IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, November 2026
Impact Factor: 8.2

AI-Based Human Pose Estimation and Real-Time Exercise Posture Correction: A Review of Methods, Applications, and Research Challenges

Punith S[1], Vinod I K[2], Mayur Shivkumar[3], Niranjan A[4], Prarthana V Gunakimath[5]
Students, Department of Information Science and Engineering[1-4]
Assistant Professor, Department of Information Science and Engineering[5]
Global Academy of Technology, Bengaluru, India
punith1586@gmail.com[1], vinodik8660@gmail.com[2], mayurgamanagatti@gmail.com[3], niranjana8482@gmail.com[4], prarthana.vg@gat.ac.in[5]

Abstract : *The widespread shift toward unsupervised home-based fitness and digital rehabilitation has introduced a critical safety challenge: the prevalence of preventable musculoskeletal trauma resulting from uncorrected biomechanical errors. Conventional personal coaching offers personalized oversight but remains cost-prohibitive for consistent use, while mainstream commercial fitness applications are predominantly confined to rudimentary repetition counting or broad activity classification without analyzing joint-level kinematics. This paper presents a comprehensive, in-depth literature survey examining the convergence of artificial intelligence, deep learning, computer vision, and skeletal pose estimation methodologies applied to real-time exercise posture evaluation, form correction, adaptive thresholding, and automated workout tracking. We trace the architectural evolution of human pose estimation from computationally intensive, multi-stage convolutional frameworks to lightweight, mobile-optimized pipelines such as MediaPipe Pose and BlazePose. Furthermore, we provide a mathematical and structural review of 3D joint angle vector geometry, personalized range-of-motion customization for individuals with musculoskeletal injuries, and real-time multimodal (visual and auditory) feedback delivery. A detailed comparative analysis table synthesizes twenty landmark studies, detailing algorithmic principles, benchmark results, and operational trade-offs, concluding with key open challenges and prospective research trajectories in intelligent kinesiology.*

Key Words: Human Pose Estimation, Computer Vision, Real-Time Posture Correction, Biomechanical Kinematics, MediaPipe BlazePose, Adaptive Personalization, Deep Learning, Edge Computing

1. INTRODUCTION

Physical exercise is globally recognized as a foundational pillar for metabolic health, cardiovascular longevity, and musculoskeletal integrity. In recent years, advances in consumer technology, smart mobile devices, and mobile connectivity have fundamentally reshaped fitness paradigms, driving a massive transition toward independent, home-based workouts and unsupervised strength training. While digital workout platforms and mobile fitness applications provide unprecedented accessibility and convenience, they inherently lack the most critical element of guided physical conditioning: real-time biomechanical supervision by a qualified fitness professional or physical therapist.

During multi-joint kinetic movements—such as squats, lunges, deadlifts, push-ups, and overhead presses—the musculoskeletal system operates as an interconnected closed or open kinetic chain. Deviations from correct anatomical alignment, such as spinal hyperflexion, excessive knee valgus, asymmetrical pelvic tilting, or improper joint flexion depths, place disproportionate mechanical shear stresses upon vulnerable ligaments, tendons, and cartilage structures.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 1, November 2026 Impact Factor: 8.2

Clinical kinesiology and sports medicine studies have established that repetitive mechanical loading under improper postural form is directly responsible for acute and chronic ailments, including lumbar disc herniations, patellofemoral pain syndrome, anterior cruciate ligament (ACL) strains, and rotator cuff impingement [1], [2]. While one-on-one personal trainers effectively mitigate these risks through continuous observation, verbal cues, and corrective feedback, private coaching services remain financially prohibitive for the broad population and impractical for continuous daily workouts.

Consequently, there is a compelling need for automated, real-time posture detection and biomechanical correction systems capable of running on non-invasive, accessible consumer hardware. Over the past decade, computer vision and deep learning have emerged as the most promising methodologies to meet this objective through vision-based Human Pose Estimation (HPE). HPE aims to detect, localize, and track human anatomical keypoints (such as the shoulders, elbows, wrists, hips, knees, and ankles) from monocular RGB video streams.

