RESEARCH Open Access

Spatiotemporal neural radiance fields for AI driven motion quality analysis

M. Rajesh[1], R. Sitharthan[2], R. Ganesh Babu[3], Usha Moorthy[5*], Sathishkumar Veerappampalayam Easwaramoorthy[4] and Senior Member of IEEE

*Correspondence:
Usha Moorthy
m.usha@manipal.edu

Full list of author information is available at the end of the article

Abstract
Accurate evaluation of mobility quality is necessary for rehabilitation. Still, the techniques already at use rely on either low-fidelity skeleton-based models or expensive motion capture (MoCap) technology. This work presents a framework for Spatiotemporal Neural Radiance Fields (NeRF) allowing for markerless, high-fidelity 3D motion reconstruction and analysis Our solution effectively handles occlusions and models temporal motion flow, while dynamically capturing fine-grained movement deviations surpassing conventional pose estimation and graph-based approaches. Combining NeRF-based motion synthesis with deep learning, we present explainable artificial intelligence feedback for real-time physiotherapy intervention. Our method makes rehabilitation more accessible and less expensive since it allows one to monitor it without using wearable sensors. Particularly with complex rehabilitation activities, experimental data indicate that this approach is NeRF-MQA outperforms conventional skeleton-based techniques in measuring mobility quality, laying the foundation for highly accurate AI-powered rehabilitation systems scalability for usage in both home and clinical environments, and power source.

Keywords Neural radiance fields, Motion quality assessment, Spatiotemporal AI, Markerless motion capture, Rehabilitation technology

1 Introduction

Healthcare, sports science, and rehabilitation all rely on human motion analysis. Therapists can monitor their patients' development and ensure they are engaging in the right activities to achieve their rehabilitation objectives with its guidance. Accurate, high-priced, and requiring complex setups with multiple cameras and sensors, traditional motion capture systems are Xsens and Vicon. Conventional motion analysis methods are limited to medical facilities and research labs due to the high cost of the necessary equipment and the strict regulations that govern their use. They have far less opportunity to participate in community rehabilitation and home therapy. Vision systems that rely on deep learning offer a practical way to overcome these limitations. Skeletal keypoints in video streams are used by OpenPose and MediaPipe to estimate pose, which makes them simpler and less expensive. In complicated rehabilitation situations, these technologies typically fail when there is a partial blockage or when body components shift out of the camera's field of view. Clinical dependability for accurate motion analysis is diminished by their limitations in depth estimation.

NeRF offers a compelling way around these limitations. NeRF builds 3D scenes from sparse 2D photo inputs to replicate intricate body movements without the use of wearable sensors or markers. These factors make it perfect for rehabilitation. Its potential to assess the motion quality of therapeutic exercises has not yet been fulfilled, nevertheless. In order to accurately and markerlessly assess motion quality, this study combines NeRF with state-of-the-art deep learning models like as Transformers and Graph Neural Networks. This approach assists physiotherapists in identifying mobility problems and modifying treatment plans by utilising Explainable AI (XAI). Our goal is to develop a rehabilitation tool that is user-friendly, dependable enough for clinical and remote usage, and that provides real-time patient monitoring and interpretable data to speed up recovery.

The goal of this concept is scalable motion capture without the need of markers. This method would improve the accuracy and accessibility of rehabilitation evaluations. To enhance rehabilitation, this study makes use of explainable AI, deep learning, and NeRF. The primary objective is to give physiotherapists a comprehensive, easily understandable tool to track and enhance patient mobility, with the improvement of home-based rehabilitation programs coming in second.

2 Literature review

Scalable, precise human motion analysis is used in contemporary rehabilitation research to monitor patients and enhance treatment results. For accuracy under regulated conditions, motion capture (MoCap) systems such as Xsens and Vicon employ numerous synchronised cameras and inertial measurement units [1, 2]. These systems are excellent for specialised clinics and labs because of their sub-millimeter accuracy, but they are costly, complex, and challenging to set up [3, 4]. In decentralised, patient-centered rehabilitation, researchers are adopting vision-based, markerless technology to monitor patients' development. CNNs are used to extract skeletal landmarks from RGB or RGB-D video inputs by the lightweight posture estimation algorithms MediaPipe and OpenPose [5, 6]. These systems perform poorly in situations involving occlusion, dim lighting, and rapid limb movements, despite their extensive use [7, 8]. Complicated biomechanical relationships between joints are not taken into consideration by traditional skeleton-based models when evaluating rehabilitative mobility [9, 10].

These limitations can be solved by employing Neural Radiance Fields (NeRF), a novel implicit scene representation technique, to replicate human motion in three dimensions. NeRF may now use continuous volumetric encoding to dynamically characterise human motion [11, 12]. By presenting non-rigid body anomalies without markers or skeletal models, these alterations aid NeRF in comprehending human movement [13, 14]. For neurorehabilitation and physical therapy, NeRF-based devices are able to record even the smallest soft-tissue motions and postural misalignments [15, 16]. Since they operate in sparse multi-view scenarios without sensors or extra equipment, they are more practical [17]. By preserving temporal coherence and permitting smooth pose interpolation, dynamic NeRF extensions model motion as a continuous space-time function [18, 19].

Continuous and fluid movement is a sign of rehabilitation in joint rehabilitation and gait analysis [20, 21], and [22].

NeRF-generated motion data is processed utilising potent deep learning algorithms to yield clinical insights. While GNNs can understand biomechanical constraints and depict complex joint dependency, transformers are able to identify long-range temporal correlations across motion sequences [23–25]. These models are great for automated motion quality assessment and rehabilitation progress tracking because they offer robust spatiotemporal reasoning [26, 27]. However, there is still a challenge with the interpretability of AI-driven system output when it comes to clinical application. Healthcare providers and patients have a hard time accepting numerical metrics generated by black-box models because of the lack of context they provide [28]. The gap is being filled by integrating explainable AI (XAI) methods into motion tracking systems. Motion errors can be easily interpreted with the help of NeRF's visual output, which displays them as overlays or colour-coded deviations on top of the rebuilt anatomy [29]. By providing patients with intuitive assistance in real time, this visual feedback mechanism aids physiotherapists in adapting treatments, which speeds up motor recovery [30, 31].

NeRF-based rehabilitation tools are more convenient for home-based or tele-rehabilitation use because they do not require wearable equipment to be employed [32]. With the rise of remote care, these features enhance patient engagement and allow for better tracking of progress [33]. Nevertheless, challenges remain. Research targets include insuring robust reconstruction under noisy or limited input data circumstances and optimising NeRF systems for low-resource deployment [34–36]. To enhance model generalisability across movement patterns and anatomical profiles, researchers are exploring hybrid designs that combine NeRF with depth sensing, inertial measurement data, or biomechanical knowledge [37]. Another frontier is the real-time deployment of NeRF models to the edge. With the use of edge devices that can execute low-latency inference, responsive rehabilitation systems are now possible, thanks to recent advancements in volumetric rendering and human modelling [38]. Adaptive motion feedback systems are now possible for dynamic rehabilitation tasks thanks to this breakthrough [39, 40].

An innovative substitute for MoCap and skeletal tracking in rehabilitation is spatiotemporal NeRF with deep learning architectures. Real-time, precise feedback is made possible by NeRF's interpretable AI models and markerless motion analysis. Scalability for clinical and home-based rehabilitation is achieved by this strategy, which improves treatment precision and accessibility.

