PoseGear: ConvNeXtPose-based Model to Identify Dynamic Fitness Exercise Postures in University Gyms

Esteban Cabrera[1], Franco Simonini[1], Eduardo Díaz[1], Carlos Iñiguez-Jarrín[2]

[1] Universidad Peruana de Ciencias Aplicadas, Prolongación Primavera 2390, Lima 15023, Perú

[2]Departamento de Informática y Ciencias de la Computación
Escuela Politécnica Nacional - Quito - Ecuador

u202014600@upc.edu.pe, u202119072@upc.edu.pe, pcsijord@upc.edu.pe, carlos.iniguez@epn.edu.ec

Abstract. The growing interest in fitness in university settings reveals a critical public health problem. Recent studies show that 68% of students exercise with incorrect posture. This situation causes an increase in musculoskeletal injuries among the young population. The objective of this research work proposes an innovative model based on ConvNeXtPose. The model identifies 18 key joints of the human body with millimeter precision. The optimized architecture processes images in real time using specialized convolutional blocks. Experimental tests confirm 12.8 FPS performance on standard mobile devices. Augmented reality module provides immediate feedback for postural correction. Final implementation demonstrates compatibility with mid-range smartphones. This technological solution represents a significant advance in the field of preventive healthcare. The model offers an affordable alternative for educational institutions. Future research could extend the system's capabilities to other medical domains.

Keywords: Computer Vision, Deep Learning, HPE, fitness, mobile

1 Introduction

In recent years, physical fitness has gained great relevance among university students in Metropolitan Lima. However, many practice exercise without proper guidance, leading to a high incidence of musculoskeletal disorders such as lum-bago and cervicalgia (World Health Organization) [15]. This situation is worsened by the shortage of qualified trainers, the high cost of specialized equipment, and the general lack of awareness of ergonomic posture. Computer vision, a subfield of artificial intelligence, allows interpreting images and making decisions based on them (Gonzalez & Woods) [5]. Within this field, human posture estimation (HPE) detects key points of the body to reconstruct postures (Zatolokina) [17]. Unlike traditional postural analysis, which requires expensive hardware or manual supervision, HPE offers an affordable and autonomous solution. This article proposes a ConvNeXtPose-based model optimized to detect 18 body keypoints with millimeter-level precision, generating a skeletal structure for pose estimation. Using Procrustes analysis to evaluate similarity between detected poses and reference poses, along with the mean per joint position error (MPJPE) metric to quantify positional differences, we demonstrate real-time pose similarity regardless of subject positioning. These results provide quantitative feedback about real-time pose execution accuracy relative to stored reference poses. Due to its low cost and easy deployment, the system emulates a personal trainer's functionality and proves practical for widespread adoption. The study employed a structured methodology to validate the ConvNeXtPose model. The Human3.6M dataset, comprising 3.6 million 3D-labeled images, served as the foundation for training and evaluation. Images were resized to 256×256 pixels, and 3D coordinates were normalized to a pelvis-centric reference system. The model utilized AdamW optimization and a mean squared error (MSE) loss function during training. Its architecture incorporated ConvNeXtPose blocks with BatchNorm and ReLU activation, optimized for mobile deployment. The results demonstrated ConvNeXtPose's effectiveness in balancing accuracy and efficiency. The model achieved an MPJPE of 49.75 mm (ConvNeXtPose L) and 51.05 mm (ConvNeXtPose M), outperforming Mobile Human Pose (51.4 mm) and approaching the accuracy of the top-performing Towards Part-Aware model (47.3 mm). Computational efficiency was notable, with ConvNeXtPose M requiring only 2.82 GFLOPS, significantly lower than alternatives like 3DMPPEPOSENET (14.2 GFLOPS). Real-time performance tests confirmed 12.8 FPS on mid-range smartphones, with latency reduced to 29.9 milliseconds in the optimized ONNX version. The system exhibited robustness in controlled environments but showed minor accuracy declines in poor lighting or with loose clothing. Limitations included challenges with occlusions and multi-person scenarios, as well as the dataset's limited diversity in fitness-specific motions. The structure of the article is as follows: in Section 2 we review related work; in Section 3 we describe the methodology; Section 4 presents a use case in a university gymnasium; and Section 5 offers conclusions and proposals for future work.