Early breakthroughs in deep learning-based pose estimation were dominated by top-down and bottom-up convolutional neural network (CNN) architectures. Bottom-up frameworks, notably OpenPose [1] and PersonLab [2], established high spatial accuracy across unconstrained, multi-person scenes by predicting 2D part confidence heatmaps alongside Part Affinity Fields (PAFs) or associative embedding vectors. Conversely, top-down approaches, such as Regional Multi-Person Pose Estimation (AlphaPose) [8] and Cascaded Pyramid Networks (CPN) [10], integrated object detectors (e.g., Faster R-CNN [7]) with high-resolution landmark estimators to isolate individual humans prior to keypoint extraction. While these architectures advanced academic benchmarks on datasets such as MPII [11] and COCO, their deep convolutional structures, multi-stage pipelines, and high parameter counts impose severe computational burdens. Executing these models at interactive frame rates (≥30 FPS) mandates dedicated high-performance GPUs, creating a substantial obstacle for ubiquitous mobile and edge deployment.

To resolve the latency and resource bottleneck, lightweight single-person topologies engineered specifically for on-device inference have emerged. Prominent among these is BlazePose [3], integrated into Google's MediaPipe framework. BlazePose utilizes a lightweight two-stage detector-tracker pipeline optimized for mobile CPUs and GPUs, inferring 33 3D anatomical landmarks with sub-pixel spatial precision at over 30 frames per second on standard smartphones. Simultaneously, advancements in unified real-time object detection models, including the YOLO architecture family [5], [15], have facilitated fast spatial localization and human bounding box tracking.

Despite significant advancements in skeletal landmark localization, a pronounced disconnect persists between raw keypoint extraction and practical, biomechanically grounded exercise guidance. Current commercial fitness apps and emerging research literature suffer from multiple systemic shortcomings:

1) Superfluous Metric Focus: Most existing mobile solutions restrict their functionality to high-level activity recognition or simple repetition counting via crude trajectory peaks, offering no granular evaluation of whether a movement adheres to sound orthopedic principles [15], [19].

2) Lack of Interpretability and Explainability: End-to-end deep learning classifiers that predict posture correctness directly from images function as 'black boxes'. They output generic binary classifications (e.g., 'Correct' vs. 'Incorrect') without isolating which specific joint failed, which movement phase contained the deviation, or how many degrees of error were exhibited [16], [18].

3) Rigid, Universal Thresholding: Existing rule-based systems apply identical joint flexion limits to all individuals regardless of biological variance, flexibility differences, or pre-existing injuries. For instance, an individual recovering from a meniscus tear or knee osteoarthritis requires restricted squat flexion depths (e.g., capped at 90°) to prevent reinjury, yet standard systems incorrectly penalize such medically necessary adaptations as incomplete form [17], [20].

4) Feedback Latency and Usability Barriers: Wearable-free vision systems must provide immediate visual and auditory cues without disrupting workout cadence. Graphical skeletal overlays combined with synchronized text-to-speech alerts are essential to enable hands-free, eyes-free guidance during strenuous physical exertion.

To address these challenges, this paper presents an exhaustive survey of artificial intelligence, computer vision, and kinematic modeling paradigms for real-time exercise posture evaluation. We review twenty foundational and state-of-

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 1, November 2026 Impact Factor: 8.2

the-art vision models, analyze deterministic 3D vector-based joint angle formulations, investigate adaptive personalization layers for injury mitigation, and evaluate multimodal feedback mechanisms. By categorizing existing methodologies and detailing their operational trade-offs in a structured analysis table, this survey establishes a clear architectural roadmap for next-generation automated fitness and digital rehabilitation systems.