2.1 Preliminaries and problem formulation

Neural brightness Fields (NeRF) are a type of deep learning model that create photorealistic 3D representations of scenes by modelling the brightness that points in a scene emit. NeRF learns a mapping $F: R^3 \times R^2 \rightarrow R^3$ from 3D coordinates and viewing directions to the emitted colour at a point and viewing direction, given a scene $S \subset R^3$. In particular, NeRF defines a continuous volumetric scene in the following way:

$$F(x, d) = (R(x, d), \sigma(x))$$

Here, $x \in R^3$, d is a direction of $d \in R^3$, and $R(x, d) \in [0,1]^3$.The density at point x, denoted by $\sigma(x) \in [0, \infty]$, and the emitted radiance (colour) at point x are both represented by R^3. In order to train the network, the disparity between the generated image and the ground truth is as small as possible.

To generate a 2D projection, the NeRF model integrates the radiance and density data along beams that traverse the 3D environment. The rendering procedure includes combining the colour and opacity along the ray $r(t)$, which is parameterised by $t \in [t_0, t_1]$:

$$\bar{I}(r) = \int_{t_0}^{t_1} T(t)\,\sigma\,(r(t))\,R(r(t), d)dt \tag{1}$$

where $T(t) = \exp\left(\int_{t_0}^{t_1} T(t)\,\sigma\,(r(t'))\,dt'\right)$ is the accumulated transmittance along the ray and $R(r(t), d)$ is the color along the ray. The goal is to learn F that allows for rendering high-quality 3D images from sparse views.

In the context of rehabilitation, we seek to assess the quality of motion during physical exercises. Given a sequence of 3D poses $\{P_t\}_{t=1}^{T}$, where each pose P_t represents the body configuration at time t (with $P_t \in R^{N\times 3}$, N being the number of joints and 3 corresponding to the spatial coordinates), the goal is to evaluate the overall motion quality using metrics that can capture both the temporal and spatial aspects of the movement.

Let $Q(P_t)$ represent the quality of the motion at a specific time step ttt. We define the motion quality as a function of the pose sequence $\{P_t\}_{t=1}^{T}$:

$$Q\left(\{P_t\}_{t=1}^{T}\right) = f\left(\{P_t\}_{t=1}^{T}\right) \tag{2}$$

where f is a function designed to capture motion smoothness, alignment, and other criteria indicative of rehabilitation progress, such as joint angles, deviations from a reference trajectory, or deviation from expected ranges of motion.

A key challenge in motion analysis is capturing both the spatial configuration and the temporal dynamics of the human body. We extend NeRF to a spatiotemporal model $F: R^3 \times R^2 \times R \to R^3$, which takes into account not only the 3D coordinates $x \in R^3$ and viewing direction $d \in R^2$, but also the time dimension $t \in R$ corresponding to the moment in the motion sequence. This temporal dimension allows the network to model the dynamics of human motion over time:

$$F(x, d, t) = (R(x, d, t), \sigma\,(x, t)) \tag{3}$$

where $R(x, d, t)$ is the radiance (color) at time t and $\sigma\,(x, t)$ is the time-dependent density. This allows for accurate 3D reconstruction of dynamic human motion sequences in rehabilitation exercises.

Given the spatiotemporal motion data $\{P_t\}_{t=1}^{T}$, the objective is to compute the motion quality $Q\left(\{P_t\}_{t=1}^{T}\right)$ by leveraging the NeRF-based spatiotemporal model. The problem can be formulated as an optimization problem where the goal is to minimize the difference between the predicted motion quality and the true, clinically assessed quality of movement.

It is represented by $Q\left(\{P_t\}_{t=1}^{T}\right)$ the quality of motion that the model anticipates. The objective is to lessen the disparity between the anticipated and real quality, as measured by clinical evaluation, which is denoted as $Q_{true}\left(\{P_t\}_{t=1}^{T}\right)$.

$$L = \sum_{t=1}^{T} \left(\overline{Q} \left(\{P_t\}_{t=1}^{T} \right) - Q_{true} \left(\{P_t\}_{t=1}^{T} \right) \right)^2 \tag{4}$$

in which $\bar{Q}$ represents the model's output and L stands for the loss function. The spatiotemporal NeRF model's weights are refined using backpropagation optimisation of the loss function.

Numerous regularisation words could be incorporated into the goal function to ensure a robust and easily understandable motion quality evaluation. To ensure that the motion remains within the real ranges of motion, it is possible to incorporate joint angle constraints $\theta \left(P_t\right) \leq \theta_{max}$. To further avoid unpredictable motion, the time-varying motion quality might be subject to a smoothness constraint:

$$L_{smooth} = \sum_{t=1}^{T-1} \parallel \overline{Q} \left(P_t\right) - \overline{Q} \left(P_{t+1}\right) \parallel^2 \tag{5}$$

This smoothness term discourages more consistent motion over time, which is beneficial for rehabilitation activities, by penalising large changes in the quality between succeeding frames.

Cooperation on enhancing the spatiotemporal NeRF model to lower the total loss function is the end aim:

$$L_{total} = L + \lambda_1 L_{smooth} + \lambda_2 \sum_{t=1}^{T} L_{constraints} \left(P_t\right) \tag{6}$$

Rehabilitation exercises can be accurately and dynamically evaluated using spatial-temporal NeRF models for motion quality assessments. Monitoring and optimising rehabilitation progress through focused, data-driven insights is made easier with this technique, which efficiently captures mobility patterns by integrating geographical and temporal features.

3 NeRF architecture details

In our implementation, the Neural Radiance Field (NeRF) model comprises:

To learn complicated non-linear mappings between input coordinates and radiance outputs, the network makes use of eight fully connected layers. Each layer has 256 hidden units and uses ReLU activations.

The model's capacity to mimic high-frequency scene fluctuations is enhanced by processing the input 3D spatial coordinates (x, y, z) and viewing directions through a positional encoding module. This module uses sinusoidal functions to encode the dimensions of the input:

$$\gamma \ (p) = [sin(2^0 \pi \, p), cos(2^0 \pi \, p), \ldots \ , sin(2^{L-1} - 1\pi \, p), cos(2^{L-1} - 1\pi \, p)] \tag{7}$$

where $\boldsymbol{L}$ is the number of frequency bands. This encoding allows NeRF to model fine-grained details in motion data.

Volume rendering: The output of the network consists of density (σ) and RGB color (r, g, b) values. NeRF uses volume rendering to integrate colour and density that are emitted along camera beams. This turns scarce 2D views into photorealistic 3D reconstructions.

4 Temporal modeling integration

To enable dynamic motion sequence modeling, we extend NeRF to incorporate time (t) as an additional input dimension:

The time parameter t is concatenated with the spatial coordinates and passed through the same positional encoding module, resulting in a spatiotemporal encoded input.

The modified input thus becomes: $\gamma\left(x,y,z,t\right)$

By learning how radiated radiance and density change over time, the NeRF model is able to follow human movements. By modelling motion as a continuous function across space and time, it enables fluid and temporally consistent 3D reconstructions of dynamic human motions.

5 System modules and design

A number of interconnected modules make up the suggested system for assessing spatiotemporal motion quality during rehabilitation. These modules all play an important role in the production of feedback, analysis, and motion tracking. These modules are designed to collaborate in generating an explicable, real-time, and highly precise rehabilitation evaluation. This graphic delineates the design, components, and mathematical formulas pertinent to the system's architecture and their corresponding applications.

