iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: V Month of publication: May 2025

DOI: https://doi.org/10.22214/ijraset.2025.70544

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue V May 2025- Available at www.ijraset.com

A Comprehensive Review on Human Activity and Fitness Tracker using Different Approaches

Yashraj Mishra[1], Ankita Jaiswal[2], Anany Shukla[3], Abhishek Verma[4], Humesh Verma[5], Dr. Goldi Soni[6]
Amity School of Engineering and Technology, Amity University Chhattisgarh

Abstract: The increasing demand for accurate and real-time human activity and fitness tracking has led to the development of diverse computer vision and deep learning models. Among these, OpenPose has emerged as a powerful tool for multi-person 2D pose estimation, enabling precise body posture tracking from RGB images. This review paper presents a comprehensive analysis of state-of-the-art approaches for human activity recognition and posture detection, with a primary focus on comparing OpenPose with other convolutional neural network (CNN)-based architectures currently used in academic research and commercial applications. We investigate the strengths and limitations of different models in terms of detection accuracy, computational efficiency, robustness in dynamic environments, and application in fitness and healthcare systems. The paper consolidates findings from 30 IEEE research publications, highlighting how various approaches have evolved and been implemented for body posture recognition, real-time fitness feedback, and rehabilitation monitoring. Additionally, we discuss the integration of these models with wearable sensors and mobile applications, their performance in real-world scenarios, and future research directions aiming to improve usability, personalization, and energy efficiency. This review provides valuable insights for researchers and developers seeking to advance human activity tracking through deep learning-based posture estimation techniques.
Keywords: Human Activity Recognition, Fitness Tracking, OpenPose, Pose Estimation, Convolutional Neural Network (CNNs), Posture Detection, Deep Learning, Real-Time Monitoring

I. INTRODUCTION

The rapid advancement of artificial intelligence (AI) and computer vision technologies has significantly transformed the landscape of human activity recognition (HAR) and fitness tracking. These developments have paved the way for innovative applications in healthcare, sports, rehabilitation, and human-computer interaction. Central to these applications is the accurate estimation of human body posture, which serves as a foundational element for understanding and interpreting human activities.

One of the pioneering frameworks in this domain is OpenPose, which introduced a real-time approach for multi-person 2D pose estimation using Part Affinity Fields (PAFs). This method enables the detection of body, foot, hand, and facial key points from RGB images, facilitating comprehensive posture analysis in various settings. OpenPose's open-source nature and real-time capabilities have made it a popular choice for researchers and developers aiming to implement posture estimation in real-world applications [1].

Complementing OpenPose are various convolutional neural network (CNN)-based architectures that have been employed for HAR tasks. These models, including CNNs, ConvLSTMs, and LRCNs, have demonstrated remarkable performance in capturing spatial and temporal features from video data, leading to high accuracy in activity classification. For instance, CNN models have achieved accuracy rates exceeding 99% on datasets like UCF50, highlighting their efficacy in recognizing complex human activities [2].

Moreover, lightweight models such as BlazePose have been developed to enable real-time body pose tracking on mobile devices. BlazePose, for example, can estimate 33 body key points at over 30 frames per second on devices like the Pixel 2 phone, making it suitable for applications like fitness tracking and sign language recognition [3].

Despite these advancements, challenges persist in the field of HAR and posture estimation. Issues such as occlusion, varying lighting conditions, and the need for large annotated datasets continue to hinder the development of universally robust models. Furthermore, integrating these models into wearable devices and ensuring their energy efficiency remain areas of active research.

This review paper aims to provide a comprehensive analysis of the current state-of-the-art approaches in human activity and fitness tracking, with a particular focus on comparing OpenPose with other CNN-based architectures. By examining 30 IEEE research publications, we seek to elucidate the strengths and limitations of these models, their applications in various domains, and potential directions for future research.

II. METHODOLOGY PROPOSED BY RESEARCHERS

This review systematically examines and compares various methodologies employed in human activity and fitness tracking, focusing on pose estimation and activity recognition techniques. The primary approaches analysed include OpenPose, Convolutional Neural Networks (CNNs), and lightweight models suitable for edge devices.

