Hybrid Deep Learning Healthcare AI Framework for Real-Time Human Pose Estimation and Remote Patient Monitoring to Support TKR Physiotherapy

Hisham AbouGrad[1], Manasa Yegamati[1], and Mimi Mather[1]

[1]Department of Computer Science and Digital Technologies, School of Architecture, Computing and Engineering, University of East London, University Way, Docklands, London, E16 2RD, UK

ABSTRACT

Total Knee Replacement (TKR) rehabilitation critically depends on precise physiotherapy exercise execution, and the rise of patient volumes and constrained clinical resources limit continuous supervision. This study presents an Artificial Intelligence (AI) framework for real-time assessment and feedback of TKR exercises using deep learning–based human pose estimation to empower remote rehabilitation. We investigate three architectures: a Dense Convolutional Neural Network (DCNN) incorporating frame decoupling for robust joint tracking; a pruned Generative Adversarial Network (Sparse GAN) optimized for computational efficiency; and a novel hybrid model that embeds the DCNN as a discriminator within the GAN model. A diverse dataset of over 10,000 annotated video clips, sourced from clinical environments and public repositories, was processed with OpenCV, and joint annotations were generated using OpenPose. Models were trained and evaluated on standard metrics (i.e. Precision, Recall, F1-score) alongside runtime and memory usage benchmarks. The hybrid architecture achieved the highest classification performance with 86.01% F1-score, which demonstrates the synergetic benefits of combining rich feature extraction with generative refinement, though it incurred elevated computational costs. The Sparse GAN provided faster inference suitable for mobile deployment, with only a marginal decrease in F1-score. The standalone DCNN provided a balance between accuracy and speed, but it did not match the hybrid's precision. These results highlight a fundamental trade-off between model complexity and real-time usability in AI-driven therapeutic monitoring. The hybrid model is optimal for clinical settings where accuracy is paramount, while the Sparse GAN provides a practical solution for resource-constrained and edge-based applications. Future work will explore model compression, hardware acceleration, and edge-computing strategies to further optimize performance. By demonstrating the viability of advanced pose estimation techniques in a physiotherapy context, this research contributes to the broader discourse on the use of AI in healthcare for scalable, autonomous rehabilitation tools across several medical and wellbeing domains.

Keywords: Healthcare AI Framework, Remote Rehabilitation, Physiotherapy Patient Monitoring MedTech, Total Knee Replacement, Dense Convolutional Neural Network, Hybrid DCNN+GAN model

INTRODUCTION

Postoperative rehabilitation is crucial for optimal recovery following TKR, where precise execution of therapeutic exercises significantly impacts functional outcomes (Cheong Chung et al., 2020; Kluzek et al., 2023). However, healthcare systems like the NHS face growing challenges in providing continuous supervision due to increasing patient volumes and limited clinical resources (Schoen, 2008).

This supervision gap raises concerns about improper exercise performance, which may delay recovery or cause complications among patients.

Recent advances in deep learning-based Human Pose Estimation (HPE) provide optimal solutions to enhance rehabilitation monitoring capabilities (Sigal, 2021; Sun et al., 2019). These computer vision systems can automatically analyse exercise movements from video by detecting deviations from correct form and providing real-time feedback. By integrating such medical technology (MedTech) into rehabilitation protocols, patients can perform guided exercises independently while maintaining clinical standards of care outside traditional healthcare settings.

This study built and validated a deep learning framework to automate physiotherapy monitoring by comparing three architectures: First, DCNN for robust feature extraction; Second, a pruned Sparse GAN optimized for efficiency; and third, a novel hybrid DCNN+GAN model combining both approaches (AbouGrad and Shabarshov, 2024). The models are trained using video datasets representing various environments and different conditions, including patient demographics, to ensure generalizability (LeCun et al., 1998; Saito et al., 2020).