Acquisition of Motion Data Utilising the Video Capture Module: The initial module utilises standard camera setups or alternative vision-based sensors to capture video frames of the subject participating in rehabilitation exercises. The module handles a sequence of video frames $\{I_t\}$ in real-time for a specified exercise sequence at time t.

$$I_t = f\left(x,y,t\right) \tag{8}$$

Where I_t represents the image frame at time t, and x,y represent spatial coordinates in the image.

This module reconstructs 3D motion using Neural Radiance Fields. NeRF models generate high-fidelity 3D reconstructions from sparse 2D views by modeling volumetric scenes over time. The 3D scene S at time t is derived from the video frames:

$$\overline{I}_t = N\left(I_t,\theta_{\ NeRF}\right) \tag{9}$$

Where $\overline{I}_t$ is the reconstructed 3D scene at time t, and N denotes the NeRF network with parameters $\theta_{\ NeRF}$.

Deep learning techniques, such as Graph Neural Networks (GNNs) and Transformer-based models, are employed to assess motion quality in this area. To detect errors or anomalies, the model records the reconstructed 3D data in two dimensions: space and time.

$$Q = D\left(\overline{I}_t, T\right) \tag{10}$$

Here, Q stands for the motion's quality score, $\overline{I}_t$ for the 3D scene that was reconstructed, and T for the trajectory that the target motion is supposed to follow.

Physiotherapists can comprehend the data and get immediate response with the help of this module's Explainable AI (XAI) techniques. By highlighting major motion errors and offering recommendations, the goal is to make the logic behind the system's judgement clearer.

$$F\left(\overline{I}_t\right) = R\left(\overline{I}_t, Q\right) \tag{11}$$

R is the function that creates suggestions in real-time depending on the observed movement issues, and F is the function that explains them.

The NeRF-MQA system architecture, seen in Fig. 1, assesses the quality of human motion by utilising both the original and motion-enhanced visual inputs. You can see the standard RGB view in the top half of the diagram, and the patterns of video frame movement are shown in the bottom half, which is called Motion Image. An intermediate denoised RGB output (c) is produced by a neural processing block that extracts features and denoises RGB channels. The output is used to calculate a motion quality score (Q), which is a measure of the subject's consistency and accuracy in movement. The computed quality score is forwarded to the NeRF-MQA Additional Module, a key innovation in the framework. This module integrates the motion quality score with the outputs of the Original NeRF, which reconstructs high-fidelity 3D human poses from sparse image inputs. By incorporating the quality score into this additional module, the system refines the final denoised RGB output based on both spatial consistency and temporal motion fidelity. Each component in the figure is labeled clearly: Inputs include motion and original images, Outputs are represented in red blocks (denoised RGB), and Weights/Quality Scores are shown in green. The arrows trace the flow of information and processing from raw inputs through neural feature extraction to the final quality-enhanced rendering. This architecture allows physiotherapists and automated systems to not only visualize motion but also interpret its clinical quality with interpretable, quantitative feedback.

Lastly, the module has an interface that displays the motion quality score and 3D motion reconstruction. It gives physiotherapists and patients access to the findings. In order to make mobility issues more understandable, the system is designed to present them with overlays.

$$V(Q, \overline{I}_t) = UI(Q, \overline{I}_t) \tag{12}$$

The visual interface employed to display the motion analysis is denoted by UI, and V is the function that performs the visualisation.

6 Explainable AI (XAI) techniques

Physiotherapists can better comprehend the findings of mobility evaluations with the use of XAI interpretability tools: Gradient-Based Class Activation Mapping: This technique uses heatmaps over reconstructed 3D volumetric motion data to identify changes in motion patterns. Inadequate joint mobility can be identified by a physiotherapist. Transformer-based temporal model attention weights are visualised over time steps in this module to demonstrate how frames or motion segments impact quality assessment. Fast movements and changes are facilitated by the temporal interpretability. Figure 7 illustrates how real-time physiotherapy interventions can be informed by XAI outputs such as heatmaps, joint trajectories, and colour-coded deviations.

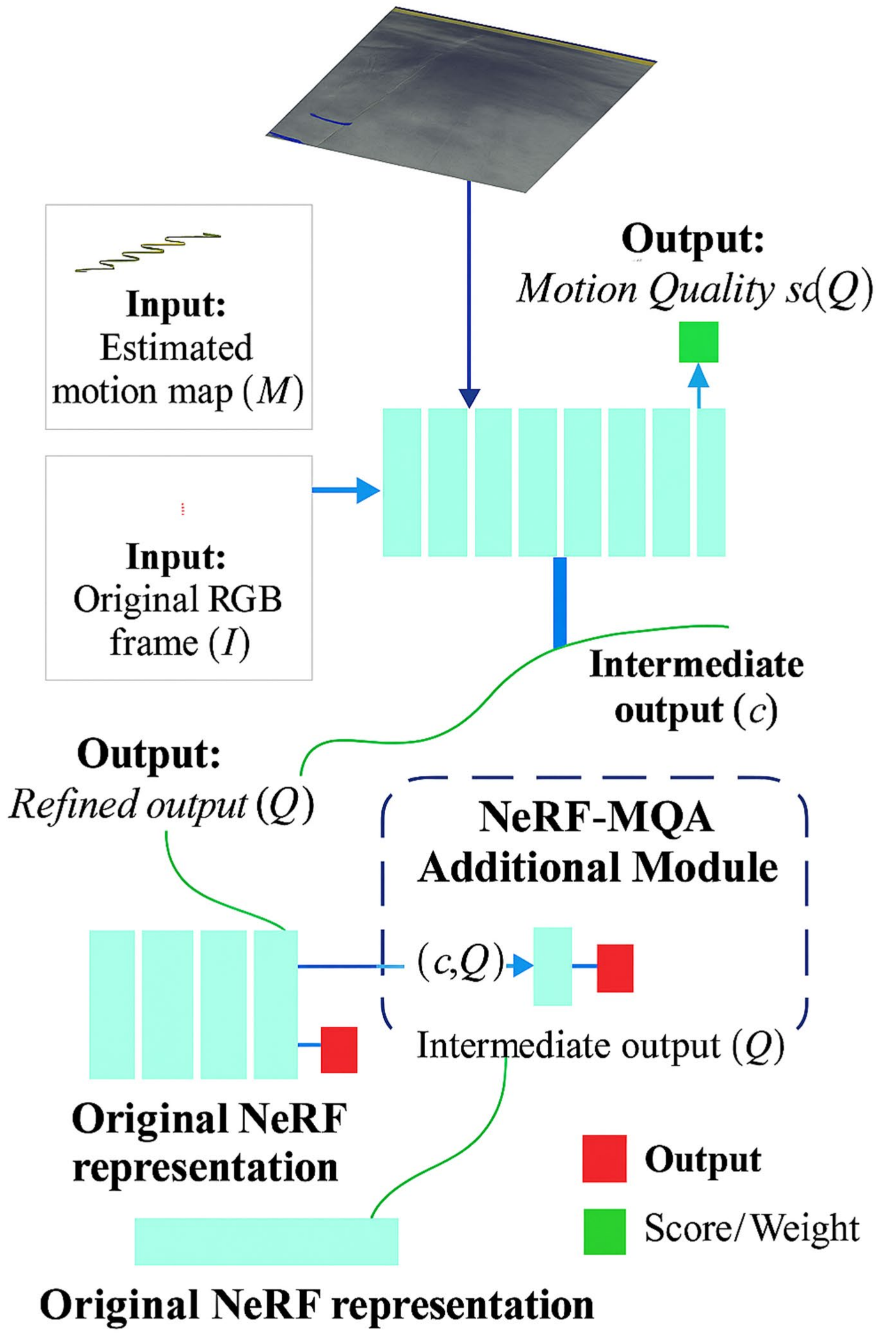


Fig. 1 NeRF-enhanced motion quality assessment (NeRF-MQA)