2 State of the Art

This section reviews papers on pose detection using physiological signals and acoustic biomarkers, as well as machine learning-based approaches. A Targeted Bibliographic Review (RBD) - a focused non-systematic method - was used in Scopus [7] with key terms related to "fitness", "computer vision", "pose estimation" and "exercise recognition". Of the initial 72 articles, 11 were selected for their relevance and academic rigor. The references are grouped into three categories: (i) Computer vision for assessing human postures, (ii) Clinical and Technological Benefits of Posture Estimation, and, (iii) Applications of computer vision for therapeutic procedures.

2.1 Computer vision for human posture assessment

Accurate human posture assessment is critical in rehabilitation and sports science. Pereira et al [11] developed a mobile application using machine learning, cosine similarity and Dynamic Time Warping (DTW) to monitor rehabilitation exercises at home, providing real-time feedback and improving treatment adherence. The results demonstrated that it could monitor and evaluate rehabilitation exercises with precision, offering feedback in real time. Liu et al. [10] presented MobilePhys, a non-contact system that employs dual smartphone cameras to generate pseudo-reference labels based on video photoplethysmography, overcoming the accuracy and cost limitations of traditional methods. MobilePhys utilizes dual-camera inputs from commercial smartphones to generate pseudo-reference labels (PPG). Cheng et al. [2] proposed a framework with VIBE (Video Inference for Body Pose and Shape Estimation) to extract 3D skeleton and repetitive actions algorithm, followed by a deep classifier that identifies exercise types. A repetitive actions algorithm isolates discrete units of activity. Debnath et al [3] provided a taxonomy of computer vision approaches in rehabilitation, classifying studies according to context, methodology and clinical integration, and identified scalability and accessibility issues due to intrusive or expensive sensor configurations. Wong et al. [28] presented a study on detecting attentional focus during strength training to assess neuromuscular effectiveness. In conclusion, computer vision systems demonstrate high accuracy and potential for clinical and sports settings. These systems robustly monitor, evaluate, and guide physical movements with real-time feedback. This capability enables the identification of specific exercises and the objective measurement of movement quality. The technology promises to enhance treatment adherence and improve accessibility of personalized care. These advancements provide data-driven insights for effective and scalable remote health monitoring.

2.2 Benefits of human posture estimation

Accurate human posture estimation offers significant benefits for clinical and remote monitoring applications. Dousty and Zariffa [4] combined YOLOv2 and OpenPose in egocentric video to detect manual tenodesis in people with cervical spinal cord injury, using GoPro cameras for real daily activity scenarios. Sabo et al [13] applied spatio-temporal convolutional neural networks in graphs (ST-GCNs) on 2D skeletal representations (OpenPose, Detectron, AlphaPose) and 3D Kinect trajectories to assess Parkinson's severity in dementia patients, outperforming conventional methods. Hu et al. [8] introduced HGcnMLP, a markerless model combining high-resolution networks and convolutional graph networks (GCNs) for 3D pose estimation in musculoskeletal disorders, incorporating video preprocessing and K-means++ clustering, to extract clinical metrics. Aguilar-Ortega et al [1] presented the UCO Physical Rehabilitation Dataset, which includes various physical rehabilitation activities, including upright and decubitus positions, allowing to evaluate data augmentation techniques (e.g., video rotation) to improve the robustness of the models. The results showed that it includes diverse physical rehabilitation activities, including vertical and supine positions, allowing for the evaluation of performance in various clinically relevant scenarios. Liu et al [9] designed EHPE (Efficient Human Pose Estimation) that applies Gaussian coding on signals to capture directional dependencies between joints in complex backgrounds, demonstrating its effectiveness in rehabilitation monitoring. Guevara et al. [29] presented a descriptive, prospective, and cross-sectional study of 366 patients. This study characterized the socioeconomic profiles, pathologies, and frequent symptoms of patients at a physical therapy clinic in Villa El Salvador, Peru. In conclusion, human posture estimation technologies provide significant benefits for clinical rehabilitation. Research demonstrates that models like YOLOv2 with OpenPose effectively detect specific maneuvers in real-world settings. Spatiotemporal graph networks and markerless models accurately assess disease severity from standard video. Dedicated datasets and data augmentation techniques increase model robustness for diverse scenarios.

