ISSN 2277-2685
IJESR/April. 2026/ Vol-16/Issue-2s1/226-231
Shahbaz Ahmed *et. al.*, /International Journal of Engineering & Science Research
Full Length Article

Smart Human Action Monitoring Using RGB And Motion Signals

Shahbaz Ahmed[1], Asim Ahmed Khan[2], Syed Maqdoom Masroor Abrar[3], Mrs. Heena Yasmin[4]
[1,2,3]B.E.Students; Dept. of CSE ISL Engineering College, Hyderabad India.
[4]Assistant Professor; Dept. of CSE ISL Engineering College, Hyderabad India.
Mail Id; shahbazahmed11004@gmail.com, ahmedkhan23607@gmail.com, sm.abrar13@gmail.com

Accepted 27-04-2026

ABSTRACT
Human action recognition is a vital component of smart monitoring systems, with applications in healthcare, surveillance, and intelligent indoor environments. This project presents a smart human action monitoring system that relies exclusively on RGB images to detect and classify human activities in real time. The system processes image frames to extract essential features such as body posture, joint positions, and movement patterns. These features are then analyzed to recognize common human actions, including sitting, walking, running, drinking, and other daily activities. By leveraging advanced image processing and deep learning techniques, the system achieves high accuracy, robustness to varying lighting conditions, and efficiency suitable for real-time deployment. Experimental results demonstrate the system's ability to monitor human activity reliably, providing a practical solution for indoor action recognition and smart environment applications.

Keywords:
Human Action Recognition, RGB Image Processing, Smart Monitoring System, Deep Learning, Real-Time Activity Detection, Human Activity Classification, Computer Vision, Indoor Surveillance, Image-Based Recognition, Motion Analysis, Pose Estimation, Intelligent Monitoring Systems, Activity Recognition, Artificial Intelligence.

INTRODUCTION

Human Action Recognition (HAR) has become an essential area of research in computer vision due to its wide range of applications in healthcare, security surveillance, elderly monitoring, and intelligent indoor environments. The ability to automatically identify and classify human activities from visual data provides significant advantages in building smart systems that can support safety, automation, and human–computer interaction. For instance, monitoring the daily activities of elderly individuals can help detect abnormal behaviors, while in surveillance, action recognition enables the identification of suspicious activities in real time.

Traditional approaches to HAR often rely on wearable sensors, depth cameras, or motion capture systems to track body movements. While these methods can provide reliable information, they suffer from limitations such as high cost, intrusiveness, and dependency on specialized hardware. In contrast, RGB cameras offer a low-cost and non-intrusive solution that can be easily deployed in indoor environments. However, recognizing human activities from RGB images alone is challenging due to factors like varying lighting conditions, occlusions, and diverse human motion patterns.

Recent advancements in image processing and deep learning have significantly improved the performance of RGB-based HAR systems. By extracting key visual features such as body posture, joint positions, and temporal movement patterns, deep learning models can achieve high accuracy and robustness in action classification tasks. These improvements open the door to practical real-time monitoring solutions suitable for everyday environments.

In this work, we present a smart human action monitoring system that leverages RGB images to detect and classify human activities in real time. The proposed system is designed to recognize common daily actions such as sitting, walking, running, and drinking, with an emphasis on efficiency and robustness under different environmental conditions. By integrating advanced feature extraction techniques with deep learning models, our approach provides a reliable and practical framework for indoor action recognition, contributing to the development of intelligent and adaptive monitoring systems.

SCOPE OF THE PROJECT

The scope of this project is to design and implement a smart human action recognition system that utilizes RGB images for real-time monitoring in

indoor environments. The system focuses on detecting and classifying common human activities such as sitting, walking, running, and drinking, which are essential for healthcare assistance, surveillance, and smart home automation. By relying exclusively on RGB cameras, the project ensures a cost-effective, non-intrusive, and widely deployable solution compared to sensor-based or depth-camera approaches. The scope also covers the development of feature extraction techniques to analyze body postures, joint positions, and motion patterns, along with the integration of deep learning models to achieve high accuracy and robustness under varying lighting and environmental conditions. Furthermore, the project aims to demonstrate real-time performance, making it suitable for practical deployment in smart monitoring systems. However, the scope is limited to recognizing predefined daily activities and indoor scenarios, leaving room for future expansion into more complex outdoor environments and a broader set of human actions.

