International Journal of Scientific Research & Engineering Trends
Volume 12, Issue 3, May-June-2026, ISSN (Online): 2395-566X

Human Activity Recognition Using OpenCV

Swami Bhagwat, Shreyash Vidhate, Darshan Shinde, Professor G. K. Gaikwad
Dept. of Artificial Intelligence and Data Science MET Institute of Engineering, Nashik, Maharashtra, India

Abstract — Human Activity Recognition (HAR) focuses on automatically identifying human actions from video streams or sensor data using computer vision and machine learning techniques. With the rapid growth of intelligent healthcare, surveillance, and smart automation systems, HAR has become an important research area. This paper presents a redesigned and implementation-oriented study of a HAR system built using OpenCV and modern learning models. The work explains the complete pipeline including video acquisition, preprocessing, feature extraction, and activity classification. Instead of relying only on theoretical descriptions, the paper emphasizes a practical modular architecture and real-time considerations. The role of deep learning models combined with OpenCV preprocessing is discussed along with system challenges such as lighting variation, occlusion, and computational cost. The proposed approach highlights how lightweight processing and hybrid models can support accurate and efficient recognition suitable for real-world deployment.

Keywords— Human Activity Recognition (HAR), OpenCV, Machine Learning, Computer Vision, Real-Time Detection, Deep Learning.

I. INTRODUCTION

Human Activity Recognition (HAR) has become an im- portant research area that lies at the intersection of com- puter vision, artificial intelligence, and ubiquitous computing systems. It focuses on building computational models that can automatically identify and categorize human activities by analyzing data obtained from cameras, video streams, or wearable motion sensors. With the growing adoption of automated and continuously monitored environments, HAR is increasingly relevant in many practical domains. These include healthcare systems for continuous patient observa- tion and fall detection, intelligent surveillance platforms for behavior monitoring, and smart home ecosystems that adapt device behavior based on user activity patterns. The main goal of HAR is not limited to detecting isolated events, but to enable continuous and reliable understanding of human behavior in real time, thereby improving operational safety, system efficiency, and human–machine interaction quality. The development of practical HAR solutions, particularly for

Identify applicable funding agency here. If none, delete this.

edge or low-resource devices, depends heavily on efficient computer vision tools. One of the most widely used libraries in this context is OpenCV, which offers optimized functions for video and image processing. It supports core operations such as frame preprocessing, foreground extraction, motion tracking, and feature computation. Using these capabilities, raw video input can be transformed into structured visual descriptors that are suitable for further learning-based analysis. This preprocessing stage plays a crucial role in reducing noise and computational overhead before classification. Approaches used in Human Activity Recognition (HAR) have changed considerably over time. Early methods primarily relied on manually crafted features such as optical flow patterns, body silhouettes, and geometric motion descriptors. While these techniques performed well in controlled environments, they often faced difficulties under real-world conditions such as changes in lighting, occlusions, and variations in camera viewpoints. To address these challenges, later HAR systems focused on improved feature extraction, motion analysis, and temporal modeling techniques that capture both spatial and time-based characteristics of human actions. In many implementations, OpenCV-based preprocessing is used to enhance visual data quality and support efficient activity recognition pipelines suitable for practical applications. Even with these advancements, deploying HAR systems in real environments remains challenging. Variations in how different individuals perform the same activity can reduce classification consistency. In addition, many high-accuracy models require substantial computational resources, making real-time execution on lightweight devices difficult. Another emerging concern is model transparency, since practical applications often require interpretable predictions. Therefore, ongoing research is focused not only on improving accuracy but also on enhancing efficiency, explainability, and adaptability. This work aims to review the progression of HAR techniques, highlight the contribution of OpenCV in system design, and outline cur- rent limitations and future research opportunities for building dependable and scalable HAR solutions.

© 2026 IJSRET

II. LITERATURE REVIEW INTRODUCTION

Human Activity Recognition (HAR) has been an active area of research for more than a decade, gradually advancing from basic motion detection to more detailed activity classification methods. Early studies primarily relied on traditional computer vision techniques that used manually designed features, while later approaches focused on improved analytical and pattern-based methods for recognizing human activities. This section reviews important studies and methodologies related to HAR, describing their techniques, datasets, strengths, and limitations. A. Early Vision-Based Approaches Initial HAR models re- lied heavily on low-level feature extraction methods such as motion history images, optical flow, and silhouette tracking. For example, traditional methods used background subtraction and contour-based analysis to differentiate between human postures. Although these approaches performed well under controlled environments, their accuracy degraded significantly under changing lighting conditions, occlusion, and complex backgrounds [6].

