints, and non-peer-reviewed work. Although this strengthens the scientific reliability of the synthesis, it may limit coverage of emerging innovations, particularly in rapidly evolving areas such as transformer-based pose estimation or multimodal hybrid systems. Despite these limitations, the review provides a valuable evidence base for understanding current capabilities, methodological trends, and future opportunities for pose estimation in healthcare applications.

7. Future Research Directions

The synthesis of current evidence highlights several strategic directions that can advance the role of pose estimation in healthcare as the field transitions from experimental prototypes to clinically deployable technologies. One of the most pressing needs is the development of large-scale, clinically annotated, and standardized datasets. Existing studies frequently rely on small, task-specific, or non-clinical datasets, which limit generalizability and complicate cross-study comparison. Establishing benchmark datasets that encompass diverse patient populations, medical conditions, and real-world clinical settings would enable more rigorous evaluation and improve methodological consistency across future research.

Another important direction involves expanding model capabilities beyond 2D pose estimation. Although 2D approaches dominate due to their simplicity and accessibility, they remain limited when analyzing multi-plane, depth-sensitive, or biomechanically complex movements. Future research should investigate hybrid 2D–3D pipelines, monocular depth inference, and lightweight multi-view reconstruction to capture more clinically accurate spatiotemporal dynamics without the need for high-cost motion capture systems. Temporal modeling also represents a critical frontier. Many medical tasks, such as gait-cycle segmentation, fall prediction, tremor analysis, and rehabilitation monitoring, are inherently sequential. Integrating temporal deep-learning architectures, including Transformers, LSTMs, GRUs, or temporal convolutional networks, may enhance motion continuity, reduce false positives, and improve stability in real-time applications. As pose estimation becomes increasingly embedded in clinical workflows, interpretability and explainability will become equally essential. Future work should focus on developing clinically meaningful Explainable AI (XAI) mechanisms that translate key point trajectories into interpretable indicators such as joint range of motion, gait-phase progression, postural asymmetry, or compensatory movement patterns. Achieving this will require close interdisciplinary collaboration to define clinically validated thresholds and decision criteria.

Improving robustness in uncontrolled environments remains another major challenge. While many studies report strong performance under laboratory conditions, real-world deployment, particularly in homes, hospitals, and long-term care facilities, introduces variability in lighting, occlusion, background clutter, camera placement, and patient behavior. Enhancing model generalization to these settings is crucial for reliable clinical use. Multi-modal data fusion also presents promising opportunities. Combining pose estimation with depth sensors, inertial measurement units, physiological monitoring, or audio cues may increase accuracy and provide richer clinical insights, enabling more comprehensive assessments of mobility, balance, exertion, and neuromuscular function.

As telemedicine and home-based monitoring continue to expand, optimization for edge devices becomes increasingly important. Future research should prioritize models that maintain clinical accuracy while supporting efficient on-device inference through techniques such as quantization, pruning, knowledge distillation, and hardware-specific acceleration. Finally, longitudinal and multi-session clinical validation remains limited in the current literature. Future studies should incorporate repeated measurements across extended periods and real patient populations to assess reliability, responsiveness, and clinical utility in real care pathways. Establishing long-term evidence will be critical for transitioning pose estimation from experimental research toward recognized and trusted clinical measurement tools.

8. Conclusion

This study addresses the lack of a structured healthcare-specific comparison of pose estimation frameworks by systematically reviewing 41 studies using the PICOC framework and PRISMA guidelines. The findings indicate that pose estimation has become an increasingly important component of digital health, particularly in fall detection, rehabilitation monitoring, gait analysis, and neuromuscular assessment. The results confirm that framework selection is strongly task-dependent. OpenPose is more suitable for high-precision biomechanical and clinical analysis, whereas MediaPipe and MoveNet provide more efficient and scalable solutions for real-time monitoring and telehealth applications. PoseNet remains relevant for lightweight deployment, although recent studies show growing preference for more advanced and versatile frameworks. This study contributes by linking technical performance with clinical applicability and providing evidence-based guidance for selecting appropriate frameworks according to healthcare requirements. However, several limitations should be acknowledged, including database restrictions, potential publication bias, and substantial heterogeneity in datasets, evaluation metrics, and experimental settings. Future research should prioritize standardized benchmarking datasets, improved real-world generalization through domain adaptation, multimodal sensor integration, and longitudinal clinical validation. Overall, pose estimation is evolving from a research-oriented technology into a practical component of intelligent healthcare systems with strong potential to improve clinical assessment, continuous monitoring, and data-driven healthcare innovation.