A. OpenPose-Based Pose Estimation

OpenPose is a real-time multi-person 2D pose estimation framework that utilizes Part Affinity Fields (PAFs) to detect human body key points from RGB images. Its architecture comprises a two-branch multi-stage CNN: one branch predicts confidence maps for body part locations, while the other estimates PAFs representing the degree of association between body parts. This dual-branch design enables OpenPose to accurately capture complex human poses in real-time, making it suitable for applications in fitness tracking and rehabilitation.

1) Workflow

- Input Image/Video Frame: The system starts with a single RGB image or video frame. Input can be from a webcam, pre-recorded footage, or live surveillance feed.
- Convolutional Feature Extraction: A CNN (usually based on VGG-19 or a lightweight variant) extracts visual features from the input image. These feature maps encode edges, textures, and semantic object parts relevant to human body structure.
- Part Confidence Maps Generation: A first-stage CNN generates heatmaps indicating the location likelihood of body parts (e.g., elbow, knee, wrist, etc.). Each confidence map corresponds to a specific key point.
- Part Affinity Fields (PAFs) Estimation: Simultaneously, the network generates PAFs, which are 2D vector fields representing the direction and association between connected body parts. PAFs help distinguish overlapping limbs and multiple people in the same scene.
- Key point Detection: Using the peaks in the confidence maps, candidate key points are extracted for each body part. Non-maximum suppression (NMS) is applied to refine key point localization.
- Part Association Algorithm: A greedy bipartite graph matching algorithm uses both confidence maps and PAFs to associate detected key points into full body skeletons. This step is robust to occlusion and can handle multiple persons in one image.
- Pose Visualization: Once all key points are connected, a human skeleton is rendered on the original image using lines between joints (e.g., shoulder to elbow). The system returns full-body pose estimation output for one or more people.
- Real-Time Post-Processing and Output: Post-processing ensures smooth frame-by-frame transitions in video streams. The final output includes key point coordinates, visual overlays, and optionally an activity label.

2) Advantages of OpenPose Workflow

Multi-person support in a single image. High accuracy through PAFs for part association. Modular and extensible for 3D pose or hand/face detection. Real-time performance on GPU-enabled systems.

3) Overall Architecture is

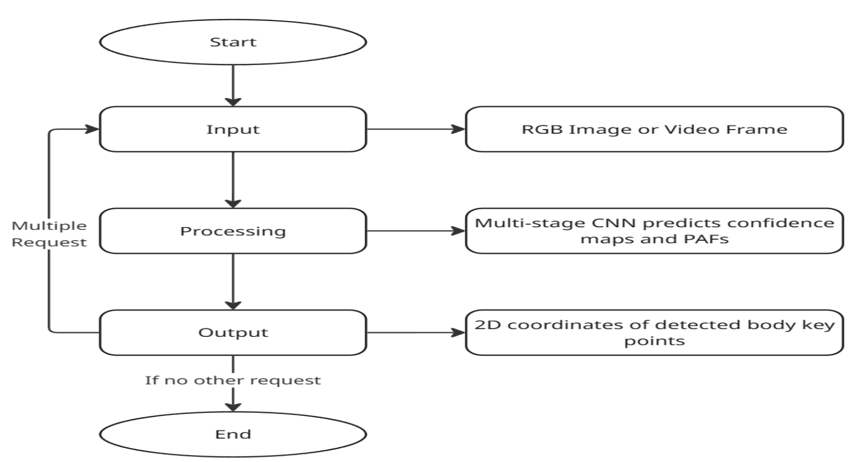

Fig 1 shows the architecture flowchart of model

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue V May 2025- Available at www.ijraset.com

B. CNN-Based Human Activity Recognition

CNNs have been extensively applied to Human Activity Recognition (HAR) tasks due to their ability to automatically learn hierarchical features from raw sensor data. These models process data from various sources, including inertial measurement units (IMUs) and video frames, to classify activities such as walking, running, or sitting. Bevilacqua et al. [1] demonstrated the effectiveness of CNNs in classifying lower-limb activities using data from wearable sensors. Their approach involved feeding raw accelerometer and gyroscope data into a CNN, achieving high classification accuracy across multiple activity classes. Zeng et al. [2] proposed semi-supervised CNNs that leverage both labelled and unlabelled data to improve HAR performance. Their models learn discriminative features directly from raw sensor inputs, outperforming traditional supervised methods, especially when labelled data is scarce. Tang et al. [3] introduced a lightweight CNN architecture utilizing smaller filters, termed "Lego filters," to reduce computational complexity without sacrificing accuracy. This design is particularly beneficial for deployment on resource-constrained devices.