B. Machine Learning-Based HAR Models With the introduction of machine learning, algorithms like Support Vector Machines (SVM), k-Nearest Neighbors (k- NN), and Hidden Markov Models (HMM) were applied to classify extracted motion features. These systems showed better generalization capabilities and required less computa- tional power compared to deep learning methods. However, they still struggled with dynamic background variations and large-scale datasets. C. Deep Learning and Hybrid Models Recent developments in deep learning have transformed HAR systems. Convolutional Neural Networks (CNNs) are used to extract spatial features from frames, while Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks model temporal dependencies between video frames. Researchers have also combined OpenCV-based preprocessing with CNN-LSTM architectures for enhanced accuracy. Such hybrid systems are capable of real-time recognition, robust against environmental changes, and adaptable to multiple camera perspectives. D. Role of OpenCV in Modern HAR OpenCV continues to play a crucial role as a preprocessing and visualization tool in HAR pipelines. It enables real- time video capture, frame segmentation, and skeletal tracking using modules such as BackgroundSubtractorMOG2, Hough Transform, and Pose Estimation APIs. Researchers commonly integrate OpenCV for feature extraction and feed the processed data into deep learning models, achieving high recognition rates in healthcare and surveillance applications.

III. MODULES AND BACKGROUND

AA Proposed Core Modules: MODULES AND BACK- GROUND The proposed Human Activity Recognition (HAR) system using OpenCV is designed with several integrated modules, each responsible for a specific function in the activity detection and classification process. The following are the key core modules of the system: 1) Video Acquisition Module: Captures live video input from a webcam or CCTV camera. Ensure real-time video capture at a consistent frame rate. 2)

Preprocessing Module: Enhances the video frames by convert- ing them to grayscale and reducing noise. Normalizes lighting variations and resizes frames for consistent model input. 3) Feature Extraction Module: Identifies important visual and motion-based features from frames using OpenCV techniques like edge detection, contour extraction, and key point iden- tification. Prepares feature vectors that represent the activity in numerical form. 4) Activity Classification Module: Applies a machine learning or deep learning model to classify activ- ities such as walking, sitting, standing, or falling. Integrates OpenCV's real-time analysis tools with trained models for immediate prediction. 5) Data Storage and Analysis Module: Stores classified activity data and timestamps in a local or cloud-based database. Supports visualization of patient activity trends and performance reports. 6) User Interface Module: Provides a graphical interface for caregivers or healthcare staff to monitor live video feeds and activity alerts. Offers options to review past logs, adjust system settings, and manage users. The deployment diagram illustrates how the Human Activity Recognition sys- tem is physically deployed across hardware nodes and devices. It helps visualize the distribution of software components on hardware, showing how the system operates in a real-world environment. The main nodes include the Camera module for capturing video, an Edge Processing Unit or Server for preprocessing, feature extraction, and ac- tivity clas- sification, and storage units for logging activity data. Caregivers and administrators ac- cess the system through client devices such as computers, tablets, or smartphones. The diagram highlights communication channels, data flow, and deployment dependencies, ensuring proper system setup, scalability, and performance.

1. Background

Human Activity Recognition (HAR) is widely recognized as an important research direction within computer vision and artificial intelligence because it enables machines to interpret human behavior automatically. Its growing relevance comes from practical needs in areas such as medical monitoring, security analytics, and intelligent interactive systems. The primary purpose of HAR is to detect and label human actions

© 2026 IJSRET

using visual streams or sensor measurements so that computer systems can react appropriately to user behavior. This section explains the foundations of HAR, its system pipeline, and how OpenCV contributes to building efficient recognition frameworks.

2. Human Activity Recognition Overview

HAR is an interdisciplinary domain that brings together image analysis, motion understanding, and learning-based prediction models to recognize physical actions like walking, sitting, running, standing, or falling. A standard HAR pipeline generally includes three core stages: capturing input data, deriving meaningful features, and performing activity prediction. Input may come from cameras or wearable sensors and is first cleaned and normalized to reduce noise and inconsistencies. The system then derives motion and posture indicators using techniques such as shape boundaries, pose estimation, and motion vector tracking. These derived descriptors are converted into structured representations that learning algorithms can interpret. Finally, trained models map these representations to specific activity labels based on learned patterns.