Acknowledgement

The authors would like to express their sincere gratitude to all researchers whose work contributed to this systematic review. The authors also extend appreciation to the developers of the OpenPose, PoseNet, MoveNet, and MediaPipe frameworks for enabling advancements in pose estimation research. Special thanks are given to colleagues and academic mentors who provided constructive feedback throughout the preparation of this manuscript. The support from institutional research and academic environments played a significant role in completing this study.

References

Abdulghani, M. M., Ghazal, M. T., & Salih, A. B. M. (2023). Discover human poses similarity and action recognition based on machine learning. *Bulletin of Electrical Engineering and Informatics*, *12*(3), 1570–1577. https://doi.org/10.11591/eei.v12i3.4930

Analia, R., Forster, A., Xie, S. Q., & Zhang, Z. (2025). Privacy-Preserving Approach for Early Detection of Long-Lie Incidents: A Pilot Study with Healthy Subjects. *Sensors*, *25*(12), 1–22. https://doi.org/10.3390/s25123836

Anggraini, I. T., Basuki, A., Funabiki, N., Lu, X., Fan, C. P., Hsu, Y. C., & Lin, C. H. (2020). A proposal of exercise and performance learning assistant system for self-practice at home. *Advances in Science, Technology and Engineering Systems*, *5*(5), 1196–1203. https://doi.org/10.25046/aj0505145

Bazarevsky, V., & Grishchenko, I. (2020). Google AI Blog: On-device, Real-time Body Pose Tracking with MediaPipe BlazePose.

Cao, Z., Hidalgo, G., Simon, T., Wei, S. E., & Sheikh, Y. (2021). OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. https://doi.org/10.1109/TPAMI.2019.2929257

Hii, C. S. T., Gan, K. B., You, H. W., & Zainal, N. (2024). Penilaian spatial dan temporal gaya berjalan dalam satah hadapan menggunakan MediaPipe pose. Jurnal Kejuruteraan, 36(4), 1561-1574.

Chang, Y. S., Zheng, Y. X., & Ku, Z. Y. (2025). Comparative Analysis of Lightweight OpenPose and MoveNet AI Models for Real-time Fall Detection and Alert Systems. *Sensors and Materials*, *37*(4 (4)), 1657–1679. https://doi.org/10.18494/SAM5420

Chen, K. Y., Shin, J., Hasan, M. A. M., Liaw, J. J., Yuichi, O., & Tomioka, Y. (2022). Fitness Movement Types and Completeness Detection Using a Transfer-Learning-Based Deep Neural Network. *Sensors*, *22*(15). https://doi.org/10.3390/s22155700

Chen, S., Kong, L., Tong, Z., Yamaguchi, Y., & Nakamura, M. (2025). Towards Sensor-Based Mobility Assessment for Older Adults: A Multimodal Framework Integrating PoseNet Gait Dynamics and InBody Composition. *Sensors*, *25*(18), 1–29. https://doi.org/10.3390/s25185878