Experimental results demonstrate that the hybrid model achieves superior classification accuracy, though with higher computational demands, revealing a critical trade-off between precision and operational efficiency. These findings contribute to developing scalable, intelligent monitoring systems capable of supporting both clinical and remote rehabilitation (PoseTrainer Project Team, 2023). By addressing real-world healthcare constraints, this research advances AI-driven solutions for accessible, patient-centered rehabilitation, and contributes to the evolving discourse on AI in healthcare by demonstrating how machine vision and pose estimation can be repurposed to support patient autonomy, improve recovery outcomes, and reduce clinical workloads. In doing so, it aligns with global efforts to create AI-driven patient-centric healthcare systems. The hybridisation approach also presents a promising pathway for developing high accuracy pose estimation models applicable in various domains, such as elderly care, sports injury recovery, and telemedicine.

Overall, this study underscores the transformative potential of deep learning in enabling more efficient, accessible, and accurate health interventions. By offering a technically feasible and socially impactful solution to physiotherapy supervision, it lays a foundation for future research in combining AI models to balance accuracy with usability. Further work should explore reducing runtime through algorithmic optimisation or hardware acceleration to enhance the model's suitability for widespread and real-time deployment.

RELATED WORK

Computer vision-based HPE has shown significant potential for revolutionizing rehabilitation assessment models (Ferraz et al., 2014; Přibyl et al., 2017). Foundational work using geometric approaches, such as Perspective-n-Point algorithms, established important mathematical principles for three-dimensional (3D) pose reconstruction, while modern improvements seek real-world robustness and computational efficiency (Wu et al., 2023; Pascual-Escudero et al., 2021).

The projection ray mapping demonstrates the use of perspective projection from the camera centre to project 3D points (x1, x2, x3) onto a 2D image plane, as shown in Figure 1. It emphasises the importance of focal length, camera settings, and spatial transformation in precise human pose estimation (Ferraz, Binefa, and Moreno-Noguer, 2014).

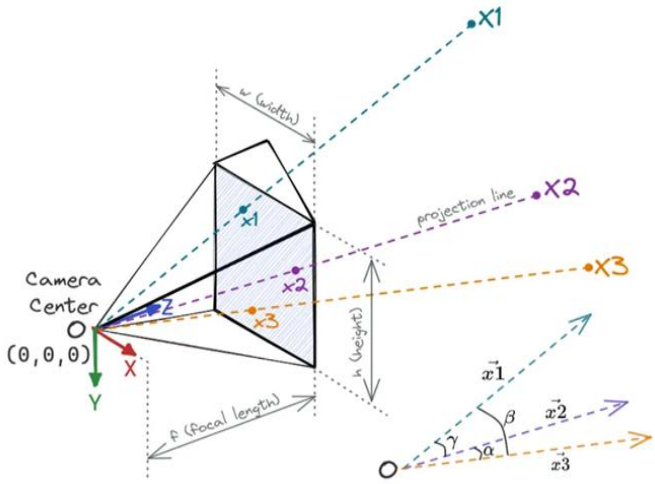

Figure 1: Projection Ray Mapping (Ferraz, L., Binefa, X. and Moreno-Noguer, F., 2014).

The field underwent a paradigm shift with the introduction of deep learning techniques, and was pioneered by LeCun et al. (1998) Convolutional Neural Network (CNN) architecture. This enables breakthroughs, such as DeepPose's end-to-end joint coordinate prediction (Toshev & Szegedy, 2014). Subsequent innovations in research, including Ouyang et al. (2014) multi-source training frameworks and He et al. (2016) residual network designs, substantially enhanced pose estimation reliability by addressing fundamental deep learning challenges.

The graphic displays the results of a convolutional layer that uses learnt filters to highlight and extract significant image features like edges and shapes. As shown in Figure 2, the higher activation levels indicate the presence of more noticeable characteristics in the input image (LeCun et al. 2015).