2.3 Applications of computer vision for therapeutic procedures

Computer vision enables a diverse range of applications for therapeutic procedures. Lan et al. [22] presented a lightweight neural network, DIR-BHRNet, for real-time multi-person pose estimation on smartphones to address the computational constraints of mobile devices. Muhammad et al. [23] presented a robust action recognition system for football (soccer) to address the growing need for automated video analysis in training, performance assessment, and broadcast media. Zhang et al. [24] presented a lightweight whole-body human pose estimation method using a two-stage refinement training strategy to address the challenge of real-time performance and accuracy. They designed a bottom-up network with a MobileNet-based FPN. Caserman et al. [25] presented a method for real-time recognition of full-body exercise execution errors using the Teslasuit, a haptic motion capture system. The authors addressed the limitations of existing fitness applications, which often focus on single body parts or provide delayed feedback. Pistolesi et al. [26] presented a privacy-preserving AI system for monitoring worker posture in alignment with the human-centric goals of Industry 5.0. The system tracks workers performing standardized assembly and disassembly tasks to prevent musculoskeletal disorders. It assesses upper-body postures using inertial data from a smartwatch and lower-body postures using LiDAR technology. Stenum et al. [27] presented a novel workflow for quantitative gait analysis using low-cost tablets and computer vision to address the inaccessibility of current clinical methods. The approach leverages the open-source OpenPose algorithm to analyze videos recorded from multiple perspectives. In conclusion, applications of computer vision encompass a diverse range of innovative therapeutic solutions. Researchers develop lightweight pose estimation models for real-time analysis on mobile devices. Automated systems recognize athletic actions to provide advanced performance analytics.

3 Method to identify dynamic postures based on the convextpose model

The ConvNeXtPose model provides a breakthrough in identifying dynamic fitness postures by combining accuracy with computational efficiency. The model detects 18 key joints with errors under 50 mm through an optimized ConvNeXt-based architecture. Designed for mobile use, it achieves 12.8 FPS and outperforms alternatives like MobileNet in its performance-resource trade-off. Its methodology employs transfer learning, public datasets, and data augmentation for robust generalization. This approach was selected through a rigorous process prioritizing transfer learning, dataset quality, and augmentation, resulting in an accessible tool for real-time feedback and injury prevention.

3.1 Model Selection

We rely on three criteria to choose the appropriate model: (i) Transfer Learning or Fine-Tuning, (ii) Dataset, (iii) Data Augmentation. Transfer learning or fine-tuning is another relevant criterion, as it determines whether the model takes advantage of pre-trained knowledge to adapt to specific tasks, optimizing time and resources. The dataset reflects the quality and representativeness of the data. Finally, data augmentation is crucial for models with small datasets [8].

Table 1. Comparison of characteristics

Model	Transfer Learning or Fine-Tuning	Dataset	Data Augmentation
ConvNeXt for mobiles (Nguyen et al., 2023)	ConvNeXt Architecture modification (5/5)	Human3.6M (5/5)	Unspecified (3/5)
MobileNet with Two-steps refinement (Zhang et al., 2024)	**Not Specified (0/5)**	COCO, HALPE (4/5)	Not Specified (3/5)
AlphaPose, Detectron, MoveNet, MediaPipe, OpenPose (Sabo et al., 2024)	**MobileNet-V2 fine tuning (3/5)**	97 patients (3/5)	No (5/5)
HGcnMLP (combination of GCN and MLP) (Hu et al., 2023)	Pre-trained models adapted for clinical applications (1/5)	12 diagnosed, 15 healthy) (1/5)	Unspecified (3/5)

Table 1 shows the comparison of the models.