3. Importance of OpenCV in HAR

OpenCV is a widely adopted open-source computer vi- sion toolkit that supports fast and efficient image and video operations. It includes optimized routines for frame transformation, foreground separation, motion highlight- ing, boundary detection, and object localization. Within HAR pipelines, OpenCV is mainly used at the prepro- cessing and feature preparation stages. It enables con- tinuous frame handling, tracking of moving subjects, and extraction of focused regions that contain relevant motion information. These steps reduce unnecessary background data and improve downstream classification reliability. OpenCV can also be connected with modern learning libraries, allowing developers to build mixed pipelines where classical vision processing is followed by neural model inference.

4. Role of Machine Learning and Deep Learning

Early HAR implementations often relied on traditional machine learning classifiers trained on manually designed features. Methods such as Support Vector Machines, Decision Trees, and ensemble models were commonly applied when feature vectors were explicitly engineered. While effective in constrained environments, their perfor- mance depended strongly on feature quality. More recent HAR systems increasingly use deep learning approaches that learn feature representations directly from raw inputs. Convolutional Neural Networks are effective for extract- ing spatial structure from images, whereas sequence- oriented models such as recurrent networks and LSTM units capture time-based dependencies across frame sequences. When these models are used together with OpenCV-based preprocessing, they can support stable real-time recognition even when the environment contains noise or visual variation.

5. Applications of HAR

Activity recognition technology supports many real-world use cases. In medical settings, it is used to observe patient movement patterns and automatically detect dangerous events such as falls. Security systems apply HAR to ana- lyze behavior and flag unusual actions. Smart home plat- forms use activity understanding to adjust device behavior based on user routines. Additional uses include athletic performance analysis, physical rehabilitation monitoring, and industrial safety supervision. The wide range of deployment scenarios shows that HAR is useful in both personal and enterprise environments.

6. Research Challenges in HAR

Although HAR methods have improved considerably, several technical issues still limit consistent performance.

Visual recognition can be affected by illumination dif- ferences, partial visibility of subjects, dynamic back- grounds, and variations in body shape or movement style. Achieving real-time processing adds further constraints because models must operate efficiently on limited hard- ware resources. Another ongoing difficulty is ensuring that trained systems work reliably across different users and environments rather than only under controlled con- ditions. Addressing these limitations remains an active research focus and motivates continued innovation in model design and data strategies.

IV. ANALYSIS AND DISCUSSION

The evaluation of modern Human Activity Recognition systems shows that the field has advanced considerably in terms of prediction accuracy, processing speed, and deployment capability. At the same time, system perfor- mance is strongly influenced by the selection of learning models, the quality of extracted features, and the effec- tiveness of preprocessing steps. Different methodological families — rule based, classical machine learning, and deep learning — demonstrate distinct trade-offs. This section examines these approaches from a performance and design perspective and explains how OpenCV-based processing contributes to practical efficiency and flexibil- ity. A. Comparison of HAR Methodologies Early activity recognition solutions depended mainly on manually de- signed motion descriptors, including flow vectors, fore- ground masks, and contour-based representations. These techniques are relatively inexpensive in terms of com- putation and can be implemented quickly using