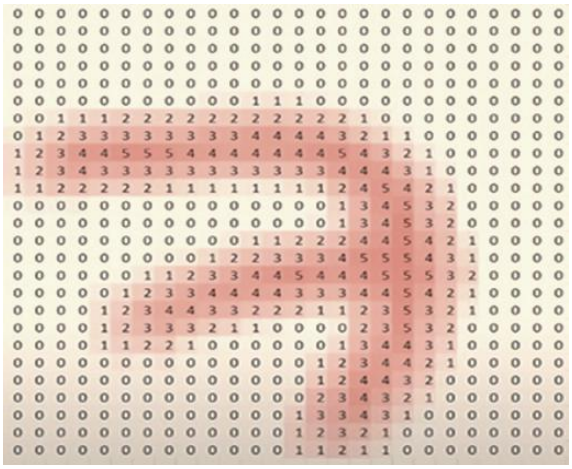

Figure 2: Post-Convolution Feature Extraction (LeCun et al., 2015).

Recent architectural advances have pushed performance boundaries further. The Sun et al. (2019) High-Resolution Networks (HRN) model has achieved unprecedented accuracy through sophisticated feature preservation, while Özyer et al. (2021) comprehensive reviews and specialized platforms, such as PoseTrainer Project Team, (2023), demonstrated the MedTech's growing clinical applicability.

Generative approaches have expanded HPE capabilities, with optimized GAN variants balancing computational demands with motion modeling fidelity (Saito et al., 2020). These healthcare systems with such MedTech abilities show promise for rehabilitation settings by maintaining functionality even in suboptimal conditions, such as low-light environments (Sevikumar et al., 2025).

Emerging evidence highlights how patient-centric digital rehabilitation platforms boost engagement and enhance recovery outcomes (Santórum et al., 2023). These findings validate the effectiveness of interactive and feedback-based systems for remote therapy purposes, particularly relevant for developing TKR physiotherapy monitoring solutions that require sustained patient involvement.

As emphasized in foundational surveys by Sigal (2021), successful clinical implementation requires careful optimization of three key factors: (1) Measurement precision; (2) Computational performance; and (3) Practical adaptability. Current evidence positions deep learning-based HPE as a transformative technology for remote rehabilitation, which can deliver real-time movement analysis to augment traditional physiotherapy while addressing critical healthcare resource limitations.

Methodology

The research approach and methods used to create a deep learning-based HPE system designed for monitoring in physical therapy are described and discussed in this section. It includes the procedures of data collection, preprocessing, model construction, training, and assessment to create a systematic framework that can accurately analyse patient movements.

Data Collection and Description: To build an effective model for analysing physiotherapy postures, a diverse dataset encompassing accurate and inaccurate poses was crucial. The dataset integrated primary data gathered from customized video recordings of patients and secondary data obtained from authorized online physiotherapy resources (Davis et al., 2011; Santórum et al., 2023).

Primary Data: Structured video recordings are captured of participants performing TKR rehabilitation exercises in a controlled setting. Each participant executed prescribed movements, including leg lifts, squats, and step-ups, with proper and intentionally flawed techniques recorded for analysis (Cheong Chung et al., 2020).

Secondary Data: Supplementary datasets were curated from publicly accessible physiotherapy training tutorials, which incorporate diverse demographics, environmental conditions, and camera perspectives (PoseTrainer Project Team, 2023). Ethical compliance has been considered for quality assurance, and proper authorisation was secured for data utilisation.

The combined datasets, which included more than 17 hours of video recordings, were subsequently divided into separate frames to perform the model training.

Data Preprocessing: The video datasets underwent frame-wise processing using OpenCV and NumPy using the following key steps:
Frame Extraction: Videos were sampled at fixed intervals to maintain temporal consistency and consistent time intervals between movements.
Normalization: Pixel intensities (p) were rescaled to [0, 1] in order to uniform all data for stable training convergence.
Augmentation: Spatial transformations, including flipping and rotation, were applied to improve dataset diversity and model robustness.
Annotation: Joint keypoints were extracted using OpenPose and classified as correct or incorrect based on biomechanical alignment to ensure proper positioning of the body's segments during movement (PoseTrainer Project Team, 2023).