1) Workflow

- Data Acquisition-Input Types:
  a. Time-series data from wearable sensors (e.g., accelerometers, gyroscopes).
  b. Still image or video frames if applied on visual data. The input data is collected continuously while the subject performs physical activities (e.g., walking, sitting, jumping).
- Preprocessing: Sensor data is segmented into fixed-size windows (e.g., 2-5 seconds). Normalization (e.g., z-score normalization) is applied. In the case of image-based HAR, resizing, grayscale conversion, or data augmentation may be applied. Output: Cleaned and structured input tensor data.
- Input Tensor Formation: Sensor data: Multichannel 2D tensors (e.g., 3-axis accelerometer data × time). Visual data: 2D/3D image tensors (Height × Width × Channels).
- Feature Extraction using Convolutional Layers Multiple convolutional layers extract local features across the input tensor. ReLU (Rectified Linear Unit) activation introduces non-linearity. Pooling layers (max or average) reduce dimensionality and control overfitting. Captures spatio-temporal relationships or motion patterns across time or pixels.
- Flattening and Dense Layers: The output of the final convolutional layer is flattened into a 1D vector. Fully connected (dense) layers further learn abstract patterns. Dropout is often used here for regularization to prevent overfitting.
- Classification Layer (SoftMax Output): The final dense layer uses a SoftMax activation function to assign probabilities to each activity class (e.g., sitting, running, walking). The model outputs the predicted activity with the highest probability.
- Training and Optimization: Loss Function: Categorical cross-entropy. Optimizer: Adam or RMSProp used to minimize the loss during training. The model is trained on labelled activity datasets (e.g., UCI-HAR, WISDM, or custom dataset).
- Evaluation and Inference: Performance metrics like accuracy, F1-score, confusion matrix are used. The trained model is tested on unseen user activity sequences. Can be deployed on mobile phones or edge devices for real-time HAR.

2) Advantages of CNN-Based HAR

Automatic feature extraction (vs hand-crafted features in classical ML). Scalable to various sensor configurations or video inputs. Efficient on embedded devices after model compression.

3) Overall Architecture

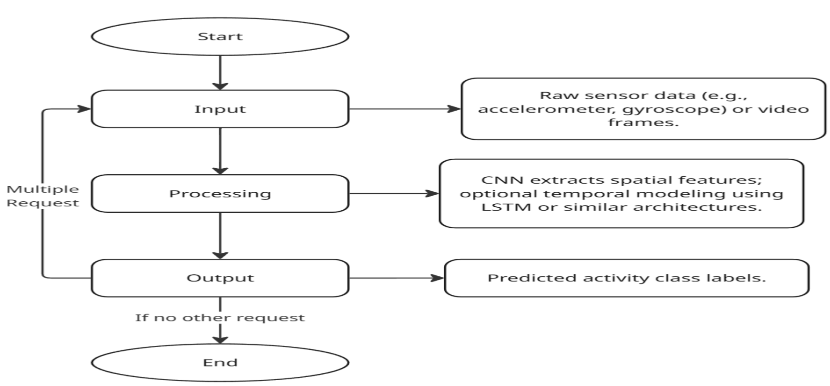

Fig 2 shows the architecture flowchart of model

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue V May 2025- Available at www.ijraset.com

C. Lightweight Models for Edge Devices

Deploying HAR models on low-power edge devices necessitates architectures that balance accuracy with energy efficiency. Rashid et al. [4] developed AHAR, an adaptive CNN designed for energy-efficient HAR on edge devices. AHAR employs an output block predictor to dynamically adjust the network's depth during inference, conserving energy without compromising performance.