Chen, W., Jiang, Z., Guo, H., & Ni, X. (2020). Fall Detection based on key points of human-

skeleton using openpose. *Symmetry*. https://doi.org/10.3390/SYM12050744
Chua, J., Ong, L. Y., & Leow, M. C. (2021). Telehealth using posenet-based system for in-home rehabilitation. *Future Internet*. https://doi.org/10.3390/fi13070173
Desai, M. M., & Mewada, H. K. (2021). Review on Human Pose Estimation & Human Body Joints Localization. *International Journal of Computing and Digital Systems*. https://doi.org/10.12785/ijcds/100181
Divya, R., & Peter, J. D. (2022). Smart healthcare system-a brain-like computing approach for analyzing the performance of detectron2 and PoseNet models for anomalous action detection in aged people with movement impairments. *Complex and Intelligent Systems*. https://doi.org/10.1007/s40747-021-00319-8
Du, Q., Bai, H., & Zhu, Z. (2023). Intelligent Evaluation Method of Human Cervical Vertebra Rehabilitation Based on Computer Vision. *Sensors*, *23*(8), 1–16. https://doi.org/10.3390/s23083825
Dudekula, K. V., Chalapathi, M. M. V., Kumar, Y. V. P., Prakash, K. P., Reddy, C. P., Gangishetty, D., … Singhu, R. (2024). Physiotherapy Assistance for Patients Using Human Pose Estimation With Raspberry Pi. *ASEAN Journal of Scientific and Technological Reports*, *27*(4). https://doi.org/10.55164/ajstr.v27i4.251096
Dutt, M., Goodwin, M., & Omlin, C. W. (2024). An Interpretable Deep Learning-Based Feature Reduction in Video-Based Human Activity Recognition. *IEEE Access*, *12*(December), 187947–187963. https://doi.org/10.1109/ACCESS.2024.3432776
Ferrer-Mallol, E., Matthews, C., Stoodley, M., Gaeta, A., George, E., Reuben, E., … Davies, E. H. (2022). Patient-led development of digital endpoints and the use of computer vision analysis in assessment of motor function in rare diseases. *Frontiers in Pharmacology*, *13*(September), 1–17. https://doi.org/10.3389/fphar.2022.916714
Francisco, J. A., & Rodrigues, P. S. (2023). Computer Vision Based on a Modular Neural Network for Automatic Assessment of Physical Therapy Rehabilitation Activities. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *31*, 2174–2183. https://doi.org/10.1109/TNSRE.2022.3226459
Ge, F., Wu, C., Ge, F., Xu, S., & Xiao, J. (2025). Reliability and validity of OpenPose for measuring HKA angle in dynamic walking videos in patients with knee osteoarthritis. *Scientific Reports*, *15*(1), 1–8. https://doi.org/10.1038/s41598-025-09627-2
Gopal, A., Torres, W. O., Winawer, I., Poole, S., Balan, A., Stuart, H. S., … Bove, R. (2023). “Self-care selfies”: Patient-uploaded videos capture meaningful changes in dexterity over 6 months. *Annals of Clinical and Translational Neurology*. https://doi.org/10.1002/acn3.51928
Grishchenko, I., & Bazarevsky, V. (2020). Mediapipe holistic—simultaneous face, hand and pose prediction, on device. *Retrieved June*.
Grishchenko, I., Bazarevsky, V., & Gabriel Bazavan, E. (2021). 3D Pose Detection with MediaPipe BlazePose GHUM and TensorFlow.js.
Gu, F., Fan, J., Cai, C., Wang, Z., Liu, X., Yang, J., & Zhu, Q. (2022). Automatic detection of abnormal hand gestures in patients with radial, ulnar, or median nerve injury using hand pose estimation. *Frontiers in Neurology*, *13*. https://doi.org/10.3389/fneur.2022.1052505
Han, K., Yang, Q., & Huang, Z. (2020). A two-stage fall recognition algorithm based on human posture features. *Sensors (Switzerland)*, *20*(23), 1–21. https://doi.org/10.3390/s20236966
Hu, X., Yan, H., Wan, C., Peng, L., & Qi, Y. (2024). Online Rapid Job Analysis and Evaluation Using Particle Swarm Optimized Random Forest. *IEEE Access*, *12*(October), 156420–156432. https://doi.org/10.1109/ACCESS.2024.3484671
Hussein Abdullah, B., Ahmed Abdullah, A., Osama Abdel Aal, R., Reda Noman, R., Khaled Darwish, A., & Al-Bahidi, W. (2020). Automatic Feedback For Physiotherapy Exercises Based On PoseNet. *Information Bulletin in Computers and Information*. https://doi.org/10.21608/fcihib.2020.116046
Jaber, R., Zahid, S. A., Rodacki, A. L. F., Qahwaji, R., & Buckley, J. G. (2025). Computer vision versus wearables assessment of the up-on-the-toes 30-second test. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging and Visualization*, *13*(1). https://doi.org/10.1080/21681163.2025.2485095