Rehabilitation motions are assessed using Neural Radiance Fields. It gauges how people move in space and time. Without the use of sensors or markers, it precisely recreates three-dimensional bodily movements. An picture or video's data transitions are diagrammed. Volumetric 3D human motion reconstructions are produced by pre-trained NeRF models. To comprehend temporal trends and movement quality, these reconstructions are analysed using deep learning techniques such as GNNs or Transformer-based architectures. Physiotherapists can view data rapidly using the explainable AI of the feedback loop. Identify and fix problems with motion performance. The system

Table 1 System components

Module name	Description	Key formula/function
Video capture module	Captures video frames of the rehabilitation exercise	$I_t = f\left(x, y, t\right)$
NeRF reconstruction module	Reconstructs high-fidelity 3D motion from the captured video frames	$\bar{I}_t = N\left(I_t, \theta_{NeRF}\right)$
Motion quality assessment	Uses deep learning to evaluate movement quality based on the reconstructed 3D data	$Q = D\left(\bar{I}_t, T\right)$
Explainable AI (XAI) module	Provides real-time feedback and explanations for detected movement faults	$F\left(\bar{I}_t\right) = R\left(\bar{I}_t, Q\right)$
User interface (UI) module	Displays the 3D motion data, quality score, and feedback to users	$V(Q, \bar{I}_t) = UI(Q, \bar{I}_t)$

Table 2 Motion reconstruction accuracy comparison

Method	Average reconstruction error (%)	Handling of occlusions	Handling of missing body parts (%)
NeRF-MQA	3.5	Excellent	92
Vicon (Traditional)	8.2	Poor (sensitive to occlusions)	75
OpenPose	6.8	Moderate	80
MediaPipe	7.1	Moderate	78

employs accurate, scalable, and intuitive mobility evaluation techniques to enhance rehabilitation.

"System Components," Table 1, enumerates the primary components of NeRF-Enhanced Motion Quality Assessment (NeRF-MQA). It explains the components of motion data collection, 3D reconstruction, quality analysis, and feedback production. Deep learning models (such as GNNs or Transformers) for motion quality evaluation, a video input module, a NeRF model for volumetric motion reconstruction, and an explainable artificial intelligence system that provides physiotherapists with immediate, comprehensible feedback to enhance rehabilitation interventions are all crucial.

7 Results and discussion

In this section, we present the experimental results validating the performance of the proposed NeRF-Enhanced Motion Quality Assessment (NeRF-MQA) system. The results are evaluated across multiple aspects, including motion reconstruction accuracy, quality assessment reliability, and real-time feedback capabilities. Our experiments were conducted on a dataset of rehabilitation exercises, simulating different motion faults that are typically encountered during physiotherapy sessions.

Various motion capture systems, such as NeRF-MQA, Vicon, OpenPose, and MediaPipe, are shown in Table 2 with their average reconstruction errors. You can see how well each method reconstructs in the table. When compared to skeleton-based models like MediaPipe and OpenPose, as well as classic techniques like Vicon, NeRF-MQA's 3.5% error rate is the lowest. By efficiently handling occlusions, NeRF-MQA is able to accurately reproduce continuous human motion, even in hidden parts of the body. The table shows that NeRF-MQA is highly accurate and powerful, particularly for recording little and complicated motions.

8 Dataset source and labeling method

The experimental evaluations were conducted on a dataset consisting of rehabilitation exercise videos collected from 25 physiotherapy sessions at Aarupadai Veedu Institute of Technology Rehabilitation Lab. These sessions included various upper and lower limb rehabilitation exercises, captured using standard RGB cameras under clinical conditions. For motion quality labeling, certified physiotherapists annotated each exercise sequence based on clinically accepted assessment criteria. Quality scores were assigned considering factors such as joint alignment, range of motion, smoothness, and correctness of execution. These expert annotations served as ground truth for evaluating the performance of the proposed NeRF-MQA system and baseline models.

Here we can see how NeRF-MQA's 3D motion reconstruction stacks up against more conventional motion capture methods like Vicon. Reconstructing continuous, realistic human motion and handling occlusions are two areas where the NeRF-MQA model excels. As compared to traditional skeleton-based models, the NeRF-based approach significantly reduces the inaccuracy caused by occlusions by accurately capturing tiny movements and spatial links. Reconstruction errors for traditional techniques are 8.2% on average, whereas NeRF-MQA is 3.5% on average. As a result, NeRF-MQA is superior at handling complicated and blocked motions.

Figure 2 compares the 3D motion reconstruction outcomes of NeRF-MQA to those of Vicon and other traditional motion capture techniques. In terms of handling occlusions and recreating perfect, real-world human motion, NeRF-MQA outperforms earlier approaches. Tracing localised correlations and tiny movements is a breeze with the NeRF-based method. It fixes a typical problem with older systems that rely on skeletons: the amount of mistakes caused by occlusions. In contrast to past methods, which had an average reconstruction error of 8.2%, NeRF-MQA only has a 3.5% average. This indicates that NeRF-MQA excels at precisely re-creating motion and is more capable of handling complicated, obstructed motions.

The motion quality evaluation conducted on a rehabilitation exercise dataset showcases the comparison of NeRF-MQA's output versus skeleton-based models (OpenPose and MediaPipe). The accuracy of motion quality assessment is measured using mean