© 2026 IJSRET

standard OpenCV operations. However, their behavior is unstable when environmental conditions change, such as variations in brightness, camera viewpoint, or partial subject visi- bility. The next stage of development introduced super- vised machine learning classifiers trained on engineered feature vectors. Methods like Support Vector Machines and tree-based classifiers improved label prediction by learning decision boundaries from annotated samples. Even so, their success remained closely tied to how well the input features were constructed and normalized. More recent deep neural approaches learn representa- tions directly from frame sequences instead of depending on handcrafted descriptors. Convolution-based models capture spatial structure, while sequence-aware architec- tures such as CNN–LSTM combinations learn motion evolution across time. These models generally produce higher recognition reliability because both appearance and temporal dynamics are learned automatically. When supported by OpenCV preprocessing steps — such as subject isolation, pose-region extraction, and noise fil- tering — deep models achieve stronger and more sta- ble real-time results. Experimental comparisons reported in multiple studies show that hybrid deep architectures typically reach accuracy in the low-to-mid 90% range, while classical pipelines often remain notably lower under comparable testing conditions. B. Contribution of OpenCV to System Efficiency OpenCV plays a practi- cal enabling role in many HAR pipelines by handling the low-level visual processing workload. Its optimized routines support fast frame reading, transformation, seg- mentation, and motion-focused filtering. By performing these operations before model inference, the amount of irrelevant visual information is reduced, which lowers the burden on the classifier. Another advantage is interoper- ability. OpenCV can be easily combined with modern learning frameworks, enabling layered pipelines where deterministic vision processing prepares the data and neural networks perform final recognition. This staged design improves throughput and stability, especially in environments where immediate response is required, such as assisted living spaces or safety monitoring systems. C. Real-Time Processing Tradeoffs Operational HAR sys- tems often need to deliver predictions with minimal delay. Lightweight vision pipelines based mainly on classical OpenCV operations can run quickly even on modest hardware, but their recognition precision may be limited in complex scenes. Deep neural models typically provide stronger predictive performance, yet they require greater computational resources. To balance these factors, devel- opers increasingly rely on optimized network families and hardware acceleration. Compact architectures such as MobileNet-style or efficiency-focused networks reduce parameter count while preserving acceptable accuracy. GPU support and model quantization further improve inference speed. Additional gains are achieved by re- ducing input complexity through preprocessing steps like foreground extraction and region cropping, which shorten model processing time without major information loss.

D. Observed Limitations and Open Problems Even with modern techniques, HAR systems are not free from weaknesses. Recognition reliability can still degrade due to shadows, cluttered backgrounds, partial occlusion, and differences in how individuals perform the same action. Another recurring issue is dataset bias — many models are trained on limited or staged datasets and fail to generalize well in uncontrolled environments. Scaling HAR solutions also introduces engineering challenges. Multi-camera deployments and distributed sensing setups require coordinated timing, consistent calibration, and efficient data handling. Without these, prediction quality and system stability may drop. Improving cross envi- ronment robustness will require better training diversity, domain adaptation strategies, and stronger augmentation methods.

V. CONCLUSION

Human Activity Recognition has developed into a key enabling technology for systems that need to interpret hu- man behavior automatically from visual or sensor-based data. Its impact is clearly visible across domains such as medical supervision, intelligent security monitoring, and automated living environments. This work reviewed how HAR approaches have progressed from manually engineered motion features to data-driven deep learning models that can analyze both spatial and temporal pat- terns in activity sequences. The discussion highlighted how OpenCV serves as a practical foundation for building such systems by supporting fast video preprocessing, motion isolation, and feature preparation. The combi- nation of OpenCV with modern learning frameworks has made it easier to design recognition pipelines that are both accurate and deployable in near real-time set- tings. In particular, layered models that join convolution- based feature learning with temporal sequence mod- eling have shown strong performance in recognizing complex actions from continuous video streams. These hybrid strategies demonstrate that careful integration of preprocessing and learning components leads to more stable and responsive HAR solutions. At the same time, several practical barriers remain. Model robustness is still affected by scene variation, partial visibility, and differences in execution style across users. Resource requirements of advanced neural models also make edge deployment challenging. Future improvements are likely to come from compact model architectures, cross-modal learning that combines video with wearable sensor data, and adaptive training techniques that

© 2026 IJSRET

generalize better across environments. Overall, HAR systems built with efficient vision libraries and intelligent learning mod- els are moving toward wider real-world adoption. With continued progress in optimization, interpretability, and multimodal design, these systems are expected to become a core component of next-generation smart and safety- aware digital environments.

Future Scope

The field of Human Activity Recognition (HAR) is rapidly advancing, yet several opportunities remain for improvement and innovation. Future research should fo- cus on enhancing system adaptability, accuracy, and real- time performance, particularly for healthcare and assistive technologies. The integration of OpenCV with emerging technologies such as deep learning, IoT, and edge com- puting will play a critical role in shaping next-generation HAR systems [4]. A. Integration with Internet of Things (IoT) and Edge Devices One of the most promising future directions is the integration of HAR systems with IoT and edge computing platforms. Deploying HAR algorithms directly on edge devices such as smart cameras or micro- controllers can significantly reduce latency and network dependency. This enables real-time activity monitoring in healthcare facilities, homes, and public areas. Combining OpenCV's lightweight image processing capabilities with IoT sensors can further enhance accuracy by providing multi-modal data such as motion, temperature, and vital signs [3]. B. Improved Dataset Diversity and Generaliza- tion Many current HAR models perform well only under