Jadeja, Y., Shafik, M., Wood, P., & Makkar, A. (2025). Enhancing Healthcare Assistance with a Self-Learning Robotics System: A Deep Imitation Learning-Based Solution. *Electronics (Switzerland)*, *14*(14). https://doi.org/10.3390/electronics14142823

Japhne, F., Janada, K., Theodorus, A., & Chowanda, A. (2024). Fitcam: detecting and counting repetitive exercises with deep learning. *Journal of Big Data*, *11*(1). https://doi.org/10.1186/s40537-024-00915-8

Juraev, S., Ghimire, A., Alikhanov, J., Kakani, V., & Kim, H. (2022). Exploring Human Pose Estimation and the Usage of Synthetic Data for Elderly Fall Detection in Real-World Surveillance. *IEEE Access*. https://doi.org/10.1109/ACCESS.2022.3203174

Kitchenham, B. (2007). Guidelines for performing systematic literature reviews in software engineering. *Technical Report, Ver. 2.3 EBSE Technical Report. EBSE*.

Kolosov, D., Kelefouras, V., Kourtessis, P., & Mporas, I. (2023). Contactless Camera-Based Heart Rate and Respiratory Rate.

Latreche, A., Kelaiaia, R., Chemori, A., & Kerboua, A. (2023). Reliability and validity analysis of MediaPipe-based measurement system for some human rehabilitation motions. *Measurement: Journal of the International Measurement Confederation*, *214*(April), 112826. https://doi.org/10.1016/j.measurement.2023.112826

Lee, B., & Kim, H. (2024). Evaluating the effects of safety incentives on worker safety behavior control through image-based activity classification. *Frontiers in Public Health*, *12*(August), 1–13. https://doi.org/10.3389/fpubh.2024.1430697

Lee, P., Chen, T. B., Liu, C. H., Wang, C. Y., Huang, G. H., & Lu, N. H. (2022). Identifying the Posture of Young Adults in Walking Videos by Using a Fusion Artificial Intelligent Method. *Biosensors*, *12*(5). https://doi.org/10.3390/bios12050295

Marchais, G., Arbane, M., Topilko, B., Brousseau, J., Brochot, C., Yaddaden, Y., … Maldague, X. (2025). SafeRespirator: Comprehensive Database for N95 Filtering Facepiece Respirator Leakage Detection Including Infrared, RGB Videos, and Quantitative Fit Testing. *IEEE Access*, *13*(December 2024), 5648–5659. https://doi.org/10.1109/ACCESS.2024.3522562

Munea, T. L., Jembre, Y. Z., Weldegebriel, H. T., Chen, L., Huang, C., & Yang, C. (2020). The Progress of Human Pose Estimation: A Survey and Taxonomy of Models Applied in 2D Human Pose Estimation. *IEEE Access*. https://doi.org/10.1109/ACCESS.2020.3010248

Olikkal, P., Pei, D., Karri, B. K., Satyanarayana, A., Kakoty, N. M., & Vinjamuri, R. (2024). Biomimetic learning of hand gestures in a humanoid robot. *Frontiers in Human Neuroscience*, *18*(July), 1–14. https://doi.org/10.3389/fnhum.2024.1391531

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*. https://doi.org/10.1136/bmj.n71

Pereira, B., Cunha, B., Viana, P., Lopes, M., Melo, A. S. C., & Sousa, A. S. P. (2024). A Machine Learning App for Monitoring Physical Therapy at Home. *Sensors*, *24*(1), 1–17. https://doi.org/10.3390/s24010158