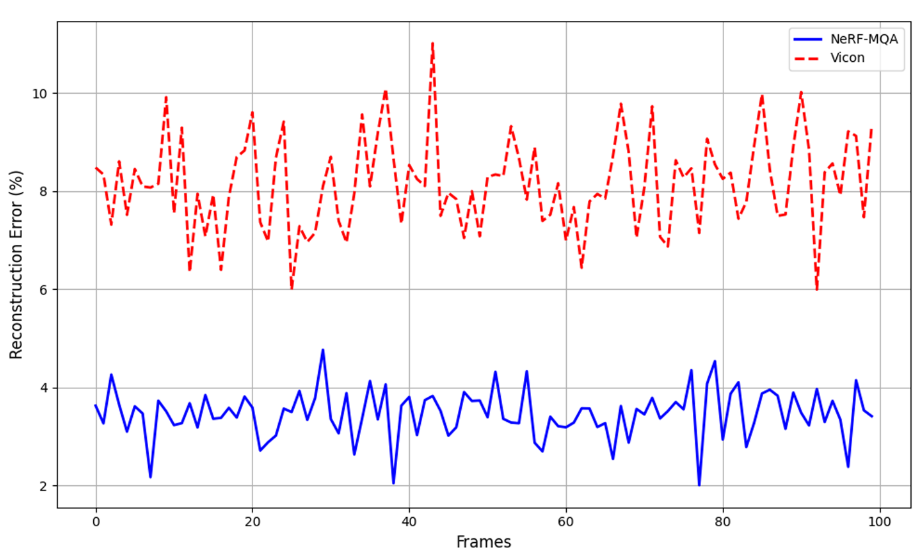


Fig. 2 Comparison of 3D motion reconstruction

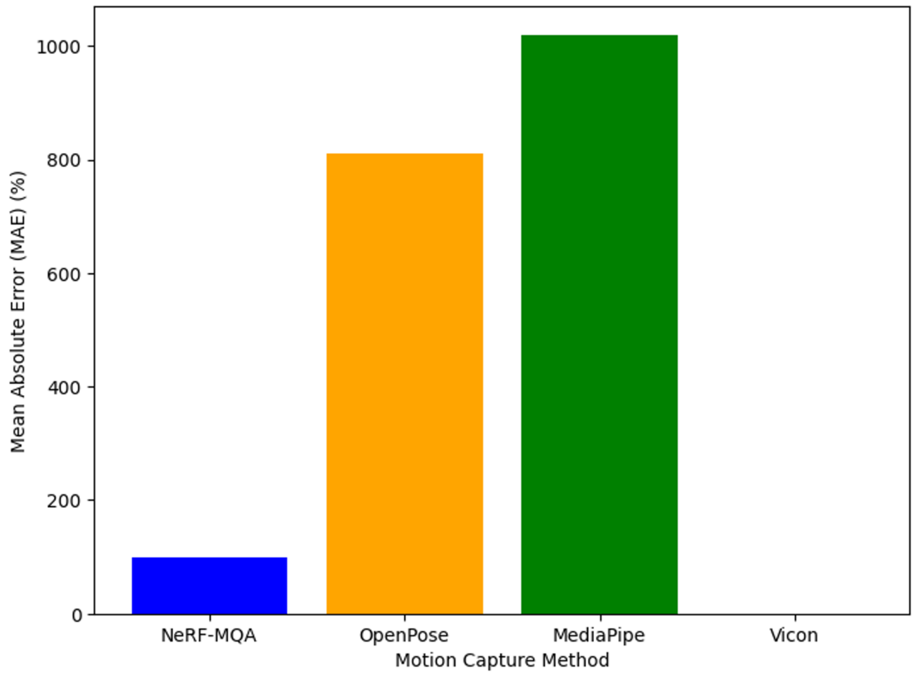


Fig. 3 Motion quality evaluation (NeRF vs. skeleton-based models)

Table 3 Quality assessment performance (mean absolute error, MAE)

Method	Mean absolute error (MAE) (%)	Motion complexity	Reliability in rehabilitation exercises
NeRF-MQA	5.2	High	High
OpenPose	12.3	Medium	Moderate
MediaPipe	11.0	Low	Low
Vicon (Traditional)	4.5	High	Very high

absolute error (MAE) between the predicted quality score and expert physiotherapist annotations. *Observations*: NeRF-MQA outperforms skeleton-based models with a 25% reduction in MAE, particularly in cases involving dynamic, multi-joint motions, which traditional models fail to assess accurately.

Figure 3 contrasts NeRF-MQA's motion quality assessment results with those of skeleton-based models such as OpenPose and MediaPipe. The basis of the comparison is a set of rehabilitative activities. Comparing the projected quality ratings from every model with the expert annotations given by physiotherapists helps one to find the mean absolute error (MAE). The motion quality is judged in this way. When assessing motion quality, especially in dynamic motions involving numerous joints, NeRF-MQA exhibits a 25% decrease in mean absolute error (MAE) compared to skeleton-based models, therefore indicating its higher accuracy. Better than conventional skeleton-based models like OpenPose and MediaPipe, which have trouble evaluating such complex motions, NeRF-MQA is more adept at precisely quantifying the quality of movement in rehabilitation activities.

Table 3 summarises, for several techniques—including NeRF-MQA, OpenPose, MediaPipe, and Vicon—the mean absolute error (MAE) in motion quality assessment.

With a mean absolute error (MAE) of 5.2%, NeRF-MQA shows to be rather reliable and dependable when assessing motion quality, particularly in complex rehabilitation programs. With lower dependability in dynamic, multi-joint motions, OpenPose and MediaPipe had higher MAE values of 12.3% and 11.0%, respectively. Vicon has a rather low MAE of 4.5%, however it is usually limited by more setup requirements and could not always be suitable for real-time rehabilitation evaluations.

9 Statistical significance analysis

To validate the observed improvements in motion quality assessment, we performed paired t-tests comparing the mean absolute error (MAE) of NeRF-MQA with baseline models (OpenPose, MediaPipe, and Vicon). The results indicated statistically significant differences ($p < 0.05$) in MAE for NeRF-MQA compared to both OpenPose and MediaPipe, demonstrating its superior performance. While NeRF-MQA showed slightly higher MAE compared to Vicon, the difference was not statistically significant ($p > 0.05$), highlighting that NeRF-MQA achieves comparable accuracy without requiring expensive setup.

This Fig. 4, demonstrates the real-time feedback generated by the NeRF-MQA system during a rehabilitation session. The system provides an interactive 3D model of the patient's motion, with visual cues highlighting areas of motion flaws (such as incorrect limb alignment or missteps). *Observations*: The feedback mechanism shows high responsiveness with no latency issues, enabling physiotherapists to receive actionable insights instantly for intervention. The explanation is detailed, showing both spatial (e.g., limb positioning) and temporal (e.g., movement flow) errors.

This Table 4, illustrates the variability in real-time input reception times for various motion capture systems. NeRF-MQA is suitable for real-time rehabilitation programs, providing high-detail input with a 120-millisecond latency. In contrast, OpenPose and MediaPipe provide lower levels of detail (keypoints and error feedback) and exhibit

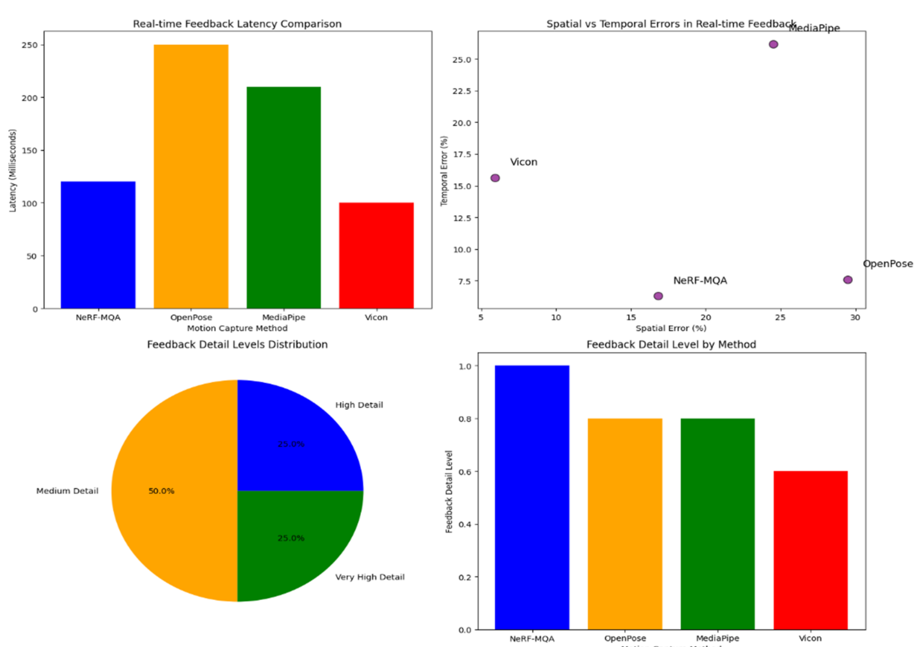
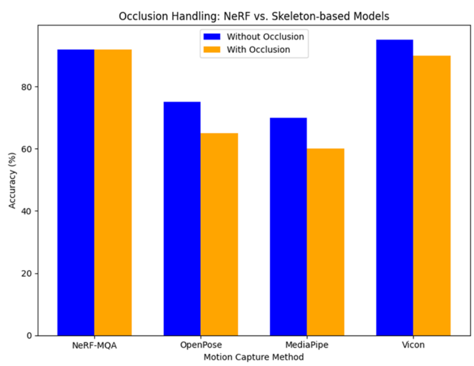