The pre-processed dataset was strategically divided into three distinct subsets to optimize model development and evaluation, where 70% was allocated for training purposes, 15% for validation, and 15% for final testing. This partitioning scheme served multiple critical purposes: (1) providing ample training data for robust model learning, (2) enabling hyperparameter optimization and overfitting detection through validation, and (3) ensuring fair assessment of generalization capability using completely unseen test data. The balanced distribution across these subsets was carefully maintained to guarantee unbiased performance comparisons between all investigated architectures while preserving the statistical integrity of the evaluation process.

Model Development: To identify the optimal approach for physiotherapy and pose classification technical solution, three deep learning architectures were developed and evaluated:

Decoupled Convolutional Neural Network: Based on LeCun et al. (1998), the DCNN model processed video frames independently using NumPy for reducing overfitting and improving error recovery, as shown in Figure 3.

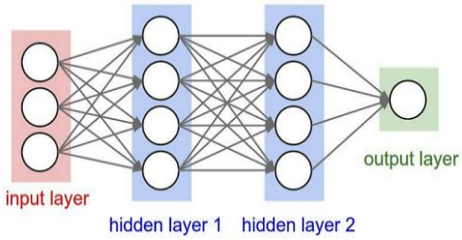

Figure 3: DCNN Architecture (LeCun et al., 1998).

The architecture consisted of:

- Input Layer: Processes normalized frame data by converting pixel values into input nodes for subsequent feature extraction.
- Hidden Layers with ReLU Activation: Two interconnected hidden layers using ReLU activation to capture non-linear pose relationships while preventing gradient vanishing and improving training efficiency.
- Softmax Output Layer: Utilizes Softmax normalization to generate classification probabilities, determining exercise correctness through probabilistic outputs.

By using the decoupled feedback principles and training in batches of 20 frames, the model was able to learn independently across batches to reduce gradient noise and enhance generalisation ability overall.

Sparse Generative Adversarial Network: Implemented in PyTorch, the Sparse GAN model, presented in Figure 4, includes the following key components:

- **Generator, G:** The generator synthesizes artificial pose sequences by progressively transforming random noise through layered neural computations. Its objective is to create movements that increasingly approximate real exercises by responding to the discriminator's evaluation of authenticity.
- **Discriminator, D:** A deep neural classifier distinguishes between genuine and synthesized poses by analysing hierarchical movement patterns. The adversarial training process simultaneously refines both networks and sharpens the discriminator's detection precision while improving the generator's ability to produce convincing simulations.

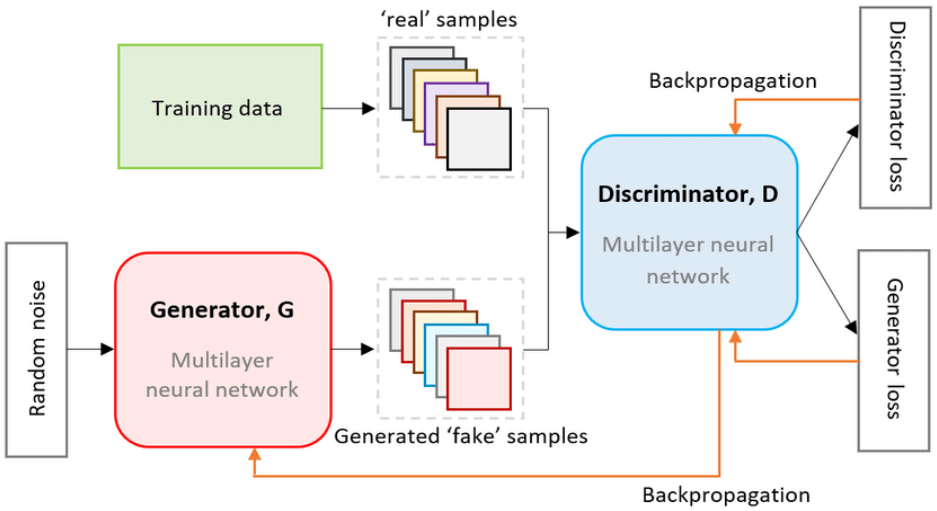

Figure 4: GAN-based synthetic data generation framework (Little et al., 2021).