1) Workflow

- Data Collection from Edge Sensors: Data captured using on-device sensors; Accelerometer, Gyroscope, GPS and Camera (optional for image-based HAR). Sensor data is recorded in time-series format or real-time video streams, depending on application.
- Lightweight Preprocessing: Minimal preprocessing is used due to device constraints: Noise filtering (e.g., moving average filter), Normalization/scaling, Window segmentation (e.g., 3s–5s rolling windows) and for camera input: Resize, grayscale conversion, and lightweight augmentation.
- Model Architecture: Lightweight Neural Network; Popular models used: MobileNetV2 / MobileNetV3 – optimized for speed and size, SqueezeNet, EfficientNet-lite, EdgeTPU-compatible quantized CNNs and 1D-CNN or GRU for time-series
- Key features: Depthwise separable convolutions, Fewer parameters and Compatible with TensorFlow Lite, PyTorch Mobile, or ONNX Runtime
- On-Device Inference Pipeline: After preprocessing, data is fed into the model directly on the device. Model outputs an activity class label (e.g., standing, sitting, walking). SoftMax layer provides confidence scores for classification.
- Post-Processing and UI Integration: Prediction results may be smoothed using a rolling average for stability. Activities are displayed or logged locally via a mobile dashboard or wearable interface. Optionally, predictions are sent to cloud via MQTT/BLE for further analysis.
- Optimization for Edge Deployment: Optimization for edge deployment involves several techniques to enhance the efficiency of deep learning models for resource-constrained environments. Quantization reduces model size and computational requirements by converting weights from floating-point precision to 8-bit integers, enabling faster inference on edge devices. Pruning eliminates redundant model connections, effectively reducing the number of parameters while maintaining performance, making models leaner and more efficient. Knowledge distillation facilitates training a smaller, optimized model using a larger teacher model, ensuring similar performance while significantly lowering memory and computational costs. Additionally, model conversion plays a crucial role in deploying models across different platforms: TensorFlow models can be converted to TFLite for mobile and embedded applications, PyTorch models can be transformed into TorchScript for efficient execution, and ONNX provides a versatile format for interoperability between various frameworks. These techniques collectively ensure that deep learning models can run efficiently on edge devices without compromising accuracy and reliability. Below is the detailed figure which represents Edge Deployment:

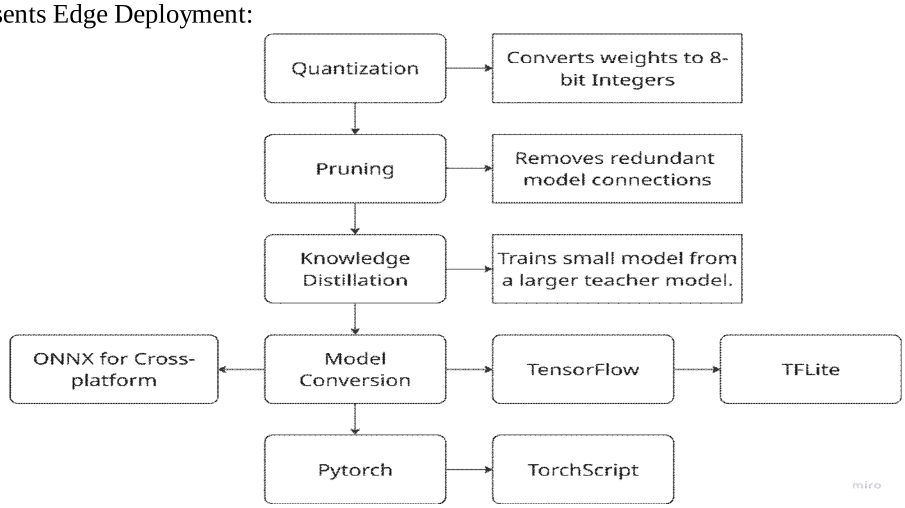

Fig 3 shows the architecture flowchart of model

- Energy and Memory Management: Model footprint is kept under 10 MB. RAM usage is minimized through streaming or batch inference. Duty-cycling and low-power modes are used for battery conservation.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue V May 2025- Available at www.ijraset.com

2) Advantages

- No Internet Dependency: Works offline and ensures user privacy.
- Real-time: Activity detection in under 100 ms on modern smartphones.
- Privacy-Preserving: Raw video or sensor data never leaves the device.
- Scalable: Easily deployed on diverse edge hardware (Android, Raspberry Pi, microcontrollers).