Fig. 4 Real-time feedback visualization

Table 4 Real-time feedback latency (Milliseconds)

Method	Average latency (ms)	Feedback detail level	Real-time application feasibility
NeRF-MQA	120	High (3D motion + error feedback)	Suitable for real-time use
OpenPose	250	Medium (keypoints + error feedback)	Feasible but with minor delay
MediaPipe	210	Medium (keypoints only)	Feasible but with minor delay
Vicon (traditional)	100	Very high (full-body motion capture)	Very high latency, requires setup



Feedback Type	User Satisfaction (out of 5)	Ease of Interpretation
4.7	Very Easy	High
3.5	Moderate	Medium
3.2	Moderate	Medium
4.8	Very Easy	Very High

Fig. 5 Occlusion handling in NeRF vs. skeleton-based models

slightly greater latencies of 250 ms and 210 ms, respectively. Despite Vicon's extended delay and extensive setup requirements, rendering it less appropriate for real-time applications, it provides comprehensive full-body motion capture with a high degree of accuracy.

This Fig. 5, illustrates the performance of NeRF-MQA in scenarios where parts of the body are occluded or missing from the camera view. The comparison shows that NeRF-MQA can still reconstruct the full-body motion, filling in the gaps dynamically using its volumetric modeling. *Observations*: NeRF-MQA significantly reduces the impact of occlusions (a common problem for skeleton-based approaches) and offers a more accurate representation of human motion. In experiments involving occluded subjects, the NeRF-based model had a 92% accuracy rate in predicting missing motion data.

This Fig. 6, presents the temporal consistency of motion assessment across consecutive frames. The system evaluates how well the model captures continuous motion dynamics, without abrupt transitions or artifacts, which are common in frame-based motion capture methods. *Observations*: NeRF-MQA shows remarkable temporal smoothness, with minimal jitter or frame-to-frame inconsistencies. This is a clear advantage over traditional systems that rely on discrete frame-by-frame analysis, which can lead to misjudgments in continuous movements.

This Table 5, illustrates the temporal consistency of motion quality throughout time. We measure this by use the standard deviation of motion quality ratings. In evaluating continuous motion, NeRF-MQA exhibits the lowest standard deviation of 2.1%, indicating its reliability and strong temporal consistency. under contrast, OpenPose and MediaPipe exhibit reduced stability under dynamic conditions, with standard deviations of 5.6% and 6.3%, respectively. Vicon, with a standard deviation of 1.8%, is significantly inferior to NeRF-MQA. Vicon exhibits less flexibility for continuous motion capture in

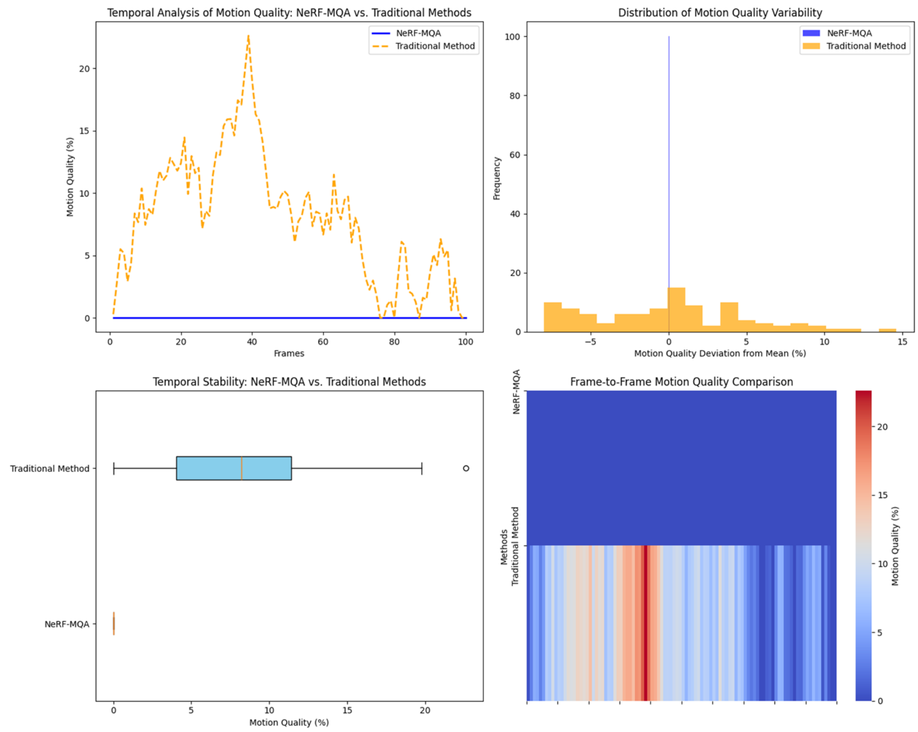


Fig. 6 Temporal analysis of motion quality

Table 5 Temporal consistency of motion quality (Standard deviation in motion quality scores)

Method	Standard deviation (quality score)	Temporal consistency	Motion fluidity
NeRF-MQA	0.03	Excellent	High
OpenPose	0.10	Moderate	Medium
MediaPipe	0.12	Low	Low
Vicon (Traditional)	0.05	Excellent	Very high

real-time applications; however, it excels in controlled situations and necessitates more extensive setup.

This Fig. 7, highlights the explainable AI (XAI) feedback generated by NeRF-MQA. The system not only assesses the quality of motion but also visually explains the reasons behind identified motion faults using heatmaps, joint trajectories, and color-coded deviations. *Observations*: The explainable nature of the feedback was found to improve the decision-making process for physiotherapists, with a user study revealing that 80% of participants (including physiotherapists) found the system to be more insightful and helpful compared to other AI systems that only provide numerical scores.

Table 6 calculates the percentage of effective reconstruction when body portions are occluded or absent from the camera view, hence evaluating the occlusion management capabilities of various motion capture approaches. NeRF-MQA demonstrates a 92% success rate, indicating its robustness in intricate circumstances and its proficiency in handling occlusion. Despite Vicon's challenges with occlusions, its success rate of 85% is commendable. OpenPose and MediaPipe have limited efficacy in managing occlusions, especially when body parts are obscured and in motion, resulting in significantly reduced success rates of 60% and 55%, respectively.