Qin, H., Cai, M., & Qin, H. (2024). NABNet: Deep Learning-Based IoT Alert System for Detection of Abnormal Neck Behavior. *Sensors*, *24*(16). https://doi.org/10.3390/s24165379

Shih, C. L., Liu, J. Y., Anggraini, I. T., Xiao, Y., Funabiki, N., & Fan, C. P. (2024). A Yoga Pose Difficulty Level Estimation Method Using OpenPose for Self-Practice System to Yoga Beginners. *Information (Switzerland)*, *15*(12), 1–24. https://doi.org/10.3390/info15120789

Sim, K. S., Wong, S. W., Low, A., Yunus, A. P., & Lim, C. P. (2024). Real-Time Digital Assistance for Exercise: Exercise Tracking System with MediaPipe Angle Directive Rules. *International Journal on Informatics Visualization*, *8*(4), 2380–2387. https://doi.org/10.62527/joiv.8.4.2993

Sirikongtham, P., & Nimkoompai, A. (2025). A New Method for Real-Time Fall Detection Based on MediaPipe Pose Estimation and LSTM. *International Journal of Advanced Computer Science and Applications*, *16*(8), 107–113. https://doi.org/10.14569/IJACSA.2025.0160811

Smith, J., Loncomilla, P., & Ruiz-Del-Solar, J. (2023). Human Pose Estimation Using Thermal Images. *IEEE Access*. https://doi.org/10.1109/ACCESS.2023.3264714

Stenum, J., Cherry-Allen, K. M., Pyles, C. O., Reetzke, R. D., Vignos, M. F., & Roemmich, R.

T. (2021). Applications of pose estimation in human health and performance across the lifespan. *Sensors*. https://doi.org/10.3390/s21217315

Tong, Y., Zhang, Z., Chen, G., Li, X., Yan, H., Xu, M., & Lin, B. (2021). Testing the Feasibility of a Multi-Model Fusion Method for Monitoring the Action of Rehabilitating Stroke Patients in Care Management. *IEEE Access*. https://doi.org/10.1109/ACCESS.2021.3083668

Upadhyay, A., Basha, N. K., & Ananthakrishnan, B. (2023). Deep Learning-Based Yoga Posture Recognition Using the Y_PN-MSSD Model for Yoga Practitioners. *Healthcare (Switzerland)*, *11*(4). https://doi.org/10.3390/healthcare11040609

Venerito, V., Manigold, T., Capodiferro, M., Markham, D., Blanchard, M., Iannone, F., & Hügle, T. (2025). Single-camera motion capture of finger joint mobility as a digital biomarker for disease activity in rheumatoid arthritis. *Rheumatology Advances in Practice*, *9*(2). https://doi.org/10.1093/rap/rkae143

Wang, J., Tan, S., Zhen, X., Xu, S., Zheng, F., He, Z., & Shao, L. (2021). Deep 3D human pose estimation: A review. *Computer Vision and Image Understanding*. https://doi.org/10.1016/j.cviu.2021.103225

Xu, Q., Huang, G. H., Yu, M., Guo, Y., Tong, Y., Zhang, Z. Z., … Peter, J. D. (2024). Contactless Camera-Based Heart Rate and Respiratory Rate. *Sensors*, *12*(1), 1–17. https://doi.org/10.1049/ipr2.12208

Xu, Q., Huang, G., Yu, M., & Guo, Y. (2020). Fall prediction based on key points of human bones. *Physica A: Statistical Mechanics and Its Applications*. https://doi.org/10.1016/j.physa.2019.123205

Yu, H., Zhu, Z., Wang, C., Wang, J., & Liu, C. (2024). Kinematic-driven human-robot interaction system with deep learning for flexible acupuncture needling manipulations. *Biomedical Signal Processing and Control*, *92*. https://doi.org/10.1016/j.bspc.2024.106098

Yun, H., Nakamura, E., Kageyama, Y., Ishizawa, C., Kato, N., Igarashi, K., & Kawamoto, K. (2024). Action Recognition of Simulated Workplace With Occlusion Based on Interpolated Skeleton Data Using Openpose. *International Journal of Innovative Computing, Information and Control*, *20*(1), 231–247. https://doi.org/10.24507/ijicic.20.01.231