OBJECTIVE

The primary objective of this project is to develop an intelligent human action recognition system that can accurately detect and classify daily human activities in real time using only RGB images. The system is designed to provide a low-cost, non-intrusive, and efficient solution for smart monitoring applications, particularly in healthcare, surveillance, and indoor automation. By leveraging advanced image processing techniques and deep learning models, the project aims to extract meaningful features such as body posture, joint positions, and movement patterns to achieve reliable action recognition. The objective is not only to ensure high accuracy and robustness under varying environmental and lighting conditions but also to demonstrate the feasibility of deploying such a system in real-world scenarios where continuous monitoring and timely recognition of human activities are essential.

EXISTING SYSTEM:

In the existing system, Support Vector Machine (SVM) is employed as a primary classification technique for human action recognition. SVM works by finding the optimal hyperplane that separates different classes of human actions based on extracted features such as body posture and motion patterns. It is widely used in machine learning due to its effectiveness in handling high-dimensional data and its strong generalization ability. For human action recognition, SVM can achieve decent accuracy when the features are well-represented; however, its performance often depends heavily on the choice of kernel functions and feature engineering, which may limit its scalability to complex and dynamic environments.

Alongside SVM, the Random Forest (RF) classifier is also utilized as part of the existing system. Random Forest is an ensemble learning method that constructs multiple decision trees and aggregates their results to improve prediction accuracy. In the context of human action recognition, RF can effectively handle feature variability and reduce the risk of overfitting, making it a robust choice for classifying activities from extracted image features. Its ability to deal with noisy data and provide feature importance rankings adds further value to the recognition process.

While the combination of SVM and Random Forest provides a strong baseline for classification, the existing system faces certain limitations. These approaches rely heavily on handcrafted features, which may not fully capture the temporal and spatial complexities of human actions in real-world scenarios. Additionally, their performance tends to degrade under challenging conditions such as varying lighting, occlusions, or complex backgrounds. As a result, the existing system lacks the adaptability and scalability required for real-time human action recognition, highlighting the need for deep learning-based approaches that can automatically learn rich feature representations directly from raw RGB images.

LITERATURE REVIEW

Human Activity Recognition (HAR) is a branch of computer science that uses raw time-series data information from embedded smartphone sensors and wearable devices to infer human actions. It has aroused considerable interest in various smart home contexts, particularly for constantly monitoring human behavior in an ecologically friendly atmosphere for elderly people and rehabilitation. Data collection, feature extraction from noise and distortion, feature selection, and pre-processing and categorization are among the operating components of a typical HAR system. Extraction of feature and selection strategies have recently been developed using cutting-edge approaches and traditional machine learning classifiers. The majority of the solutions, on the other hand, rely on simple feature extraction algorithms that are unable to detect complex behaviors. Deep learning techniques are often utilized in different HAR approaches to recover features and classification swiftly because of the introduction and development of vast computing resources. The vast majority of solutions, on the other hand, depend on simplistic feature extraction algorithms incapable of recognizing complicated behaviors. Due to advancements in high computational capabilities, deep learning algorithms are now often utilized in HAR methods to efficiently extract meaningful features which can successfully categorize sensor data. In this chapter, we present a

hybrid deep learning-based classification model comprising of Convolutional Neural Network (CNN) and Long-Short Term Memory (LSTM), which is named CNN-LSTM. The proposed hybrid deep learning model has been tested over three benchmark HAR datasets: MHEALTH, OPPORTUNITY, and HARTH. On the aforementioned datasets, the proposed hybrid model obtained 99.07%, 95.2%, and 94.68% classification accuracies, respectively, which is quite impressive.

METHODOLOGIES

Modules Name:

- **Data Acquisition Module**
- **Preprocessing Module**
- **Feature Extraction Module**
- **Action Detection & Classification Module**
- **Real-Time Monitoring Module**
- **Result Visualization Module**
- **Performance Evaluation Module**

MODULES EXPLANATION:

1) Data Acquisition Module:
This module is responsible for collecting input data in the form of RGB images or video streams from cameras or datasets. It serves as the foundation of the system by providing continuous visual input for analysis. The module ensures that the captured data maintains sufficient quality for further processing, even under varying environmental conditions. By supporting both live camera feeds and stored video files, the system can be applied in real-time monitoring as well as offline testing.