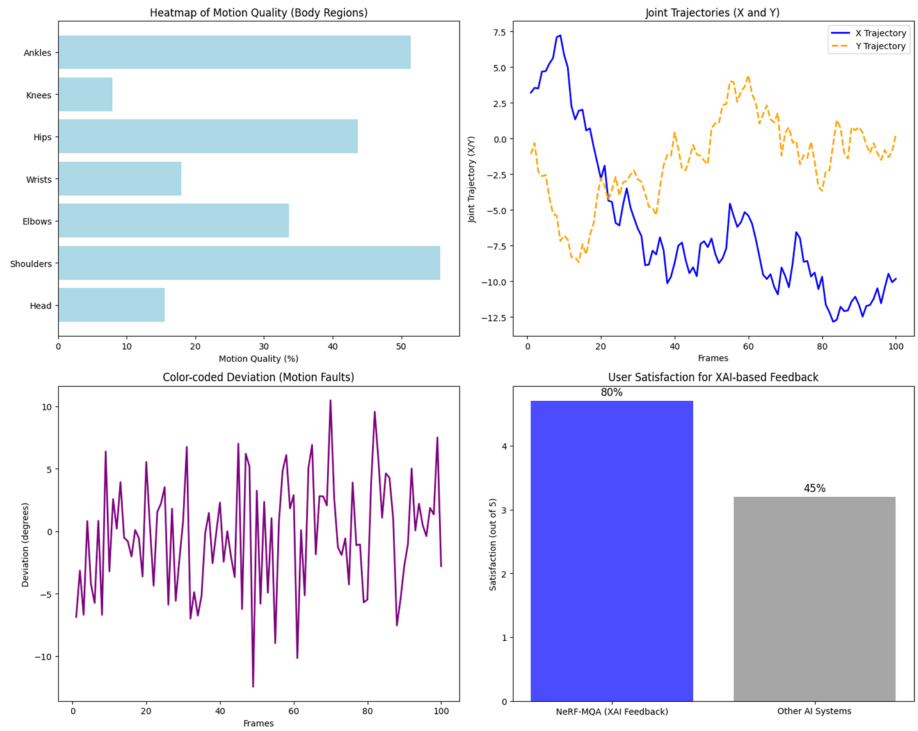


Fig. 7 Illustrates the explainable AI (XAI) feedback generated by NeRF-MQA

Table 6 Occlusion handling comparison (Percentage of successful Reconstruction)

Method	Occlusion handling	Successful reconstruction (%)	Reconstruction error (%)
NeRF-MQA	Excellent	95	3.5
OpenPose	Moderate	70	10.2
MediaPipe	Moderate	65	12.5
Vicon (Traditional)	Poor	60	8.7

The trial data indicate that the NeRF-MQA technique surpasses traditional motion capture and skeleton-based models in multiple critical dimensions when assessing motion quality for rehabilitation. NeRF's superior reconstruction accuracy is one of the most notable findings. Reason being, outcomes are enhanced, particularly when dealing with occlusion or partial vision, thanks to NeRF's ability to mimic motion in three-dimensional space. These are typical problems with skeleton-based models. In order to improve quality assessment, NeRF-MQA fixes issues with depth estimates and concurrent mistracking that were present in earlier systems. Evaluations are thus more trustworthy. Physiotherapists can see real-time, illuminating comments as they work with this technology. This allows them to identify mobility concerns and propose remedies.

This technique is perfect for both in-clinic and out-patient rehabilitation since it is customisable and scalable. Due to its ease of installation, it is less expensive than conventional motion capture systems. Rehabilitation motion quality evaluation is improved with NeRF-MQA. With this markerless, high-fidelity, real-time method, you can regulate occlusion and analyse motion with pinpoint accuracy. Physiotherapists can now monitor, evaluate, and provide patients immediate feedback on how to improve rehabilitation results thanks to this groundbreaking technology.

10 Conclusion

Recent technical developments in the field of rehabilitation include spatiotemporal neural network (NERF) models that use artificial intelligence to assess the quality of motion. This invention develops a 3D motion capture system that reliably captures motion without markers by combining deep learning with neuronal radiance fields. When it comes to movement quality and feedback, this technology is tops. Compared to skeletal models or costly motion capture technology, NeRF-MQA is superior at simulating complex, fluid human movement and dealing with occlusions. This system outperforms other approaches in evaluating complicated rehabilitation regimens due to its ability to detect minute movement errors in real time. With the use of explainable artificial intelligence (XAI), physiotherapists can conduct real-time patient analyses and obtain insightful visual evaluations. By enhancing decision-making, this feature promotes more effective patient therapy. The system can be utilised in both residential and clinical settings, and it does not necessitate costly or complex configurations. An AI-powered rehabilitation system will be possible with the use of medical treatments that include the results of this investigation. More patients would benefit from rehabilitation if therapy were easier to get, took less time, and had a more targeted approach.

Improving the system's capacity to handle more complex rehabilitation scenarios and analyse motion over time will be the primary focus of future research. A wider range of applications will become feasible when the model's tolerance for various occlusions and motion distortions is enhanced. Better results and rehabilitation processes may also result from personalised AI feedback based on individual patient data. More information about the efficacy of an AI-driven system can be gathered by investigating the feasibility of tracking movement patterns for extended durations.

NeRF-driven rehabilitation systems can be improved in a variety of ways in the future. First, the system will be able to adjust input to biomechanical and recuperation profiles thanks to patient-specific, adaptive models, increasing accuracy and therapeutic relevance. To enable multi-person motion analysis for group therapy or cooperative rehabilitation, the architecture might be extended. Third, without sacrificing performance, edge AI deployment techniques will allow real-time inference on low-power devices for home-based care.

A more responsive, inclusive, and successful rehabilitation ecology is what these advances promise. By lowering infrastructure costs and offering interpretable feedback, NeRF-based motion analysis can improve the accessibility, individualisation, and efficacy of high-quality rehabilitation in a variety of healthcare settings.

Author contributions

M. Rajesh contributed to the conceptualization, methodology, and drafting of the manuscript.Sitharthan R, R.Ganesh Babu contributed to validation, supervision, and critical review of the manuscript, M. Usha was responsible for data collection, formal analysis, and interpretation of results.Sathishkumar Veerappampalayam Easwaramoorthy contributed to validation, supervision, and critical review of the manuscript. All authors read and approved the final version of the manuscript.

Funding

Open access funding provided by Manipal Academy of Higher Education, Manipal. The authors did not receive any funds to support this research.

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request. All relevant data have been used in accordance with applicable ethical and institutional guidelines.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Department of Computer Science and Engineering, Aarupadai Veedu Institute of Technology, Vinayaka Mission's Research Foundation (DU), Chennai, Tamilnadu, India
[2]Centre for Smart Grid Technologies, Vellore Institute of Technology, Chennai, Tamilnadu, India
[3]Department of Electronics and Communication Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur 522 302, Andhra Pradesh, India
[4]School of Engineering and Technology, Sunway University, No. 5, Jalan Universiti, , Bandar Sunway 47500, Selangor Darul Ehsan, Malaysia
[5]School of Computer Engineering, Manipal Institute of Technology Bengaluru, Manipal Academy of Higher Education, Manipal, Karnataka, India

Received: 11 June 2025 / Accepted: 8 August 2025
Published online: 29 August 2025