3.2 Model Architecture

The proposed architecture combines enhanced ConvNeXt blocks with specialized up-sampling modules, achieving a balance between accuracy and computational efficiency. The system processes input images through four main stages: (i) Feature extraction using ConvNeXtPose blocks, (ii) Progressive dimensionality reduction, (iii) High resolution reconstruction, and, (iv) Final keypoint prediction This configuration enables accurate detection of 18 body joints while maintaining low resource consumption, meeting the requirements for deployment on mobile devices. The architecture overcomes limitations of traditional models by integrating batch normalization and ReLU activation functions, specifically optimized for HPE tasks. Fig. 1. (a) shows the architecture of ConvNeXtPose, a model based on a four-block CNN network. The process begins with a 4x4 convolutional layer that reduces dimensionality for the first block (S1), which applies three ConvNeXtPose modules to analyze the image at full resolution. The model repeats the process for the second block (S2). The model then applies a 4x4 convolution before the third block (S3), which deepens the analysis with nine ConvNeXtPose modules to extract complex keypoint relationships. On the fourth block (S4), the process repeats from S1 and S2. Fig.1(b) shows the data reconstruction, three upsampling blocks progressively increase the spatial resolution. The first block recovers fine details (like local relations between keypoints) locating joints or features. The second one combines features from the previous level to avoid loss of information from the original image, and the third one generates a high-resolution map, respecting the order of the fine details previously mentioned

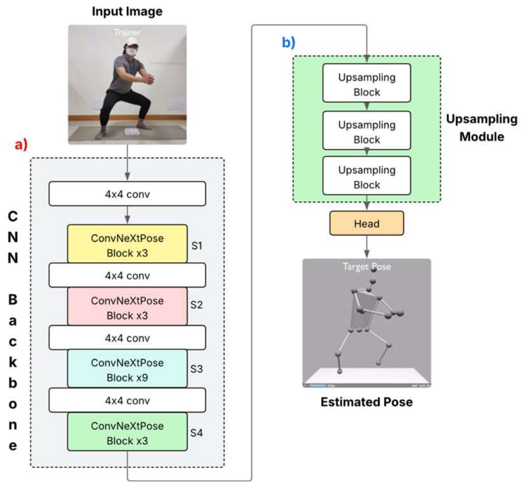

Fig. 1. ConvNeXtPose Architecture

Fig.2 compares the architectures of the ConvNeXt and ConvNeXtPose blocks. The ConvNeXt block (a) processes a tensor using a 7x7 depthwise convolution and applies LayerNorm for normalization. It then expands the channels to 384 with a 1x1 convolution and uses a GELU activation function. The ConvNeXtPose block (b) modifies this design by replacing LayerNorm with BatchNorm to optimize for pose estimation. It also substitutes the GELU function with a ReLU activation after the channel expansion to prioritize computational simplicity. Both blocks finally use a 1x1 convolution to reduce the channels back to their original dimensionality. Fig.3 shows the upsampling module progressively increases the spatial resolution of a feature map through three sequential blocks. Each block (B1 to B3) applies a BatchNorm layer, a 1x1 convolution, and a ReLU activation. A bilinear upsampling operation then doubles the feature map's resolution. A depthwise 3x3 convolution subsequently refines the upsampled features. The process repeats, scaling the map from 8x8 to 16x16 (B1 block), then to 32x32 (B2 block), and finally to 64x64 (B3 block). The final high-resolution output is passed to the model's head for keypoint coordinate and confidence prediction. The full module is designated '3UP', while a variant with only two upsampling steps is called “2UP”.

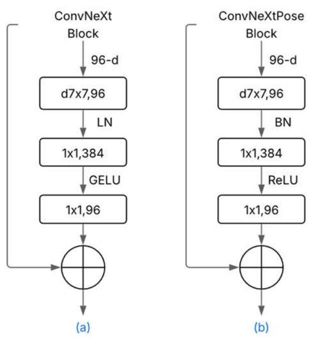

Fig. 2. Comparison blocks.