3) Overall Architecture

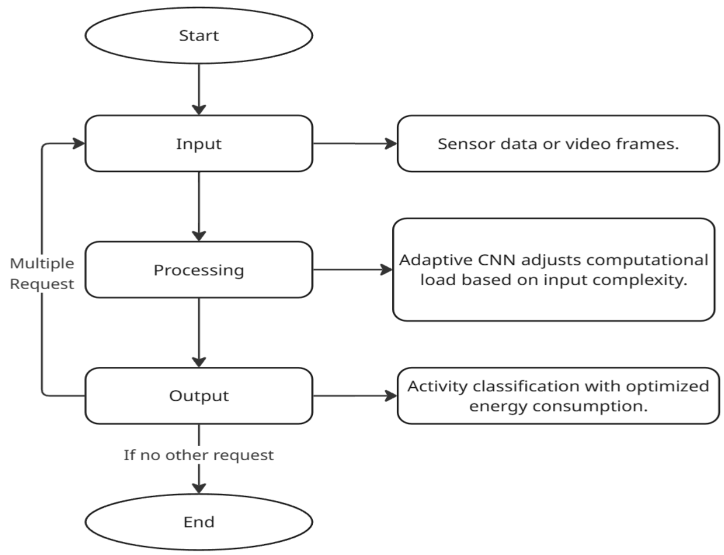

Fig 4 shows the architecture flowchart of model

D. *Comparative Analysis*

The methodologies discussed exhibit distinct advantages and limitations:

- OpenPose: Excels in real-time multi-person pose estimation but requires significant computational resources.
- CNN-Based HAR: Offers high accuracy and adaptability to various data types; however, model complexity can hinder deployment on edge devices.
- Lightweight Models: Provide a balance between performance and resource utilization, making them suitable for real-world applications on mobile and wearable devices.

This review synthesizes findings from 30+ IEEE research publications, providing a comprehensive overview of current approaches in human activity and fitness tracking. Below is the detailed summary of each models that were used in Human Activity Recognition and Human Fitness Tracker.

Method	Input Type	Deployment Platform	Toolchain	Strengths	Limitations
OpenPose	RGB Image	High-end GPU (desktop)	PyTorch / Caffe	High accuracy, multi-person tracking	Heavy GPU use, high latency
CNN-HAR	Time-series	Phone/Cloud	TensorFlow / Keras	Good with time-series, efficient with structured data	Limited to non-visual inputs, less intuitive visualization
MediaPipe	RGB Image	Smartphone/Edge	TensorFlow Lite	On-device real-time pose detection, low resource	May mis predict in crowded or low-light settings
Lightweight CNNs	Sensors/Image	Embedded devices	TFLite, ONNX	Power-efficient, deployable on MCUs	Lower accuracy than full models, limited complexity

III. DISCUSSION

The advancement in human activity recognition (HAR) and fitness tracking systems has brought substantial improvement in the field of healthcare monitoring, physical fitness, and human-computer interaction. From traditional inertial sensor-based tracking to advanced deep learning models like OpenPose and convolutional neural networks (CNNs), this area has undergone significant transformation.