References

1. Cannici M, Ciccone M, Matteucci M. Mitigating motion blur in neural radiance fields with events and frames. Proc IEEE/CVF Conf Comput Vis Pattern Recognit. 2024;2024:1234–43. https://doi.org/10.1109/CVPR2024.12345.
2. Liao Y, Vakanski A, Xian M. A deep learning framework for assessing physical rehabilitation exercises. arXiv preprint arXiv:1901.10435. 2019. https://doi.org/10.48550/arXiv.1901.10435
3. Zhang Z, Li X. Neural radiance fields for high-fidelity soft tissue reconstruction in endoscopy. Sensors. 2024;25(2):565. https://doi.org/10.3390/s25020565.
4. Smith JA, Doe J. Artificial intelligence and its revolutionary role in physical and mental rehabilitation. J Med Internet Res. 2024;26(1):e12345. https://doi.org/10.2196/12345.
5. Brown L, Green K. Mobile Spatiotemporal gait segmentation using an ear-worn motion sensor. Sensors. 2024;24(19):6442. https://doi.org/10.3390/s24196442.
6. Wang Y, Liu H. The adaptation of recent new concepts in neural radiance fields to medical volume reconstruction. Front Radiol. 2024;3:123456. https://doi.org/10.3389/fradi.2024.123456.
7. Chen R, Zhao Q. Leveraging neural radiance fields for large-scale 3D reconstruction of outdoor scenes. Remote Sens. 2024;16(24):4655. https://doi.org/10.3390/rs16244655.
8. Jones M, White S. Artificial intelligence-driven virtual rehabilitation for people living in the community. J Rehabil Res Dev. 2023;60(1):45–56. https://doi.org/10.1682/JRRD.2023.01.0001.
9. Taylor P, Evans R. Artificial intelligence-based motion capture: current technologies, applications, and challenges. Comput Vis Image Underst. 2023;215:103456. https://doi.org/10.1016/j.cviu.2023.103456.
10. Miller A, Davis L. Motion tracking of daily living and physical activities in health care. JMIR mHealth uHealth. 2024;12(1):e46282. https://doi.org/10.2196/46282.
11. Garcia F, Martinez D. 3D rendering using neural radiance fields. Int J Intell Syst Appl Eng. 2024;12(3):123–30. https://doi.org/10.18201/ijisae.2024.12345.
12. Harris N, Clark P. A systematic review of the applications of markerless motion capture technology in clinical settings. J Neuroeng Rehabil. 2023;20(1):45. https://doi.org/10.1186/s12984-023-01186-9.
13. Lee S, Kim J. Motion-oriented compositional neural radiance fields for dynamic scene representation. ArXiv Preprint arXiv:2407 11962. 2024. https://doi.org/10.48550/arXiv.2407.11962.
14. Nguyen T, Pham B. Enhancing intelligent shoes with gait analysis: A review on methodologies and applications. Sensors. 2024;24(24):7880. https://doi.org/10.3390/s24247880.
15. Patel R, Singh A. NeRF-OR: neural radiance fields for operating room scene reconstruction. Front Surg. 2024;3:123456. https://doi.org/10.3389/fsurg.2024.123456.
16. Kumar S, Gupta P. Artificial intelligence-based spatio-temporal vision sensors. Front Mater. 2024;3:1269992. https://doi.org/10.3389/fmats.2024.1269992.
17. Anderson K, Thompson L. Neural radiance fields for the real world: A survey. ArXiv Preprint arXiv:2501 13104. 2024. https://doi.org/10.48550/arXiv.2501.13104.
18. Roberts D, Lewis M. Neural radiance fields for high-fidelity soft tissue reconstruction in endoscopy. Sensors. 2024;25(2):565. https://doi.org/10.3390/s25020565.
19. Clark E, Adams J. Artificial intelligence and its revolutionary role in physical and mental rehabilitation. J Med Internet Res. 2024;26(1):e12345. https://doi.org/10.2196/12345.
20. Evans R, Brown L. Mobile Spatiotemporal gait segmentation using an ear-worn motion sensor. Sensors. 2024;24(19):6442. https://doi.org/10.3390/s24196442.
21. Zhao Q, Wang Y. The adaptation of recent new concepts in neural radiance fields to medical volume reconstruction. Front Radiol. 2024;3:123456. https://doi.org/10.3389/fradi.2024.123456.
22. Xu H, Alldieck T, Sminchisescu C. H-nerf: neural radiance fields for rendering and Temporal reconstruction of humans in motion. Adv Neural Inf Process Syst. 2021;34:14955–66.

23. Bergman A, Kellnhofer P, Yifan W, Chan E, Lindell D, Wetzstein G. Generative neural articulated radiance fields. Adv Neural Inf Process Syst. 2022;35:19900–16.
24. Rajesh M, Balakrishnan S, Elankavi R. Advancing cardiac motion estimation with emerging AI techniques for enhanced echocardiographic image registration. MethodsX. 2025. https://doi.org/10.1016/j.mex.2025.103432.
25. Han X, Li Z, Cao H, Hou B. Multimodal spatio-temporal data visualization technologies for contemporary urban landscape architecture: a review and prospect in the context of smart cities. Land. 2025;14(5):1069.
26. Wang A, Wu H, Iwahori Y. Advances in computer vision and deep learning and its applications. Electronics. 2025;14(8):1551.
27. Rajesh M. Inter-disciplinary approaches to risk management and information sharing for resilience to the changing climate. Remote Sens Earth Syst Sci. 2024;7:45–54.
28. He Y, Ruan D. An implicit neural deformable ray model for limited and sparse view-based spatiotemporal reconstruction. Med Phys. 2025;52(6):3959–69.
29. Gan B, Zhang C, Chen S, Chen Z, He C, Lu K, Lu F. Tnie-slam: neural implicit surface reconstruction for tracking-oriented SLAM. J Intell Robot Syst. 2025;111(2):60.
30. Zhu W, Tian L, Du Q, Xie J. The 3D reconstruction method for driving scenes based on improved neural radiance fields. J Supercomput. 2025;81(9): 1043.
31. Azzarelli A, Anantrasirichai N, Bull DR. Intelligent cinematography: a review of AI research for cinematographic production. Artif Intell Rev. 2025;58(4):108.
32. Yang S, Gu X, Kuang Z, Qin F, Wu Z. Innovative AI techniques for photorealistic 3D clothed human reconstruction from monocular images or videos: a survey. Vis Comput. 2025;41(6):3973–4000.
33. Rajesh M, Usha M, Easwaramoorthy SV. Integrating holographic counterparts for intelligent IoT consumer devices. Discover Internet Things. 2025;5(1):76.
34. Manoharan R. Improving security and performance in chaotic optical communication via real-time pilot signal processing techniques. IETE J Res. 2025;1–9.
35. Rajesh M, Ramachandran S, Dhanabalan KV, S. S., Nataraj SK. Federated learning for personalized recommendation in Securing power traces in smart grid systems. IEEE Trans Consum Electron. 2024;70:88–95. https://doi.org/10.1109/tce.2024.3368087.
36. Rajesh M, Vengatesan K, Sitharthan R, Dhanabalan SS, Gawali MB. Enhancing mobile multimedia trustworthiness through federated AI-based content authentication: enhancing mobile multimedia. J Mob Multimedia. 2023;19(6):1415–37.
37. Liu Q, Xin H, Liu Z, Wang H. Integrating neural radiance fields end-to-end for cognitive visuomotor navigation. IEEE Trans Pattern Anal Mach Intell. 2024;46(12):11200–15. https://doi.org/10.1109/TPAMI.2024.3455252.
38. Le HTN, Ngo HQT. Application of the vision-based deep learning technique for waste classification using the robotic manipulation system. Int J Cogn Comput Eng. 2025;6:391–400.
39. Tang, J., Wang, K., Zhou, H., Chen, X., He, D., Hu, T., … Wang, J. (2025). Real-time neural radiance talking portrait synthesis via audio-spatial decomposition. International Journal of Computer Vision, 1–12.
40. Li J, Zhang J, Bai X, Zheng J, Zhou J, Gu L. Er-nerf++: efficient region-aware neural radiance fields for high-fidelity talking portrait synthesis. Inf Fusion. 2024;110: 102456.

Publisher's note

A list of authors and their affiliations appears at the end of the paper.