Zhang, J., Xie, J., Zhang, F., Gao, J., Yang, C., Song, C., … Zhang, Y. (2024). Greenhouse tomato detection and pose classification algorithm based on improved YOLOv5. *Computers and Electronics in Agriculture*. https://doi.org/10.1016/j.compag.2023.108519

Zhang, Weijia, Li, Y., Cai, S., Wang, Z., Cheng, X., Somjit, N., … Chen, F. (2024). Combined MediaPipe and YOLOv5 range of motion assessment system for spinal diseases and frozen shoulder. *Scientific Reports*, *14*(1), 1–14. https://doi.org/10.1038/s41598-024-66221-8

Zhang, Wenwen, Xu, Y., Bai, R., & Li, L. (2023). Animal Pose Estimation Algorithm Based on the Lightweight Stacked Hourglass Network. *IEEE Access*. https://doi.org/10.1109/ACCESS.2022.3231750

Zhu, N., Zhao, G., Zhang, X., & Jin, Z. (2022). Falling motion detection algorithm based on deep learning. *IET Image Processing*, *16*(11), 2845–2853. https://doi.org/10.1049/ipr2.12208

Appendix A. Descriptive Summary of Included Studies

Author/ Year	Title	Application Domain	Framework Used
(S. Chen, Kong, Tong, Yamaguchi, & Nakamura, 2025)	*Towards Sensor-Based Mobility Assessment for Older Adults: A Multimodal Framework Integrating PoseNet Gait Dynamics and InBody Composition*	Mobility assessment & early sarcopenia screening	PoseNet + InBody
(Chua et al., 2021)	*Telehealth Using PoseNet-Based System for In-Home Rehabilitation*	Home-based rehabilitation & telehealth	PoseNet; statistical metrics (MAD, MSE, MAPE)
(Tong et al., 2021)	*Testing the Feasibility of a Multi-Model Fusion Method for Monitoring the Action of Rehabilitating Stroke Patients in Care Management*	Stroke rehabilitation monitoring	Multi-model fusion (PDSN: OpenPose + GRU attention + 1D-CNN)
(Ge et al., 2025)	*Reliability and Validity of OpenPose for Measuring HKA Angle in Dynamic Walking Videos in Patients with Knee Osteoarthritis*	KOA gait assessment & alignment monitoring	OpenPose + smartphone video; radiographic comparison
(Xu et al., 2024)	*Patient-Led Development of Digital Endpoints and the Use of Computer Vision Analysis in Assessment of Motor Function in Rare Diseases*	Motor-function assessment (DMD)	OpenPose (25 keypoints) + mobile video
(Qin, Cai, & Qin, 2024)	*NABNet: Deep Learning-Based IoT Alert System for Detection of Abnormal Neck Behavior*	Abnormal neck posture detection & IoT monitoring	YOLOv5s + Lightweight OpenPose + Coordinate Attention + EKF
(P. Lee et al., 2022)	*Identifying the Posture of Young Adults in Walking Videos Using a Fusion Artificial Intelligent Method*	Gait analysis & posture identification	OpenPose + DJNP Fusion DL + ResNet101 + Naïve Bayes
(Japhne, Janada, Theodorus, & Chowanda, 2024)	*Fitcam: Detecting and Counting Repetitive Exercises with Deep Learning*	Exercise monitoring & fitness tracking	OpenPose + LSTM + MLP
(W. Chen et al., 2020)	*Fall Detection Based on Key Points of Human Skeleton Using OpenPose*	Elderly fall detection	OpenPose (2D keypoints)
(B. Lee & Kim, 2024)	*Evaluating the Effects of Safety Incentives on Worker Safety Behavior Control Through Image-Based Activity Classification*	Construction safety monitoring	OpenPose + ST-GCN
(Abdulghani, Ghazal, & Salih, 2023)	*Discover Human Poses Similarity and Action Recognition Based on Machine Learning*	Pose similarity matching; action recognition	OpenPose (18 keypoints), SVM (RBF), Cosine Distance
(Francisco & Rodrigues, 2023)	*Computer Vision Based on a Modular Neural Network for Automatic Assessment of Physical Therapy Rehabilitation Activities*	Physical therapy assessment	OpenPose (25 keypoints), Modular Neural Network
(Yun et al., 2024)	*Action Recognition of Simulated Workplace with Occlusion Based*	Workplace action recognition under occlusion	OpenPose; skeleton interpolation