controlled conditions due to limited dataset diversity. Future work should focus on developing large, heteroge- neous datasets that include variations in human posture, lighting, background, and camera angles. Synthetic data generation using computer graphics or generative adver- sarial networks (GANs) can help address data scarcity and improve model generalization across different en- vironments and users [2]. C. Multimodal and Context- Aware Recognition Systems Next-generation HAR sys- tems are expected to be multimodal—combining video data with other sensor inputs such as accelerometers, gyroscopes, or wearable sensors. Context-aware HAR models could adapt to user behavior, surroundings, and activity context, providing intelligent decision-making capabilities. For example, a healthcare HAR system could distinguish between normal lying down and accidental falling by combining motion data with contextual cues such as time, heart rate, or patient activity logs [3]. D. Use of Explainable AI (XAI) in HAR As HAR systems increasingly influence critical domains like healthcare and security, transparency in decision-making becomes essential. Future research should incorporate Explainable AI (XAI) frameworks to provide visual or textual expla- nations for model predictions. This will help users and medical professionals understand why a certain activity was detected, thereby improving trust and accountability in automated systems [4]. E. Optimization for Real- Time and Low-Power Devices Although deep learning- based HAR systems provide high accuracy, they often require significant computational resources. Future work must focus on model optimization through techniques like pruning, quantization, and knowledge distillation to make HAR systems more efficient on low-power devices. Integration with OpenCV's hardware acceleration and GPU support will further enhance real-time performance without compromising accuracy [5]. Expansion Toward Human Emotion and Gesture Recognition: Beyond basic activities like walking or sitting, future HAR systems could extend to recognizing human emotions and com- plex gestures. Combining facial expression analysis with body movement tracking using OpenCV and deep learn- ing can create emotionally intelligent systems capable of interacting naturally with humans. Such systems will find applications in mental health monitoring, robotics, and adaptive user interfaces [6].

REFERENCES

1. J. Wang, Y. Chen, S. Hao, X. Peng, and L. Hu, "Deep Learning for Sensor-Based Human Activity Recognition: Overview, Challenges and Opportu- nities," Pattern Recognition Letters, vol. 119, pp. 3–11, 2019.
2. S. Bhattacharya and P. Das, "Real-Time Human Activity Recognition Using OpenCV and Convo- lutional Neural Networks," Journal of Ambient Intelligence and Humanized Computing, vol. 11, no. 9, pp. 3571–3582, 2020.
3. Krishna, S. B. Siva; Reddy, K. Rusheendra; Satish, V.; Babu, E. Anka; Vinod, G. V., "Human Activity Recognition With OpenCV and Deep Learning," International Journal of Engineering Research & Technology (IJERT), vol. 13, issue 02, Feb. 2024.
4. Putra, I. A.; Nurhayati, O. D.; Eridani, D., "Hu- man Action Recognition (HAR) Classification Us- ing MediaPipe and Long Short-Term Memory (LSTM)," TEKNIK Journal, vol. 43, no. 2, 2022.
5. Host, K. and Ivasˇic´-Kos, M., "An Overview of Human Action Recognition in Sports Based on Computer Vision," Heliyon, 2022.
6. Janapati, M.; Allamsetty, L. P.; Potluri, T. T.; Mogili, K. V., "Gait-Driven Pose Tracking and Movement Captioning Using OpenCV and Medi- aPipe Machine Learning Framework," Engineering Proceedings (MDPI), 2024.
7. Gupta, S.; Ramya, K. R.; Karnati, R., "A Re- view Work: Human Action Recognition in Video Surveillance Using

Deep Learning Techniques," Informatics & Automation, vol. 23, no. 2, 2024.

8. Wang, X.; Wu, Z.; Jiang, B.; Bao, Z.; Zhu, L.; Li, G.; Wang, Y.; Tian, Y., "HARDVS: Revisiting Human Activity Recognition with Dynamic Vision Sensors," arXiv Preprint, 2022.
9. R. Singh and A. Gupta, "Human Activity Recog- nition Using Multi-Camera OpenCV Framework," International Journal of Computer Applications, vol. 183, no. 35, pp. 1–6, 2021.

© 2026 IJSRET