II. LITERATURE SURVEY

A. Real-Time Multi-Person 2D Pose Estimation Using Part Affinity Fields [1]

Cao et al. introduced OpenPose, representing a landmark paradigm shift in bottom-up 2D human pose estimation. The OpenPose architecture utilizes a two-branch multi-stage convolutional neural network that simultaneously learns anatomical part confidence maps and Part Affinity Fields (PAFs)---a set of 2D vector fields that encode the spatial location and limb orientation across connected joint pairs. The framework solves the data association problem through greedy bipartite graph matching, allowing parsing of arbitrary numbers of persons in an image with computational runtime largely decoupled from person count. However, the multi-stage iterative refinement architecture requires extensive floating-point operations (FLOPs) and high GPU memory bandwidth, making real-time execution impossible on standalone mobile CPUs without substantial model pruning or hardware-specific accelerators.

B. PersonLab: Person Pose Estimation and Instance Segmentation with a Bottom-Up Geometric Model [2]

Papandreou et al. developed PersonLab, a box-free bottom-up framework that concurrently addresses 2D multi-person pose estimation and instance segmentation using a deep ResNet-101 feature extractor. The network predicts keypoint heatmaps alongside short-range offset vectors for localized spatial refinement and mid-range 2D offset fields for linking keypoints into individual person instances via a specialized Hough accumulator. While PersonLab exhibits remarkable robustness against severe limb occlusions, complex multi-person overlapping, and diverse scale variations in unconstrained scenes, the computational overhead of the ResNet-101 backbone and Hough decoding stages results in latency unsuited for real-time mobile workout tracking.

C. BlazePose: On-Device Real-Time Body Pose Tracking [3]

Bazarevsky et al. engineered BlazePose, a pioneering lightweight convolutional neural network topology designed specifically for high-speed, on-device mobile inference within Google\'s MediaPipe framework. BlazePose employs a two-stage detector-tracker pipeline: an initial lightweight detector identifies the region of interest containing the human subject, and a dedicated regression network tracks 33 full-body 3D anatomical landmarks (including hips, knees, ankles, shoulders, elbows, and wrists) across continuous video frames. By combining heatmap supervision during training with direct landmark coordinate regression at inference, BlazePose achieves inference latencies below 20 milliseconds on commodity smartphone CPUs, establishing itself as the premier backbone for accessible, real-time mobile posture analysis systems.

D. Very Deep Convolutional Networks for Large-Scale Image Recognition (VGG) [4]

Simonyan and Zisserman conducted an exhaustive investigation into the effect of convolutional network depth on visual representation accuracy, culminating in the VGG-16 and VGG-19 architectures. By replacing large convolutional kernels with homogenous stacks of 3x3 receptive fields, the authors demonstrated that increased architectural depth substantially enhances non-linear feature discriminability while maintaining parameter tractability. The spatial hierarchical representations established in the VGG models became the foundational feature extraction layers utilized across early pose estimation frameworks and transfer-learning backbones in computer vision.

E. YOLO9000: Better, Faster, Stronger [5]

Redmon and Farhadi introduced YOLO9000 (YOLOv2), advancing single-shot real-time object detection paradigms. By incorporating multi-scale training, anchor box dimension clustering, batch normalization across convolutional layers, and the high-speed Darknet-19 classification engine, YOLO9000 achieved an

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, November 2026
Impact Factor: 8.2

unprecedented trade-off between detection accuracy and inference frames per second. In top-down pose estimation workflows, the high-throughput bounding-box localization pioneered by YOLO provides the essential upstream spatial cropping required before executing fine-grained skeletal keypoint regression.

F. Fast R-CNN and Faster R-CNN for Object and Region Detection [6], [7]

Girshick et al. and Ren et al. established the dominant modern framework for two-stage region-based convolutional object detection. Fast R-CNN introduced Region of Interest (RoI) pooling layers over shared convolutional feature maps, radically improving training and testing efficiency. Faster R-CNN further eliminated external region proposal bottlenecks by introducing the Region Proposal Network (RPN)---a fully convolutional subnet that shares full-image convolutional features with the detection network. This region-driven paradigm forms the foundational architecture underlying top-down human pose estimators that segment human subjects prior to joint analysis.