Network sparsity was optimized using PyTorch's pruning utilities with 20% pruning yielding the best efficiency-performance balance (Saito et al., 2020).

Hybrid DCNN + GAN Model: The hybrid architecture integrated the DCNN as the discriminator within the GAN framework for combining the DCNN's robust classification performance with the GAN's data generation capabilities. By implementing a fully trained DCNN instead of a conventional shallow discriminator, the model achieved more precise discrimination between real and synthetic pose sequences, which significantly enhanced the adversarial learning mechanism. This integration utilised the DCNN's sophisticated feature extraction to have detailed movement analysis while benefiting from the GAN's ability to generate diverse pose variations.

The combined architecture demonstrated superior classification performance, evidenced by higher F1-scores compared to standalone models, indicating improved accuracy in detecting correct and incorrect therapeutic movements. However, this performance gain came with increased computational demands, including longer processing times, greater memory usage, and more extensive training requirements. These trade-offs underscore the need to balance model complexity with practical deployment constraints, particularly in clinical settings where both accuracy and efficiency are critical considerations.

Model Training: The three models were trained on annotated pose dataset for binary classification (correct vs. incorrect form). Training spanned 50 epochs with early stopping to mitigate overfitting. Performance was evaluated using Precision, Recall, and F1-score. Results demonstrated that the hybrid DCNN+GAN model outperformed others with an F1-score of 86.01%, followed by the standalone DCNN with 81.68% and Sparse GAN with 71.09%.

This study implemented a multi-architecture deep learning approach to evaluate automated physiotherapy posture assessment. By systematically collecting and processing motion data, then developing specialized neural networks, the research successfully demonstrated accurate automated evaluation of therapeutic exercises.

Experiments and Findings

During the research study, a comprehensive evaluation of three distinct neural network architectures was conducted for automated physiotherapy monitoring: DCNN model, GAN model, and a hybrid combination DCNN+GAN model. Our quantitative analysis used standard classification metrics precision, recall, and F1-score supplemented by graphical representations of model performance. The comparative assessment reveals critical insights into the relative strengths and limitations of each approach for clinical posture analysis applications.

The Proposed DCNN+GAN Model Performance: The assessment applied three clinically relevant performance metrics, which are Precision, Recall, and the F1 Score. These measures were deliberately selected given their importance in medical diagnostics, where both false identifications and missed detections could potentially impact patient outcomes. Table 1 presents the proposed DCNN+GAN model performance comparison.

Table 1. The Proposed DCNN+GAN Model Performance Comparison.

Model	Precision (%)	Recall (%)	F1 Score (%)
DCNN	82.36	81.01	81.68
Sparse GAN	72.03	70.10	71.09
Combined DCNN+GAN	87.90	84.20	86.01

Inference Time Comparison: Processing latency represents a crucial operational parameter for real-world practice and implementation. We conducted rigorous benchmarking of computational performance across all architectures by measuring execution time for standardized 30-second clinical video segments. Table 2 presents benchmarking of computational performance across the models.

Table 2. Computational Performance Benchmarking.

Model	Inference Time (min)
DCNN	8.2
Sparse GAN	7.1
Combined DCNN+GAN	14.0

While our hybrid model achieved superior classification accuracy with an 86.01% F1-score, its computational overhead presents significant deployment challenges for time-sensitive clinical applications. It has been obvious that Sparse GAN implementation demonstrated the most favourable balance between processing speed, with 7.1 minutes, and acceptable F1-score accuracy of 71.09%, suggesting its potential as a candidate for real-time applications where latency constraints outweigh marginal accuracy gains, see Table 2.