	on Interpolated Skeleton Data Using OpenPose		(symmetric, linear, facial); LSTM
(Shih et al., 2024)	*A Yoga Pose Difficulty Level Estimation Method Using OpenPose for Self-Practice System to Yoga Beginners*	Yoga training & safety	OpenPose BODY25; difficulty metrics (support type, COG, inclination)
(Han, Yang, & Huang, 2020)	*A Two-Stage Fall Recognition Algorithm Based on Human Posture Features*	Fall detection (home monitoring)	OpenPose; SVM, KNN, DT, RF
(Anggraini et al., 2020)	*A Proposal of Exercise and Performance Learning Assistant System for Self-Practice at Home*	Home-based exercise learning	OpenPose (18 keypoints), Least Squares, Euclidean scoring
(Zhu et al., 2022)	*Falling Motion Detection Algorithm Based on Deep Learning*	Fall detection for elderly monitoring	YOLOv3 + OpenPose + DNN
(Xu, Huang, Yu, & Guo, 2020)	*Fall Prediction Based on Key Points of Human Bones*	Fall prediction	OpenPose; Inception-ResNet-v2 (transfer learning)
(Venerito et al., 2025)	*Single-Camera Motion Capture of Finger Joint Mobility as a Digital Biomarker for Disease Activity in Rheumatoid Arthritis*	RA digital biomarkers; telemedicine	MediaPipe (hand landmarks), smartphone camera
(Marchais et al., 2025)	*SafeRespirator: Comprehensive Database for N95 Filtering Facepiece Respirator Leakage Detection*	Respirator-leak detection; occupational health	IR camera, RGB camera, Bluetooth sync; MediaPipe (pilot face tracking)
(Latreche et al., 2023)	*Reliability and Validity Analysis of MediaPipe-Based Measurement System for Some Human Rehabilitation Motions*	Telerehabilitation; remote ROM assessment	MediaPipe (33 landmarks); goniometer; digital angle ruler; ICC & Bland–Altman analysis
(Sim, Wong, Low, Yunus, & Lim, 2024)	*Real-Time Digital Assistance for Exercise Tracking System with MediaPipe Angle Directive Rules*	Exercise tracking; fitness monitoring	MediaPipe (33 landmarks, 11 used); angle-based rules; MLP baseline & expand–shrink MLP
(Analia, Forster, Xie, & Zhang, 2025)	*Privacy-Preserving Approach for Early Detection of Long-Lie Incidents*	Fall monitoring; long-lie detection; home care	Thermal imaging; MediaPipe keypoints; handcrafted features; rule-based temporal monitoring; soft-voting ensemble; Raspberry Pi 5
(Dudekula et al., 2024)	*Physiotherapy Assistance for Patients Using Human Pose Estimation With Raspberry Pi*	Physiotherapy assistance; rehab monitoring	Raspberry Pi 4; MediaPipe; joint-angle tracking; audio–visual feedback
(Du, Bai, & Zhu, 2023)	*Intelligent Evaluation Method of Human Cervical Vertebra Rehabilitation*	Cervical spine rehabilitation; posture evaluation	MediaPipe (head pose, 3-DOF angles); YOLOv5 (privacy protection)