2) Preprocessing Module:
Before feeding the captured frames into the model, preprocessing is essential to improve recognition accuracy. This module handles tasks such as resizing frames to match the YOLOv8n input dimensions, normalizing pixel values, and enhancing image quality to reduce the impact of lighting variations. Noise removal and frame stabilization may also be applied to ensure consistent input. By standardizing the data, the preprocessing module ensures that the system remains robust across diverse environments and conditions.

3) Feature Extraction Module:
In traditional systems, feature extraction requires manual design, but in this project, the YOLOv8n model automatically extracts meaningful features from RGB frames. This module focuses on identifying spatial features like body posture and joint positions, along with temporal movement patterns. The deep learning backbone of YOLOv8n captures high-level representations of human activity without explicit hand-crafted feature engineering. These learned features form the basis for accurate action detection and classification.

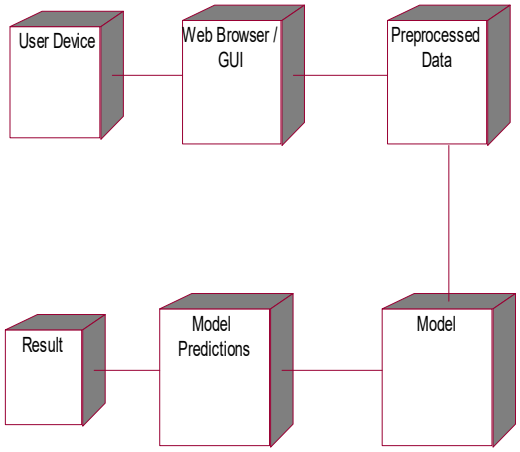
4) Action Detection & Classification Module:
This module is the core of the proposed system. Using YOLOv8n, it detects humans within the input frame and classifies their actions into predefined categories such as sitting, walking, running, and drinking. The model processes frames in real time and assigns bounding boxes, labels, and confidence scores to each detected action. This integration of detection and classification ensures efficiency and simplifies the overall pipeline, making the system highly suitable for real-time applications.

5) Real-Time Monitoring Module:
The real-time monitoring module ensures continuous tracking of human actions for applications like healthcare monitoring, surveillance, and smart home automation. It processes the video feed frame by frame, updating action recognition results instantly. This module is designed to handle multiple individuals simultaneously and adapt to dynamic indoor environments. Its ability to deliver immediate feedback makes it highly practical for safety-critical scenarios, such as fall detection in elderly care or anomaly detection in security surveillance.

6) Result Visualization Module:
To make the system user-friendly and interpretable, the result visualization module overlays detection results directly on the video stream. It displays bounding boxes around individuals, labels for recognized actions, and confidence scores that indicate the reliability of predictions. This visual feedback not only aids in monitoring but also provides transparency in the system's decision-making process. The module can be extended to generate logs, reports, or alerts depending on the application.

7) Performance Evaluation Module:
To make the system user-friendly and interpretable, the result visualization module overlays detection results directly on the video stream. It displays bounding boxes around individuals, labels for recognized actions, and confidence scores that indicate the reliability of predictions. This visual feedback not only aids in monitoring but also provides transparency in the system's decision-making process. The module can be extended to generate logs, reports, or alerts depending on the application.

IMPLEMENTATION



PROPOSED TECHNIQUE USED OR ALGORITHM USED:

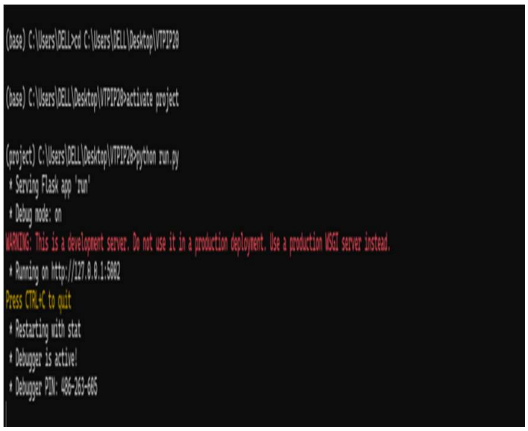
➢ **YOLOv8n:**

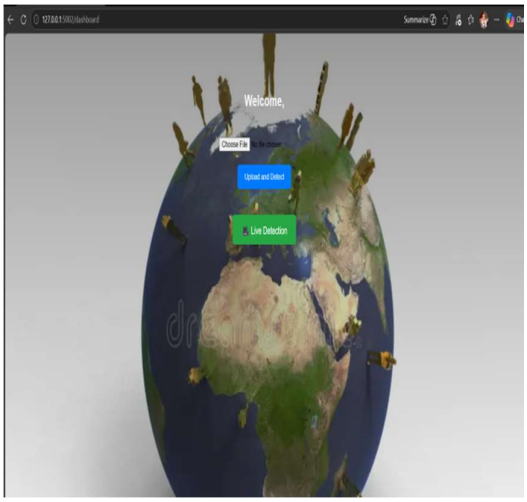
The proposed technique employs YOLOv8n (You Only Look Once, version 8 – nano variant), a state-of-the-art deep learning model, to perform real-time human action recognition using RGB images. Unlike traditional methods that rely on handcrafted features and separate classification models, YOLOv8n integrates feature extraction, detection, and classification into a single end-to-end framework. This allows the system to directly process input frames, identify humans, and classify their actions simultaneously with high efficiency and accuracy.

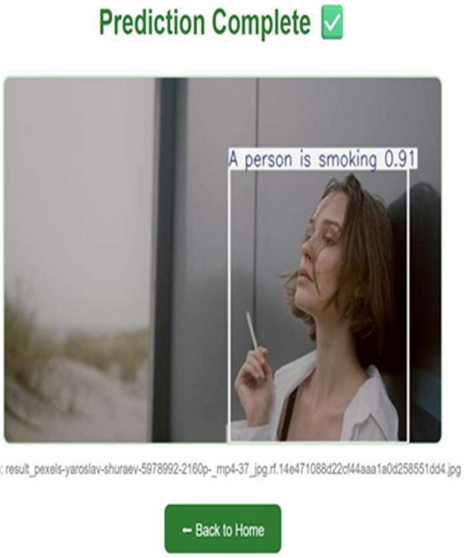
YOLOv8n is designed with a lightweight yet powerful architecture that makes it suitable for real-time applications, even in resource-constrained environments. The model automatically learns spatial and temporal features such as body posture, movement patterns, and joint alignments from the input frames. By leveraging convolutional neural networks (CNNs) and advanced feature representation, it overcomes the limitations of handcrafted features used in existing systems. The nano variant (YOLOv8n) ensures reduced computational overhead while maintaining competitive accuracy, enabling deployment on edge devices like surveillance cameras or embedded systems.

Furthermore, the proposed technique enhances robustness against challenges like varying lighting conditions, background clutter, and partial occlusions. The system outputs bounding boxes, action labels, and confidence scores in real time, making it highly practical for applications in healthcare monitoring, surveillance, and smart indoor environments. By combining speed, accuracy, and adaptability, the YOLOv8n-based approach provides a scalable solution for human action recognition that significantly outperforms traditional machine learning techniques such as SVM and Random Forest.

RESULT SNAPSHOTS







CONCLUSION

In this project, a smart human action recognition system has been developed using **YOLOv8n** to detect and classify human activities in real time based on RGB images. Unlike traditional machine learning techniques such as SVM and Random Forest, which rely on handcrafted features, the proposed system leverages deep learning to automatically extract meaningful spatial and temporal features from image frames. This enables accurate recognition of daily activities like sitting, walking, running, and drinking with high robustness against environmental challenges such as varying lighting, occlusions, and cluttered backgrounds.

The system demonstrates significant improvements in terms of accuracy, efficiency, and adaptability compared to existing methods. Its lightweight architecture ensures real-time performance while maintaining competitive accuracy, making it suitable for deployment in practical scenarios such as healthcare monitoring, smart homes, and surveillance applications.

Overall, the project presents a cost-effective, non-intrusive, and scalable solution for human action recognition. With future enhancements such as support for more complex activities, multimodal data integration, and edge deployment, the system holds great potential to evolve into a comprehensive intelligent monitoring framework capable of addressing real-world challenges effectively.

FUTURE SCOPE:

While the proposed system using YOLOv8n demonstrates high accuracy and efficiency in recognizing common daily human actions, there are several opportunities for future enhancement. One key direction is to expand the system's capability to recognize a wider range of complex and fine-grained activities, such as hand gestures, group interactions, or context-dependent actions, which are essential for more advanced applications in healthcare and smart environments.