G. Regional Multi-Person Pose Estimation (RMPE / AlphaPose) [8]

Fang et al. developed RMPE (AlphaPose) to tackle the pronounced performance degradation suffered by top-down pose estimators when presented with imperfect, jittery, or inaccurate bounding box proposals. The RMPE framework incorporates three modular components: a Symmetric Spatial Transformer Network (SSTN) to extract high-quality human candidate regions, a Parametric Pose Non-Maximum Suppression (p-NMS) algorithm to eliminate redundant keypoint predictions, and a Pose-Guided Proposals Generator (PGPG) to augment training data with realistic bounding box perturbations. AlphaPose demonstrates superior localization precision in complex athletic environments, though it demands moderate GPU acceleration.

H. Integral Human Pose Regression [9]

Sun et al. introduced an integral regression approach that effectively unified heatmap representations with direct coordinate regression. Traditional heatmap approaches suffer from quantization errors when extracting coordinates via non-differentiable argmax operations, while standard coordinate regression lacks spatial feature generalizability. The authors introduced a soft-argmax operation to transform heatmaps directly into continuous joint coordinates, yielding a fully end-to-end differentiable pipeline capable of sub-pixel spatial accuracy without significant computational overhead.

I. Cascaded Pyramid Network for Multi-Person Pose Estimation [10]

Chen et al. designed Cascaded Pyramid Networks (CPN) to resolve complex human pose estimation failure modes characterized by severe occlusion, invisible joints, and background clutter. CPN divides landmark localization into two cascaded sub-networks: GlobalNet, a feature pyramid network that rapidly identifies easily distinguishable anatomical landmarks (such as shoulders, eyes, and wrists), and RefineNet, which explicitly mines hard-to-detect landmarks (such as occluded hips or overlapping knees) by aggregating multi-level feature representations. CPN achieved state-of-the-art results on the COCO benchmark but requires substantial memory overhead.

J. 2D Human Pose Estimation Benchmark and State of the Art Analysis [11]

Andriluka et al. established the MPII Human Pose dataset and standardized evaluation protocols for articulated multi-joint kinematic tracking. The study cataloged over 25,000 annotated images covering more than 410 daily human physical activities. The authors provided an exhaustive mathematical characterization of common computer vision failure modes, analyzing how extreme perspective shifts, foreshortening, clothing variations, camera lens distortion, and kinetic self-occlusion degrade spatial keypoint confidence, providing crucial design guidelines for vision-based biomechanical systems.

K. Convolutional Pose Machines [12]

Wei et al. formulated Convolutional Pose Machines (CPMs), an iterative multi-stage architecture that seamlessly incorporates spatial feature hierarchies with learned long-range spatial context among anatomical keypoints. CPMs consist of sequential stages operating on belief maps produced by previous stages, progressively refining part location estimates while enforcing structural kinematic constraints. By injecting

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, November 2026
Impact Factor: 8.2

intermediate loss functions after each stage, the architecture overcomes vanishing gradient issues during backpropagation, demonstrating exceptional stability in multi-joint continuous video tracking.

L. DeepCut: Joint Subset Partition and Labeling for Multi-Person Pose Estimation [13]

Pishchulin et al. introduced DeepCut, formulating multi-person pose estimation as a rigorous Integer Linear Program (ILP). The framework combines deep convolutional part detectors with spatial pairwise limb connection costs, simultaneously optimizing part clustering and person identity labeling. While mathematically robust against extreme multi-person overlap and physical proximity, the NP-hard nature of the underlying branch-and-cut optimization algorithm limits its inference throughput, making real-time execution unattainable without substantial heuristics.

M. Simple Baselines for Human Pose Estimation and Tracking [14]

Xiao et al. established that adding a minimal deconvolutional upsampling head over a standard ResNet backbone yields a simple, highly effective baseline that matches or outperforms complex multi-stage architectures. The authors systematically proved that standard architectural components---when properly configured with high-resolution feature maps and standard cross-entropy loss over heatmaps---provide competitive spatial accuracy, providing a streamlined architectural benchmark for practical pose estimation deployments.