(Hii, C. S. T., Gan, K. B., You, H. W., & Zainal, N, 2024)	*Frontal Plane Spatial and Temporal Gait Assessment using MediaPipe Pose*	Gait analysis; frontal-plane gait monitoring	MediaPipe Pose; cubic spline interpolation; Butterworth filter
(K. Y. Chen et al., 2022)	*Fitness Movement Types and Completeness Detection Using a Transfer-Learning-Based Deep Neural Network*	Fitness tracking; exercise classification	YOLOv4; MediaPipe (33 keypoints); MLP (1D waveform)
(Jadeja et al., 2025)	*Enhancing Healthcare Assistance with a Self-Learning Robotics System*	Healthcare robotics; assistive HRI	YOLOv8; MediaPipe + LSTM; Deep Imitation Learning (BC + IRL)
(Kolosov, Kelefouras, Kourtessis, & Mporas, 2023)	*Contactless Camera-Based Heart and Respiratory Rate Monitoring Using AI on Hardware*	Remote vital sign monitoring; edge-AI	EVM; RPPG; BlazeFace; MediaPipe FaceMesh; CNN ROI extraction
(Jaber et al., 2025)	*Computer Vision versus Wearables Assessment of the Up-on-the-Toes 30-Second Test*	Functional ankle assessment; geriatric mobility	MediaPipe Pose Landmarker; IMU
(Weijia Zhang et al., 2024)	*Combined MediaPipe and YOLOv5 Range of Motion Assessment System*	Musculoskeletal assessment; spinal disorders	MediaPipe Pose; YOLOv5 + CBAM
(Olikkal et al., 2024)	*Biomimetic Learning of Hand Gestures in a Humanoid Robot*	Human–robot interaction; gesture imitation; assistive robotics	MediaPipe Hand Landmarks, Gaussian-based gesture encoding, PCA-based kinematic synergies
(Gu et al., 2022)	*Automatic Detection of Abnormal Hand Gestures in Patients with Radial, Ulnar, or Median Nerve Injury Using Hand Pose Estimation*	Neurological assessment; impairment screening; telemedicine	MediaPipe Hands (21 landmarks), rule-based classifier, Logistic Regression, SVM, Random Forest
(Dutt et al., 2024)	*An Interpretable Deep Learning-Based Feature Reduction in Video-Based Human Activity Recognition*	HAR; elderly care monitoring; clinical activity analysis	MediaPipe Pose, Interpretable Feature Reduction Function (IFRF), 1D-CNN
(Sirikongtham & Nimkoompai, 2025)	*A New Method for Real-Time Fall Detection Based on MediaPipe Pose Estimation and LSTM*	Fall detection; elderly monitoring	MediaPipe Pose (33 keypoints), LSTM sequence classifier
(Gopal et al., 2023)	*"Self-care selfies": Patient-uploaded videos capture meaningful changes in dexterity over 6 months*	Remote neurological monitoring; dexterity assessment	Smartphone-based pose/movement estimation tools
(Pereira et al., 2024)	*A Machine Learning App for Monitoring Physical Therapy at Home*	Home-based physical therapy; shoulder rehabilitation	MediaPipe Pose, cosine similarity, Dynamic Time Warping (DTW), QTM comparison
(Hu et al., 2024)	*Online Rapid Job Analysis and Evaluation Using Particle Swarm Optimized Random Forest*	Ergonomics; workplace safety; WMSD prevention	MoveNet Thunder, PSO-RF, REBA scoring

(Upadhyay, Basha, & Ananthakrishnan, 2023)	*Deep Learning-Based Yoga Posture Recognition Using the Y_PN-MSSD Model*	Yoga posture recognition; virtual coaching	PoseNet (TFLite MoveNet), MobileNet-SSD, hybrid Y_PN-MSSD model
(Chang et al., 2025)	*Comparative Analysis of Lightweight OpenPose and MoveNet AI Models for Real-time Fall Detection and Alert Systems*	Real-time fall detection; alert systems	Lightweight OpenPose, MoveNet, MLP classifier, LINE Notify
(Divya & Peter, 2022)	*Smart Healthcare System: A Brain-Like Computing Approach for Anomalous Action Detection in Aged People with Movement Impairments*	Anomalous action detection in older adults; intelligent healthcare monitoring	Detectron2 (object detection & instance segmentation), PoseNet (real-time pose estimation)