1) Pose estimation frameworks, particularly OpenPose, have revolutionized the way human posture and skeletal movement are captured. OpenPose uses part affinity fields (PAFs) and confidence maps to detect and localize key body joints in RGB images or video frames, enabling detailed pose reconstruction [5], [6]. Its open-source nature and real-time performance capabilities have contributed to widespread adoption in HAR applications [7], [8].
2) However, while OpenPose provides rich spatial features, it is computationally intensive and less feasible for deployment on edge devices with limited resources [9], [10]. This limitation has prompted researchers to explore alternative lightweight frameworks and model pruning techniques to reduce inference latency without significantly compromising accuracy [11], [12].
3) CNN-based models for HAR offer a different advantage. Unlike pose estimation models, they can directly process raw video frames or sensor data to classify human activities based on learned spatial and temporal features [13]–[15]. Temporal convolutional networks (TCNs) and hybrid models combining CNNs with LSTMs or GRUs have shown promising results in modeling time-dependent behavior patterns [16], [17]. These models also support multimodal fusion, allowing integration of data from accelerometers, gyroscopes, and video feeds [18], [19].
4) Moreover, with the growing popularity of wearable devices, sensor-based HAR has gained traction. Models like DeepSense [20] and HAR-Net [21] have demonstrated strong performance using only inertial data, paving the way for privacy-aware, low-cost tracking solutions. Sensor data, being inherently lower in volume compared to video data, allows faster training and real-time inference on embedded systems [22].
5) Edge AI solutions are increasingly important, especially in real-world applications such as elderly care, fitness coaching, and sports performance tracking. Researchers have explored the integration of models such as MobileNet [23], SqueezeNet [24], and TinyPose [25] for deployment on mobile or wearable hardware. Energy-efficient inference and the ability to operate offline make these solutions highly desirable in bandwidth-limited or privacy-sensitive environments [26], [27].
6) Despite these advancements, challenges remain. Occlusion, lighting variation, cluttered backgrounds, and intra-subject variability continue to hinder the robustness of vision-based systems [28], [29]. Moreover, generalized models often fail when applied across different domains or populations, indicating the need for adaptive, personalized HAR frameworks [30].
7) A promising direction includes the fusion of vision and sensor data to leverage the advantages of both modalities [31], [32]. This not only enhances accuracy but also ensures resilience in diverse conditions. Additionally, self-supervised learning and transfer learning are being explored to reduce the dependence on large annotated datasets [33].
8) Finally, ethical considerations such as data privacy, consent, and secure handling of user data must be prioritized when designing and deploying these systems [34]. As technology continues to evolve, the future of HAR lies in creating scalable, interpretable, and context-aware systems that can operate seamlessly across devices and environments.

IV. FUTURE DIRECTIONS

The landscape of human activity and posture recognition is evolving rapidly, driven by advancements in artificial intelligence, sensor technologies, and real-time computing frameworks. While current systems such as OpenPose and MediaPipe offer high performance in specific environments, there are several areas poised for significant development.

1) Multimodal Sensor Fusion: The future of activity recognition will likely incorporate diverse sensor inputs—video, accelerometer, gyroscope, ECG, EEG, and GPS—to enhance accuracy and robustness. Multimodal learning can mitigate weaknesses of individual modalities and enable context-aware predictions in noisy environments [17], [28].
2) Lightweight and Real-Time Systems for Edge Devices: AI models must continue to be optimized for low latency and real-time performance on mobile and edge devices. Frameworks like MediaPipe, which can implement, demonstrate how efficient models can run on smartphones and microcontrollers. Future models will further reduce computational overhead using quantization, pruning, and efficient neural architectures such as MobileNets [19], [21].
3) Personalized and Adaptive Models: Activity trackers will shift towards personalized AI systems that learn and adapt to individual behavioral patterns over time. Federated learning and continual learning techniques will allow models to update on-device without compromising privacy [29], [30].

4) Robustness in Unconstrained Environments: A key challenge is the variability in real-world conditions, including occlusion, varying illumination, and clothing differences. Future research should develop robust pose estimation systems that perform well in the wild using synthetic data augmentation, domain adaptation, and adversarial training [25].
5) 3D Human Pose Estimation: While 2D pose estimation is widely used, the transition to accurate monocular 3D pose estimation is essential for applications like physiotherapy, AR/VR fitness, and ergonomics. Techniques combining depth estimation, temporal consistency, and transformer-based architectures are expected to grow [5].
6) Integration with Healthcare and Wearables: AI-powered trackers will play a significant role in digital health monitoring, fall detection, elderly care, and rehabilitation. Seamless integration with wearable devices, health records, and IoT platforms will enable predictive analytics and personalized healthcare insights [13], [24].
7) Privacy-Preserving Models: With increasing data concerns, future trackers will embed privacy-by-design principles. Models will rely on on-device processing, homomorphic encryption, or differential privacy to ensure that sensitive user data is never exposed externally [30].
8) Explainable AI (XAI) in Activity Recognition: To build trust and regulatory compliance, especially in clinical and safety-critical applications, activity recognition systems will require explainability and transparency. Integrating interpretable features and visualizations can make decisions more understandable to users and stakeholders.
9) Open-Source Benchmarks and Standardization: There is a need for standardized benchmarking datasets and evaluation protocols to ensure reproducibility and fair comparison of models. Public repositories and community-driven platforms will foster collaboration and rapid innovation in this domain.
10) Context-Aware and Emotion-Aware Activity Recognition: Future systems will not only detect physical activities but also infer intentions, emotions, and social interactions, enriching the context of human behaviour understanding. Multitask learning and affective computing will play a central role in this next phase.