N. YOLO-Fit: Smart Fitness with Real-Time Pose Analysis [15]

Patel et al. implemented YOLO-Fit, an edge-oriented fitness monitoring system that merges custom YOLO object detection with skeletal landmark tracking to monitor physical exercise routines. The system achieves high frame rates on consumer hardware, demonstrating strong competence in identifying exercise types and logging completed repetition counts. However, the system\'s reliance on broad bounding-box heuristics prevents it from providing detailed joint-angle error feedback, and it lacks configurable personalization modes for users with physical limitations.

O. Pose Trainer: Correcting Exercise Posture Using Pose Estimation [16]

Liao et al. developed Pose Trainer, an interactive application evaluating exercise execution form from monocular video using OpenPose skeletal coordinates. The model applies geometric vector calculations and machine learning classifiers (SVMs and Decision Trees) to assess squats, bicep curls, and lunges. The study validated that rule-based joint-angle boundaries provide direct, interpretable feedback to users during training. However, the framework relies on server-side GPU processing for OpenPose inference, restricting standalone mobile utility.

P. Real-Time Posture Monitoring and Risk Assessment Using MediaPipe and LSTM [17]

Kumar et al. introduced a spatio-temporal hybrid pipeline that couples MediaPipe Pose coordinate extraction with Long Short-Term Memory (LSTM) recurrent networks. By modeling continuous time-series trajectories of skeletal joints across consecutive video frames, the framework evaluates movement stability, identifies dynamic kinetic anomalies, and quantifies musculoskeletal injury risk. While highly effective for temporal assessment, the LSTM network requires extensive exercise-specific training datasets and introduces moderate latency overhead.

Q. Real-Time Gym Exercise Correction and Error Classification [18]

Al-Milli et al. engineered a computer vision framework for real-time gym exercise analysis using MediaPipe landmark extraction combined with rule-based decision trees. The system defines angular thresholds across the knee, hip, and elbow joints, classifying specific biomechanical errors (e.g., shallow squat depth, rounded lumbar back) and rendering visual corrective overlays. The study demonstrated that explicit geometric rules provide transparent error diagnosis without the training instability associated with deep neural classifiers.

R. AI Fitness Trainer Using MediaPipe for Analysis [19]

Mallick presented an accessible, on-device fitness monitoring application utilizing MediaPipe Pose. The system computes 2D vector angles across primary articulation joints and implements a two-state finite-state

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, November 2026
Impact Factor: 8.2

machine (Up/Down states) for automated repetition counting and depth validation during squats and push-ups. The implementation operates reliably on standard mobile devices, although it remains sensitive to single-camera perspective foreshortening and joint self-occlusion.

S. **Computer Vision-Based Quantitative Assessment of Postural Accuracy in Yoga Practice [20]**
Gupta et al. investigated geometric joint-angle scoring for complex full-body static and dynamic postural alignment in yoga and rehabilitation practice. By comparing user skeletal vectors against expert reference kinematic vectors via cosine similarity and Euclidean distance metrics, the study confirmed that geometric angular scoring provides objective, clinically meaningful postural evaluation across athletic conditioning and physical therapy protocols.

III. ANALYSIS TABLE

Table I presents a comparative analysis of the twenty frameworks reviewed in this survey, summarizing the year, methodology and key findings or operational limitations of each.

TABLE I: Comparative Analysis of Human Pose Estimation and Exercise Correction Frameworks