The Research Key Findings: The experimental results demonstrate that the hybrid DCNN+GAN model achieved the highest classification accuracy, making it particularly suitable for clinical applications requiring precise movement assessment. While the sparse GAN model showed the fastest inference speeds advantageous for real-time mobile implementations, this came at the cost of reduced accuracy compared to the hybrid approach. The standalone DCNN provided intermediate performance, being outperformed by the combined model's enhanced capabilities. These findings suggest that the hybrid architecture offers the most promising solution for accurate physiotherapy monitoring in resource-sufficient environments, whereas the sparse GAN represents a viable alternative when real-time processing is prioritized over maximum precision. The research study establishes an important benchmark for future development of AI-assisted rehabilitation systems, which reveals the need for continued optimization to bridge the current accuracy-speed trade-off in practical deployments. This research work provides valuable insights for implementing adaptive pose estimation technologies across different clinical and home-based rehabilitation scenarios using MedTech.

Future research directions could investigate integrating explainable AI methods and algorithms to improve model interpretability. For healthcare applications, such as physiotherapy monitoring, transparent deep learning predictions are crucial for building clinical confidence and patient participation (Bairy and Fränzle, 2023). Emerging work on attention-based explanation models presents viable approaches for enhancing transparency in future rehabilitation systems.

Conclusion

The application of deep learning techniques for automated monitoring of physiotherapy exercises, specifically targeting TKR rehabilitation, has been investigated with significant research findings. The study aimed to develop a system capable of providing real-time postural feedback to patients, thereby reducing dependency on clinical supervision while maintaining exercise quality. As a result, three neural network architectures were valuated: a DCNN model, a computationally optimised GAN model, and the DCNN+GAN hybrid model.

The hybrid architecture demonstrated the highest analytical performance with an F1-score of 86.01%, confirming that model fusion techniques can enhance classification accuracy. In contrast, this improved performance came with significant computational costs, requiring 14 minutes to process a 30-second video segment. On the other hand, the Sparse GAN achieved the fastest processing times, but with less F1-score accuracy of 71.09%, while the DCNN provided intermediate results on both performance metrics. These findings underscore the inherent trade-off between model sophistication and practical deployability in clinical settings, presenting important considerations for implementing AI-assisted rehabilitation systems. The results provide valuable insights for developing tailored solutions based on specific clinical requirements and operational constraints.

Future work will focus on integrating explainable AI approaches and machine learning algorithms to improve model interpretability to enable physiotherapists and patients to better understand the rationale behind exercise feedback. Attention-based explanation models will be explored to highlight key joint movements influencing predictions. This transparency is essential for improving clinical trust, enhancing patient engagement, and supporting personalised rehabilitation.

ACKNOWLEDGMENT

The authors thank the ACE School Research Ethics Team and the School of Health, Sports and Bioscience for their support. Special thanks to Liz Nicholls, Physiotherapy Course Leader, for her valuable clinical insights.

REFERENCES

AbouGrad, H. and Shabarshov, A. 2024. AI-Framework to Detect eCommerce Fake Reviews: A Hybrid Neural Network Machine Learning Model. Artificial Intelligence and Computational Technologies: Innovations, Usage Cases, and Ethical Considerations. Near East University (NEU), Turkey 25-26 Nov 2024 Springer Nature.

Bairy, A., Fränzle, M. (2023). Optimal Explanation Generation using Attention Distribution Model. In: Tareq Ahram and Redha Taiar (eds) Human Interaction and Emerging Technologies (IHIET-AI 2023): Artificial Intelligence and Future Applications. AHFE (2023) International Conference. AHFE Open Access, vol 70. AHFE International, USA. http://doi.org/10.54941/ahfe1002928

Cheong Chung, K.J., Lo, N.N. and Yeo, S.J. (2020). Total knee arthroplasty in patients with stiff knees: outcomes and risk factors for persistent stiffness. Journal of Arthroplasty, 35(1), pp.81–87.