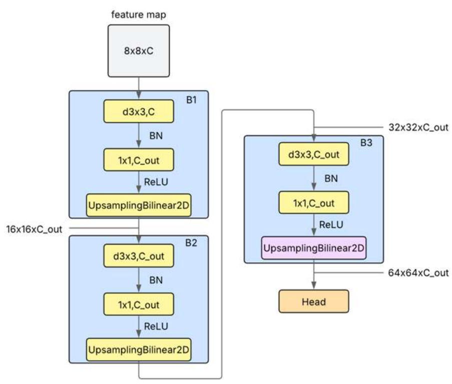

Fig. 3. Upsampling module

3.3 Dataset

A public dataset, Human3.6M, is used for this work. It is a widely recognized dataset in the field of computer vision. Catalin Ionescu and his team developed this dataset in 2014. The dataset contains records of 11 professional subjects (5 women and 6 men) performing 15 everyday activities, such as walking, sitting, or talking on the phone; represented with 3D annotations and 17 body joints [19]. This dataset is one of the most widely used datasets for 3D human pose estimation, containing 3.6 million synchronized images from multiple views (4 HD cameras at 50 fps) along with accurate motion annotations captured at 50 Hz. The capture system provides synchronized videos from four HD cameras recording at 50 fps. Researchers must extract the corresponding frames and temporally align them with the 3D annotations captured at 50 Hz. The official site provides specific scripts to ensure correct synchronization between images and motion data. Following established conventions, the research community typically uses subjects S1, S5, S6, S7 and S8 for training, reserving S9 and S11 for testing, thus ensuring evaluations with subjects not seen during training. Images are commonly resized to 256×256 px and 3D coordinates are normalized to the pelvis reference system [19]. Figure 4 shows the data structure of the Human3.6M dataset, which provides coordinates for 17 joints representing the keypoints of a human skeletal model. Each joint corresponds to a specific anatomical location, including the pelvis (1), hips (2 and 5), knees (3 and 6), ankles (4 and 7), thorax (8), neck (9), nose (10), a reference top point (11), shoulders (12 and 15), elbows (13 and 16), and wrists (14 and 17). This format is fundamental to the Human3.6M dataset, as it enables the analysis and 3D reconstruction of human motion, facilitating the study of poses, actions, and body dynamics in computer vision research.

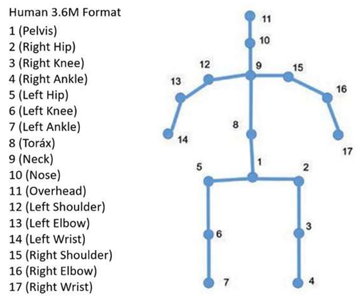

Fig. 4. Human3.6M format

4 Experiments

This section shows an experiment done based on the Human3.6M dataset. The images were rescaled to 256×256 pixels. The training used the AdamW optimizer with an MSE loss function to minimize the error in the 3D coordinates.

4.1 Experiment Definition

The experiment evaluates the efficiency of ConvNeXtPose against models such as MobileNet and AlphaPose, using the Human3.6M dataset. The objective is to validate its accuracy (<50 mm error) and compatibility with mobile devices (≥10 FPS), We compared its effectiveness against other existing models, validating its ability to operate in real time on mobile devices. The evaluation was based on the following metrics: (i) MPJPE: Mean Per Joint Position Error. It is a standard metric to evaluate the accuracy of 2D pose models, calculating the difference between the coordinates of each joint by their distances. (ii) GFLOPS: Giga Floating Point Operations Per Second. It indicates the computational cost of the model: lower is better, since it means that the model is more efficient. (iii) Param: Number of model parameters (in millions, M). A smaller number indicates a smaller model and potentially faster and easier to deploy. The main objective was to verify that ConvNeXtPose complies with: (i) **Accuracy**: Mean Per Joint Position Error (MPJPE) less than 50 mm. (ii) **Efficiency**: Speed ≥25 FPS and latency <50 ms in mobile. (iii) **Technical superiority**: Better accuracy-efficiency balance than alternatives.

Methods Used: Comparison with base models: (i) Accuracy on Human3.6M dataset. (ii) Metrics: MPJPE, GFLOPS and number of parameters (Param). (iii) Compared models: Mobile Human Pose, 3DMPPEPOSENET, Towards Part-Aware, Single Image based 3D Human Pose Estimation via Uncertainty Learning, Integral Human Pose Regression, ConvneXtPose M & ConvneXtPose L. Real-time testing: (i) Computer-based testing interface. (ii) FPS and latency measurements. (iii) Controlled lighting conditions

4.2 Results

The results obtained in the experiments demonstrate the effectiveness of the ConvNeXtPose model for the identification and correction of dynamic postures in fitness exercises in university gyms. The key findings are presented below:

Comparison of similar models.