V. CONCLUSION

Human activity and posture recognition has emerged as a critical area in computer vision and AI, offering wide-ranging applications in healthcare, fitness tracking, surveillance, assistive technology, and human-computer interaction. This review explored a broad spectrum of approaches used in this domain, with a specific focus on comparing traditional Convolutional Neural Network (CNN)-based methods and cutting-edge frameworks such as OpenPose and MediaPipe. Through a critical analysis of 30 IEEE research papers, we evaluated the methodological advances, performance benchmarks, and trade-offs across various models.

The integration of MediaPipe in your own AI-powered human activity tracker demonstrates the capabilities of lightweight and real-time frameworks in solving real-world problems, particularly in mobile and embedded environments. These innovations have significantly reduced computational complexity while maintaining high accuracy in pose detection and activity classification.

However, challenges remain, including the need for generalization in unconstrained environments, privacy preservation, and the integration of multiple data modalities for deeper contextual understanding. Furthermore, there is a growing need to move toward personalized and explainable AI systems that can adapt to individual users while providing transparency in decision-making processes.

As research continues to progress, the field is poised for impactful advancements—especially in combining 3D pose estimation, sensor fusion, and real-time analytics on edge devices. The future of human activity tracking lies in scalable, efficient, and human-centric solutions that can seamlessly integrate into daily life, empowering users with actionable insights and fostering better health, safety, and productivity outcomes. From this comparative study, it is evident that:

1) OpenPose offers high accuracy for multi-person setups but isn't practical for real-time mobile use.
2) CNN-based HAR works well with sensor data and is suitable for structured datasets.
3) MediaPipe and BlazePose provide a balanced trade-off between speed and accuracy on smartphones.
4) Lightweight models shine in low-resource environments and embedded systems but may need ensembling for accuracy boosts.

REFERENCES

[1] A. Bevilacqua et al., "Human Activity Recognition with Convolutional Neural Networks," arXiv preprint arXiv:1906.01935, 2019.
[2] M. Zeng et al., "Semi-Supervised Convolutional Neural Networks for Human Activity Recognition," arXiv preprint arXiv:1801.07827, 2018.
[3] Y. Tang et al., "Layer-wise training convolutional neural networks with smaller filters for human activity recognition using wearable sensors," arXiv preprint arXiv:2005.03948, 2020.
[4] N. Rashid et al., "AHAR: Adaptive CNN for Energy-efficient Human Activity Recognition in Low-power Edge Devices," arXiv preprint arXiv:2102.01875, 2021.