Davis, A.M., Kennedy, D., Wong, R., Robarts, S., Skou, S.T. and McGlasson, R. (2011). A pilot randomized controlled trial of a therapeutic exercise and education program for patients with knee osteoarthritis. Physiotherapy Canada, 63(3), pp.259–269.

Ferraz, L., Binefa, X. and Moreno-Noguer, F. (2014). Very fast solution to the PnP problem with algebraic outlier rejection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp.501–508.

K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 770-778. https://doi.org/10.1109/CVPR.2016.90

Kluzek, S., Palmer, D., Smith, H.E. and Watt, F.E. (2023). Total Knee Replacement: What Patients Should Know. BMJ, 380, p.e072100.

LeCun, Y., Bottou, L., Bengio, Y. and Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), pp.2278–2324.

Little, Claire & Elliot, Mark & Allmendinger, Richard & Samani, Sahel. (2021). Generative Adversarial Networks for Synthetic Data Generation: A Comparative Study. https://doi.org/10.48550/arXiv.2112.01925

Ouyang, W., Chu, X. and Wang, X. (2014). Multi-source Deep Learning for Human Pose Estimation. CVPR, pp.2337–2344.

Özyer, Tansel & Ak, Duygu & Alhajj, Reda. (2021). Human action recognition approaches with video datasets — A survey. Knowledge-Based Systems, Volume 222, 106995. https://doi.org/10.1016/j.knosys.2021.106995

Pascual-Escudero, B., Garcia-Hernando, G., Min, C., Kim, T.K. and Cardenas, J.J. (2021). On the use of PnP algorithms for 3D human pose estimation. Computer Vision and Image Understanding, 208, p.103213.

PoseTrainer Project Team. (2023). PoseTrainer: Human Pose Estimation for Therapy.

Přibyl, B., Zemčík, P. and Čadík, M. (2017). Camera pose estimation from lines using Plücker coordinates. Computer Vision and Image Understanding, 161, pp.120–136.

Saito, M., Saito, S., Koyama, M. and Kobayashi, S. (2020). Pruned GAN for Efficient Human Pose Estimation. Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp.1895–1903.

Santórum, M., Toro, M., Martinez, D., Vargas, P., Maldonado-garcés, V., Acosta-vargas, G., Ayala-chauvin, M., Ortiz-prado, E., González-rodríguez, M. (2023). User Centered Design of a Digital Platform for Therapeutic Education and Respiratory Rehabilitation in Patients with Post-COVID-19. In: Tareq Ahram and Redha Taiar (eds) Human Interaction & Emerging Technologies (IHIET 2023): Artificial Intelligence & Future Applications. AHFE (2023) International Conference. AHFE Open Access, vol 111. AHFE International, USA. http://doi.org/10.54941/ahfe1004010

Sevikumar, Arun & Shibu, Abin & Krishnan, Sreekumar. (2025). Human action recognition using pose estimation in low light conditions. 030029. 10.1063/5.0247064.

Schoen, D.C. (2008). Gender Differences in Total Knee Arthroplasty. Orthopedics, 31(1), pp.21–23.

Sigal, L. (2021). Human Pose Estimation: Foundations and Trends. Foundations and Trends® in Computer Graphics and Vision, 13(2–3), pp.143–258.

Sun, K., Xiao, B., Liu, D. and Wang, J. (2019). Deep High-Resolution Representation Learning for Human Pose Estimation. CVPR, pp.5693–5703.

Toshev, A. and Szegedy, C. (2014). DeepPose: Human Pose Estimation via Deep Neural Networks. CVPR, pp.1653–1660.

Wu, Z., Yang, C., Su, X. and Yuan, X. (2023). Revisiting the Perspective-n-Point Problem for Human Pose Estimation in Deep Learning. Neurocomputing, 519, pp.28–40.