Another enhancement could involve integrating multimodal data sources such as depth sensors, thermal cameras, or wearable devices alongside RGB input to improve robustness under challenging conditions like poor lighting or crowded scenes. Additionally, incorporating temporal models such as LSTMs, Transformers, or 3D CNNs on top of YOLOv8n can further improve the system's ability to capture sequential patterns and dependencies in continuous human motion.

Finally, future work may focus on optimizing the system for edge deployment and IoT environments, ensuring low-power operation on embedded devices without sacrificing performance. Features such as real-time alerts, cloud-based analytics, and predictive modeling can be added to make the system more interactive and proactive. These enhancements would make the human action recognition system more versatile, scalable, and effective for real-world intelligent monitoring applications.

REFERENCES

[1] M. Muneeb, H. Rustam, and A. Jalal, ''Automate appliances via gestures recognition for elderly living assistance,'' in Proc. 4th Int. Conf. Advancement Comput. Sci. (ICACS), Feb. 2023, pp. 1–6.

[2] I. A. Abro and A. Jalal, ''Multi-modal sensors fusion for fall detection and action recognition in indoor environment,'' in Proc. 3rd Int. Conf. Emerg. Trends Electr., Control, Telecommun. Eng. (ETECTE), Nov. 2024, pp. 1–6.

[3] A. Nadeem, A. Jalal, and K. Kim, ''Accurate physical activity recognition using multidimensional features and Markov model for smart health fitness,'' Symmetry, vol. 12, no. 11, p. 1766, Oct. 2020.

[4] S. Hafeez, A. Jalal, and S. Kamal, ''Multi-fusion sensors for action recognition based on discriminative motion cues and random forest,'' in Proc. Int. Conf. Commun. Technol. (ComTech), Sep. 2021, pp. 91–96.

[5] A. Jalal, A. Nadeem, and S. Bobasu, ''Human body parts estimation and detection for physical sports movements,'' in Proc. 2nd Int. Conf. Commun., Comput. Digit. Syst. (C-CODE), Mar. 2019, pp. 104–109.

[6] A. Jalal, Y. Kim, and D. Kim, ''Ridge body parts features for human pose estimation and recognition from RGB-D video data,'' in Proc. 5th Int. Conf. Comput., Commun. Netw. Technol. (ICCCNT), Jul. 2014, pp. 1–6.

[7] A. Nadeem, A. Jalal, and K. Kim, ''Human actions tracking and recognition based on body parts detection via artificial neural network,'' in Proc. 3rd

Int. Conf. Advancements Comput. Sci. (ICACS), Feb. 2020, pp. 1–6.
[8] M. A. K. Quaid and A. Jalal, ‘‘Wearable sensors based human behavioral pattern recognition using statistical features and reweighted genetic algorithm,’’ Multimedia Tools Appl., vol. 79, nos. 9–10, pp. 6061–6083, Dec. 2019.
[9] A. Jalal and M. Mahmood, ‘‘Students’ behavior mining in e-learning environment using cognitive processes with information technologies,’’ Educ. Inf. Technol., vol. 24, no. 5, pp. 2797–2821, Sep. 2019.
[10] A. Jalal, N. Sarif, J. T. Kim, and T.-S. Kim, ‘‘Human activity recognition via recognized body parts of human depth silhouettes for residents monitoring services at smart home,’’ Indoor Built Environ., vol. 22, no. 1, pp. 271–279, Feb. 2013.
[11] M. G. Morshed, T. Sultana, A. Alam, and Y.-K. Lee, ‘‘Human action recognition: A taxonomy-based survey, updates, and opportunities,’’ Sensors, vol. 23, no. 4, p. 2182, Feb. 2023, doi: 10.3390/s23042182.
[12] M. Pervaiz and A. Jalal, ‘‘Artificial neural network for human object interaction system over aerial images,’’ in Proc. 4th Int. Conf. Advancements Comput. Sci. (ICACS), Feb. 2023, pp. 1–6.
[13] B. Ren, M. Liu, R. Ding, and H. Liu, ‘‘A survey on 3D skeleton-based action recognition using learning method,’’ Cyborg Bionic Syst., vol. 5, Jan. 2024, Art. no. 0100, doi: 10.34133/cbsystems.0100.
[14] Y. Jia, G. Chen, and H. Chi, ‘‘Retinal fundus image super-resolution based on generative adversarial network guided with vascular structure prior,’’ Sci. Rep., vol. 14, no. 1, p. 22786, Oct. 2024, doi: 10.1038/s41598-024- 74186-x.