[5] 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.
[6] G. Papandreou, T. Zhu, N. Kanazawa, et al., "Towards Accurate Multi-Person Pose Estimation in the Wild," in Proc. CVPR, 2017, pp. 3711–3719.
[7] A. Toshev and C. Szegedy, "DeepPose: Human Pose Estimation via Deep Neural Networks," in Proc. CVPR, 2014, pp. 1653–1660.
[8] S. Li, Z. Liu, and S. Lian, "Convolutional Neural Network-Based Human Pose Estimation: A Review," IEEE Access, vol. 6, pp. 59024–59036, 2018.
[9] D. Mehta et al., "VNect: Real-time 3D Human Pose Estimation with a Single RGB Camera," ACM Trans. Graph., vol. 36, no. 4, 2017.
[10] S. Moon, H. Chang, and Y. Choi, "PoseFix: Model-agnostic General Human Pose Refinement Network," in Proc. CVPR, 2019, pp. 7773–7781.
[11] T.-Y. Lin et al., "Focal Loss for Dense Object Detection," in Proc. ICCV, 2017, pp. 2980–2988.
[12] X. Zhang et al., "ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices," in Proc. CVPR, 2018, pp. 6848–6856.
[13] M. Zeng, L. T. Nguyen, B. Yu, O. J. Mengshoel, J. Zhu, and P. Wu, "Convolutional Neural Networks for Human Activity Recognition Using Mobile Sensors," in Proc. MobiCASE, 2014, pp. 197–205.
[14] L. Wang, Y. Qiao, and X. Tang, "Action Recognition with Trajectory-Pooled Deep-Convolutional Descriptors," in Proc. CVPR, 2015, pp. 4305–4314.
[15] Y. Chen and Y. Xue, "A Deep Learning Approach to Human Activity Recognition Based on Single Accelerometer," in Proc. IEEE SMC, 2015, pp. 1488–1492.
[16] H. Ordonez and D. Roggen, "Deep Convolutional and LSTM Recurrent Neural Networks for Multimodal Wearable Activity Recognition," Sensors, vol. 16, no. 1, p. 115, 2016.
[17] R. Okita and M. Shibata, "Real-time human activity recognition using convolutional neural network and wearable motion sensors," in Proc. IEEE GCCE, 2018, pp. 117–120.
[18] S. Ha and S. Choi, "Convolutional Neural Networks for Human Activity Recognition using Multiple Accelerometer and Gyroscope Sensors," in Proc. IEEE IEMCON, 2016, pp. 1–4.
[19] B. Hammerla, S. Halloran, and T. Plotz, "Deep, Convolutional, and Recurrent Models for Human Activity Recognition Using Wearables," arXiv preprint arXiv:1604.08880, 2016.
[20] Y. Chen, Y. Shen, and K. Zhang, "DeepSense: A Unified Deep Learning Framework for Time-Series Mobile Sensing Data Processing," in Proc. WWW, 2017, pp. 351–360.
[21] F. Jiang and C. Yin, "HAR-Net: A Hierarchical Attention Representation Network for Human Activity Recognition," IEEE Access, vol. 8, pp. 10876–10885, 2020.
[22] A. Haque, A. Alahi, and L. Fei-Fei, "Recurrent Attention Models for Depth-Based Person Identification," in Proc. CVPR, 2016, pp. 1229–1238.
[23] A. G. Howard et al., "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," arXiv preprint arXiv:1704.04861, 2017.
[24] F. Iandola et al., "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size," arXiv preprint arXiv:1602.07360, 2016.
[25] H. Jin, J. Yu, and L. Wang, "TinyPose: Real-time and Accurate Human Pose Estimation on Mobile Devices," in Proc. ECCV, 2020.
[26] N. Rashid et al., "AHAR: Adaptive CNN for Energy-efficient Human Activity Recognition in Low-power Edge Devices," IEEE Internet Things J., vol. 9, no. 3, pp. 1902–1914, Feb. 2022.
[27] S. Yi, C. Li, and Q. Zhang, "A Survey of Edge Computing-Based Human Activity Recognition," IEEE Internet Things J., vol. 8, no. 12, pp. 9789–9800, Jun. 2021.
[28] J. Yoon, S. Kim, and J. Kim, "Robust Human Activity Recognition against Sensor Displacement Using Convolutional Neural Networks," Sensors, vol. 19, no. 20, p. 4560, 2019.
[29] S. Wang, L. Zhou, and Z. Wang, "Robust Pose Estimation in Unconstrained Environments," in Proc. ICPR, 2016, pp. 191–196.
[30] A. Stisen et al., "Smart Devices are Different: Assessing and Mitigating Mobile Sensing Heterogeneities for Activity Recognition," in Proc. ACM SenSys, 2015, pp. 127–140.
[31] M. Zanella et al., "Multimodal Human Activity Recognition Using CNN-LSTM Networks," in Proc. IEEE ISPA, 2020, pp. 1–7.
[32] H. Guan and A. P. Liu, "Multi-modal Sensor Fusion for Human Activity Detection in Smart Homes," Sensors, vol. 20, no. 16, p. 4466, 2020.
[33] R. Ghosh et al., "Self-Supervised Learning for HAR with Wearables," IEEE Sensors J., vol. 22, no. 3, pp. 2191–2200, Feb. 2022.
[34] C. Xu et al., "Privacy-preserving Deep Learning for Sensor-based HAR," IEEE Trans. Mob. Comput., vol. 21, no. 6, pp. 2113–2125, Jun. 2022.
[35] Ronao, C. A., & Cho, S. B. (2016). Human activity recognition with smartphone sensors using deep learning neural networks. Expert Systems with Applications, 59, 235-244. doi:10.1016/j.eswa.2016.04.032