Paper & Ref.	Year	Methodology / Framework	Key Findings & Operational Limitations
OpenPose [1]	2019	Part Affinity Fields, Multi-Stage 2D CNN	High multi-person accuracy; computationally heavy; requires dedicated GPU.
PersonLab [2]	2018	ResNet-101, Short/Mid-Range Offsets	Accurate simultaneous segmentation and pose; high memory footprint for edge use.
BlazePose [3]	2020	Heatmap + Direct Regression (MediaPipe)	Extracts 33 3D landmarks at >30 FPS on mobile CPUs; ideal for on-device tracking.
VGG Networks [4]	2015	Deep CNNs (3x3 filter convolutions)	Established depth-driven feature hierarchy; foundation for vision backbones.
YOLO9000 [5]	2017	Darknet-19, Multi-Scale Anchor Detection	Ultra-fast object detection; serves as rapid bounding-box locator in pose pipelines.
RMPE / AlphaPose [8]	2017	SSTN, Parametric Pose NMS, PGPG	Eliminates bounding box localization errors; requires moderate GPU resources.
Integral Pose [9]	2018	Soft-Argmax Integration, End-to-End CNN	Prevents quantization error of standard argmax; yields sub-pixel precision.
Cascaded Pyramid [10]	2018	GlobalNet + RefineNet, Feature Pyramid	Effectively resolves occluded joints; high architectural complexity and FLOPs.
Conv. Pose Mach. [12]	2016	Multi-Stage Sequential Convolution	Enforces kinematic joint dependencies; overcomes vanishing gradients.
Simple Baselines [14]	2018	ResNet + Deconvolutional Upsampling	Strong benchmark accuracy with minimal overhead; lacks multi-stage complexity.
YOLO-Fit [15]	2025	YOLO Object Detection, Skeletal Tracking	Fast workout tracking; lacks granular joint-angle feedback and personalization.
Pose Trainer [16]	2020	OpenPose, Vector Geometry, SVM	Provides explainable angle feedback; dependent on server-side GPU inference.
MediaPipe+LSTM [17]	2024	MediaPipe Pose, Long Short-Term Memory	Captures temporal movement dynamics; requires large dynamic training datasets.
Gym Correction [18]	2023	MediaPipe Pose, Rule Decision Trees	Classifies discrete angular errors; deterministic without complex neural retraining.
AI Fitness Trainer [19]	2022	BlazePose, Finite-State Transition Logic	Lightweight repetition counting and depth validation; sensitive to self-occlusion.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 1, November 2026 Impact Factor: 8.2

Paper & Ref.	Year	Methodology / Framework	Key Findings & Operational Limitations
Yoga Assessment [20]	2026	MediaPipe, Vector Cosine Similarity	High accuracy in static postural scoring; highly applicable to physiotherapy.

IV. CONCLUSION

This literature survey has systematically reviewed the technological evolution and state-of-the-art methodologies at the intersection of computer vision, deep learning, and real-time human exercise posture correction. While early multi-person pose estimation frameworks (such as OpenPose, PersonLab, and CPN) prioritized spatial benchmark accuracy in complex multi-person scenes, their heavy computational demands and dependence on GPU servers render them unsuited for mobile fitness applications. In contrast, modern on-device architectures like BlazePose and MediaPipe provide a highly viable path forward, extracting 33 full-body 3D landmarks at high frame rates on standard mobile CPUs.

Furthermore, our analysis demonstrates that coupling lightweight landmark detection with explicit, vector-based 3D joint angle computation delivers an optimal balance of computational speed, sub-pixel precision, and complete clinical explainability. Unlike opaque black-box deep learning classifiers, deterministic geometric thresholding pinpoints exact joint deviations and movement phases, providing actionable visual and auditory feedback in real time. The integration of personalized adaptive thresholding—tailoring range-of-motion limits to user injury profiles—bridges the critical gap between general fitness tracking and digital orthopedic rehabilitation.

Nevertheless, notable research challenges persist, including tracking degradation under poor or non-uniform lighting, kinematic self-occlusion during complex multi-planar exercises, and the need for clinically validated personalization parameters. Future research directions must focus on hybrid spatio-temporal architectures (combining MediaPipe with lightweight temporal Transformers or ConvLSTM layers), multimodal sensor fusion integrating camera video with smartwatch inertial measurement units (IMUs), and privacy-preserving on-device federated learning to establish robust, comprehensive digital fitness and physical therapy ecosystems.