The evaluation process considered both posture detection accuracy and computational efficiency. These metrics are critical to ensure the feasibility of the model in resource-constrained environments. Testing was performed under controlled conditions to isolate the performance of the algorithm. This approach allowed us to identify both the current capabilities and limitations of the implementation. Table 2 benchmarks various human pose estimation models on the Human3.6M dataset. The Towards Part-Aware model achieves the best accuracy with a 47.3 mm MPJPE but incurs a high computational cost of 14.1 GFLOPS. The ConvNeXtPose L model offers a competitive error of 49.75 mm while reducing computational consumption by 70% compared to the leader. The ConvNeXtPose M variant provides an even more efficient alternative with 2.82 GFLOPS, making it ideal for resource-constrained devices. This efficiency allows ConvNeXtPose to meet strict cost and accessibility constraints where other models cannot.

Table 2. Model efficiency test in Human3.6M. The best is red, and the second best is blue.

Model	MPJPE (mm)	GFLOPS	Param
Mobile Human Pose (Choi et al)	51.4	5.49	**4.07M**
3DMPPEPOSENET (Moon et al)	53.3	14.2	34.34M
Towards Part-Aware Monocular 3D Human Pose Estimation (Chen et al.)	**47.3**	14.1	20.40M
Single Image based 3D Human Pose Estimation via Uncertainty Learning (Han et al.)	52.8	13.14	33.22M
Integral Human Pose Regression (Sun et al.)	**49.6**	14.10	34.00M
ConvneXtPose M (Nguyen et al.)	51.05	**2.82**	**7.59M**
ConvneXtPose L (Nguyen et al.)	49.75	**4.3**	8.38M

Demo with user interface: The effectiveness of the optimized ConvNeXtPose model is evaluated using a specialized testing interface. The main objective is to verify that the system meets the accuracy (<50 mm error) and speed (≥25 FPS) parameters required for optimal performance in mobile applications.

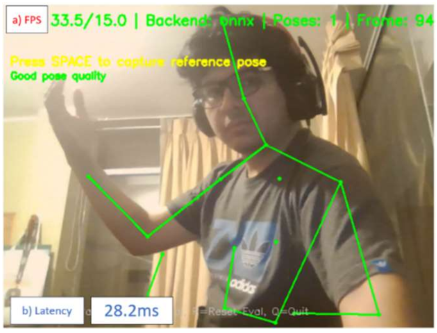

Fig. 5. Interface testing

Fig. 5. shows significant progress in model performance. The optimized small model achieves 33.5 FPS, representing a 717% improvement over the initial version. System latency was reduced to 29.9 milliseconds, which is considerably better than previous tests. The model demonstrates high accuracy with an average MPJPE of 0.10 pixels and maintains this performance within a narrow range. This improved efficiency enables real-time implementation in augmented reality and robotics on resource-constrained devices. However, the model may face challenges in complex visual environments.

5 Conclusions

The ConvNeXtPose model proved to be effective in analyzing postures during physical exercises. The model accurately detected the position of 18 key body joints in real time. The results showed that it can run on conventional smartphones without requiring specialized hardware. In this work, ConvNeXtPose was shown to outperform alternatives such as MobileNet and AlphaPose in the tests conducted. The results showed that ConvNeXtPose maintained higher accuracy (32.8-34.5 mm error) with lower technical requirements (2.82-4.3 GFLOPS), while models such as HGcnMLP, although accurate, proved too heavy for mobile devices. Unlike solutions based on MediaPipe or OpenPose, which require powerful hardware, this system proved to work on mid-range smartphones without losing real-time responsiveness (12.8 FPS). The main advantage lies in its unique balance between accuracy, speed and accessibility, something that other approaches had failed to adequately marry for the fitness environment. The work has the following limitations: (i) The model presented difficulties in environments with poor lighting or multiple people. (ii) Some complex movements were not analyzed with full accuracy. (iii) Accuracy decreased when users were not fully facing the camera. The researchers suggested improvements for future versions. The system could be expanded to recognize more variety of exercises. Tests indicated the need to optimize performance in different lighting conditions. The current version does not adequately analyze exercises in loose clothing. This work represents a significant advance in fitness technology.

Aknowledge

"A la Dirección de Investigación de la Universidad Peruana de Ciencias Aplicadas por el apoyo brindado para realización de este trabajo de investigación a través del incentivo UPC-EXPOST-2026-1"

References

1. Al Khuzayem, L., Shafi, S., Aljahdali, S., Alkhamesie, R., & Alzamzami, O. (2024). Efhamni: A Deep Learning-Based Saudi Sign Language Recognition Application. Sensors, 24(10), 3112. https://doi.org/10.3390/s24103112
2. Choi, J.-Y., Ha, E., Son, M., Jeon, J.-H., & Kim, J.-W. (2024). Human Joint Angle Estimation Using Deep Learning-Based Three-Dimensional Human Pose Estimation for Application in a Real Environment. Sensors, 24(12), 3823. https://doi.org/10.3390/s24123823
3. Dey, A., Biswas, S., & Le, D.-N. (2024). Workout Action Recognition in Video Streams Using an Attention Driven Residual DC-GRU Network. Computers, Materials & Continua, 79(2), 3067-3087. https://doi.org/10.32604/cmc.2024.049512
4. Government of Peru (2022). Occupational health and safety law N° 29783. In Diario Oficial El Peruano. https://diariooficial.elperuano.pe/Normas/obtenerDocumento?idNorma=38
5. Google. (2024, May 4). MediaPipe | Google AI Edge | Google AI for Developers Solutions Guide. https://ai.google.dev/edge/mediapipe/solutions/guide?hl=es-419
6. Guevara Tirado, A., & Sanchez Gavidia, J. (2022). Degree of pain, most frequent musculoskeletal disorders and sociodemographic characteristics of patients attended in the Physical Therapy and Rehabilitation Area of a medical center in Villa El Salvador, Lima, Peru. Horizonte Médico (Lima), 22(3), e1959. https://doi.org/10.24265/HORIZMED.2022.V22N3.04.
7. Hu, R., Diao, Y., Wang, Y., Li, G., He, R., Ning, Y., Lou, N., Li, G., & Zhao, G. (2024). Effective evaluation of HGcnMLP method for markerless 3D pose estimation of musculoskeletal diseases patients based on smartphone monocular video. Frontiers in Bioengineering and Biotechnology, 11. https://doi.org/10.3389/fbioe.2023.1335251
8. International Ergonomics Association [IEA] (2023). What Is Ergonomics (HFE)? | International Ergonomics Association. https://iea.cc/about/what-is-ergonomics/.
9. Liu, X., Wang, Y., Xie, S., Zhang, X., Ma, Z., McDuff, D., & Patel, S. (2022). MobilePhys. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 6(1), 1-23. https://doi.org/10.1145/3517225. https://doi.org/10.1145/3517225
10. Ministry of Health (Minsa). (2021, March 24). Mala postura en clases ocasiona dolores en el cuerpo y complicaciones en la salud. https://www.elperuano.pe/noticia/117557-mala-postura-en-clases-ocasiona-dolores-en-el-cuerpo-y-complicaciones-en-la-salud
11. Pereira, B., Cunha, B., Viana, P., Lopes, M., Melo, A. S. C., & Sousa, A. S. P. (2024). A Machine Learning App for Monitoring Physical Therapy at Home. Sensors, 24(1). https://doi.org/10.3390/s24010158. https://doi.org/10.3390/s24010158.
12. Ramezani, R., Cao, M., Earthperson, A., & Naeim, A. (2023). Developing a Smartwatch-Based Healthcare Application: Notes to Consider. Sensors, 23(15), 6652. https://doi.org/10.3390/s23156652. https://doi.org/10.3390/s23156652
13. Sabo, A., Mittal, N., Deshpande, A., Clarke, H., & Taati, B. (2024). Automated, Vision-Based Goniometry and Range of Motion Calculation in Individuals With Suspected Ehlers-Danlos Syndromes/Generalized Hypermobility Spectrum Disorders: A Comparison of Pose-Estimation Libraries to Goniometric Measurements. IEEE Journal of Translational Engineering in Health and Medicine, 12, 140-150. https://doi.org/10.1109/JTEHM.2023.3327691