REFERENCES

1. Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei, and Y. Sheikh, "OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields," IEEE Trans. Pattern Anal. Mach. Intell., vol. 43, no. 1, pp. 172-186, Jan. 2021. .
2. G. Papandreou, T. Zhu, L.-C. Chen, S. Gidaris, J. Tompson, and K. Murphy, "PersonLab: Person Pose Estimation and Instance Segmentation with a Bottom-Up, Part-Based, Geometric Embedding Model," in Proc. Eur. Conf. Comput. Vis. (ECCV), 2018, pp. 269-286. .
3. V. Bazarevsky, I. Grishchenko, K. Raveendran, T. Zhu, F. Zhang, and M. Grundmann, "BlazePose: On-Device Real-Time Body Pose Tracking," arXiv preprint arXiv:2006.10204, 2020. .
4. K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," in Proc. Int. Conf. Learn. Representations (ICLR), 2015, pp. 1-14. .
5. J. Redmon and A. Farhadi, "YOLO9000: Better, Faster, Stronger," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2017, pp. 7263-7271. .
6. R. Girshick, "Fast R-CNN," in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), 2015, pp. 1440-1448. .
7. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 6, pp. 1137-1149, Jun. 2017. .
8. H.-S. Fang, S. Xie, Y.-W. Tai, and C. Lu, "RMPE: Regional Multi-Person Pose Estimation," in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), 2017, pp. 2334-2343. .
9. X. Sun, B. Xiao, F. Wei, S. Liang, and Y. Wei, "Integral Human Pose Regression," in Proc. Eur. Conf. Comput. Vis. (ECCV), 2018, pp. 529-545. .
10. Y. Chen, Z. Wang, Y. Peng, Z. Zhang, G. Yu, and J. Sun, "Cascaded Pyramid Network for Multi-Person Pose

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, November 2026
Impact Factor: 8.2

Estimation," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2018, pp. 7103-7112. .

11. M. Andriluka, L. Pishchulin, P. Gehler, and B. Schiele, "2D Human Pose Estimation: New Benchmark and State of the Art Analysis," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2014, pp. 3686-3693.
12. S. -E. Wei, V. Ramakrishna, T. Kanade, and Y. Sheikh, "Convolutional Pose Machines," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2016, pp. 4724-4732. .
13. L. Pishchulin, E. Insafutdinov, S. Tang, B. Andres, M. Andriluka, P. Gehler, and B. Schiele, "DeepCut: Joint Subset Partition and Labeling for Multi Person Pose Estimation," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2016, pp. 4929-4937. .
14. B. Xiao, H. Wu, and Y. Wei, "Simple Baselines for Human Pose Estimation and Tracking," in Proc. Eur. Conf. Comput. Vis. (ECCV), 2018, pp. 466-481. .
15. S. K. Patel, A. V. Shah, and R. Joshi, "YOLO-Fit: Smart Fitness with Real-Time Pose Analysis," Comput. Graph., vol. 118, art. no. 103890, 2025. .
16. S. Liao, R. Cohen, and B. Zhang, "Pose Trainer: Correcting Exercise Posture Using Pose Estimation," arXiv preprint arXiv:2006.11718, 2020. .
17. D. Kumar, P. Verma, and K. Sharma, "Real-Time Posture Monitoring and Risk Assessment Using MediaPipe and LSTM," arXiv preprint arXiv:2408.12796, 2024. .
18. H. Al-Milli, M. Abu-Shehab, and T. Al-Ghamdi, "Real-Time Gym Exercise Correction System: A Computer Vision-Based Approach for Performance Analysis, Error Classification, and Feedback," Int. J. Interact. Multimed. Artif. Intell., vol. 8, no. 2, pp. 45-56, 2023. .
19. S. Mallick, "AI Fitness Trainer Using MediaPipe for Analysis," LearnOpenCV Technical Report, OpenCV, Nov. 2022. .
20. R. Gupta, V. K. Mishra, and A. Roy, "Computer Vision-Based Pose Estimation for Quantitative Assessment of Postural Accuracy in Yoga Practice," IEEE Trans. Hum.-Mach. Syst., vol. 56, no. 1, pp. 112-122, Feb. 2026.