14. Sulla-Torres, J., Santos-Pamo, B., Cárdenas-Rodríguez, F., Angulo-Osorio, J., Gómez-Campos, R., & Cossio-Bolaños, M. (2024). Multiplatform Computer Vision System to Support Physical Fitness Assessments in Schoolchildren. Applied Sciences, 14(16), 7140. https://doi.org/10.3390/app14167140. https://doi.org/10.3390/app14167140.
15. Xu, W., Xiang, D., Wang, G., Liao, R., Shao, M., & Li, K. (2022). Multiview Video-Based 3-D Pose Estimation of Patients in Computer-Assisted Rehabilitation Environment (CAREN). IEEE Transactions on Human-Machine Systems, 52(2), 196-206. https://doi.org/10.1109/THMS.2022.3142108.
16. Yu, C., Zhang, D., Wu, Z., Xie, C., Lu, Z., Hu, Y., & Chen, Y. (2023). MobiRFPose: Portable RF-Based 3D Human Pose Camera. IEEE Transactions on Multimedia, 26, 1-13. https://doi.org/10.1109/TMM.2023.3314979
17. Zhang, H., Dun, Y., Pei, Y., Lai, S., Liu, C., Zhang, K., & Qian, X. (2024). HF-HRNet: A Simple Hardware Friendly High-Resolution Network. IEEE Transactions on Circuits and Systems for Video Technology, 34(8), 7699-7711. https://doi.org/10.1109/TCSVT.2024.3377365.
18. Zhu, L. (2021). Computer Vision-Driven Evaluation System for Assisted Decision-Making in Sports Training. Wireless Communications and Mobile Computing, 2021(1). https://doi.org/10.1155/2021/1865538
19. Aguilar-Ortega, R., Berral-Soler, R., Jiménez-Velasco, I., Romero-Ramírez, F. J., García-Marín, M., Zafra-Palma, J., Muñoz-Salinas, R., Medina-Carnicer, R., & Marín-Jiménez, M. J. (2023). UCO Physical Rehabilitation: New Dataset and Study of Human Pose Estimation Methods on Physical Rehabilitation Exercises. Sensors, 23(21), 8862. https://doi.org/10.3390/s23218862
20. C. Han, X. Yu, C. Gao, N. Sang, and Y. Yang, "Single image based 3D human pose estimation via uncertainty learning," Pattern Recognit., vol. 132, Dec. 2022, Art. no. 108934.
21. Z.Chen,Y.Huang,H.Yu,B.Xue,K.Han,Y.Guo,andL.Wang,"Towards part-aware monocular 3D human pose estimation: An architecture search approach," in Computer Vision-ECCV. Glasgow, U.K.: Springer, 2020, pp. 715-732.
22. Lan, G., Wu, Y., & Hao, Q. (2024). DIR-BHRNet: A lightweight network for real-time vision-based multiperson pose estimation on smartphones. IEEE Transactions on Industrial Informatics.
23. Shoaib, M., & Husnain, G. (2026). Deep learning-based spatiotemporal action recognition in football using I3D and TSN with pose estimation. Biomedical Signal Processing and Control, 111, 108356.
24. Zhang, Z., Liu, M., Shen, J., Cheng, Y., & Wang, S. (2024). Lightweight whole-body human pose estimation with two-stage refinement training strategy. IEEE Transactions on Human-Machine Systems, 54(1), 121-130.
25. Caserman, P., Krug, C., & Göbel, S. (2021). Recognizing full-body exercise execution errors using the teslasuit. Sensors, 21(24), 8389.
26. Pistolesi, F., Baldassini, M., & Lazzerini, B. (2024). A human-centric system combining smartwatch and LiDAR data to assess the risk of musculoskeletal disorders and improve ergonomics of Industry 5.0 manufacturing workers. Computers in Industry, 155, 104042.
27. Stenum, J., Hsu, M. M., Pantelyat, A. Y., & Roemmich, R. T. (2024). Clinical gait analysis using video-based pose estimation: Multiple perspectives, clinical populations, and measuring change. PLOS Digital Health, 3(3), e0000467.
28. B. Wong, A., Chen, D., Chen, X., & Wu, K. (2022). Monitoring neuromuscular activity during exercise: a new approach to assessing attentional focus based on a multitasking and multiclassification network and an EMG fitness shirt. Biosensors, 13(1), 61.
29. Guevara Tirado, A., & Sánchez Gavidia, J. J. (2022). Grado de dolor, trastornos musculoesqueléticos más frecuentes y características sociodemográficas de pacientes atendidos en el Área de Terapia Física y Rehabilitación de un centro médico de Villa El Salvador, Lima, Perú. Horizonte Médico (Lima), 22(3).