International Journal of Engineering Research and Science & Technology

SMART HUMAN ACTION MONITORING USING RGB AND MOTION SIGNALS

RAVULA ANITHA[1], KOPPELA NIRISHA REDDY[2], LATTUPALLY SWECHA REDDY[3], PAILLA THULASI REDDY[4]

[1]Assistant Professor, Dept. of CSE, Guru Nanak Institutions Technical Campus**,** Ibrahimpatnam, Telangana, India

[2, 3, 4] B.Tech, Dept. of CSE, Guru Nanak Institutions Technical Campus**,** Ibrahimpatnam, Telangana, India

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.

Index Terms – Human action recognition, RGB images, deep learning, computer vision, activity recognition, real-time monitoring, image processing, pose estimation, smart surveillance, healthcare monitoring, indoor activity recognition, intelligent monitoring systems.

I. 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 monitor systems.

II. RELATED WORK

A. A Hybrid Deep Learning-Based Approach for Human Activity Recognition Using Wearable Sensors

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.

B. Video-based human activity recognition: theory, methods and applications

Video-based human activity recognition (HAR) is an important task in many fields, such as healthcare monitoring, video surveillance, and sports analysis. This review paper aims to give an in-depth look at the current state of the art in HAR from 2018 to 2024. This will include a discussion of the different methods and models used for extracting, representing, and classifying human actions from video, as well as the challenges and limits of this field. The paper will also discuss recent improvements and plans for making HAR systems more accurate and useful. Even though there has been a lot of progress, a few knowledge gaps still need to be filled to make recognition more accurate and efficient. The purpose of this review paper is to offer scholars and professionals an overview of the theory, methods, and applications of HAR in videos. Through a critical analysis of the extant literature, this paper seeks to identify

prospective avenues for future research and contribute towards advancing HAR systems that are more precise and efficient. By showing the different ways that HAR can be used, the paper shows how important this field is in many different areas.

C. A Real-Time Human Action Recognition Model for Assisted Living

Ensuring the safety and well-being of elderly and vulnerable populations in assisted living environments is a critical concern. Computer vision presents an innovative and powerful approach to predicting health risks through video monitoring, employing human action recognition (HAR) technology. However, real-time prediction of human actions with high performance and efficiency is a challenge. This research proposes a real-time human action recognition model that combines a deep learning model and a live video prediction and alert system, in order to predict falls, staggering and chest pain for residents in assisted living. Six thousand RGB video samples from the NTU RGB+D 60 dataset were selected to create a dataset with four classes: Falling, Staggering, Chest Pain, and Normal, with the Normal class comprising 40 daily activities. Transfer learning technique was applied to train four state-of-the-art HAR models on a GPU server, namely, UniFormerV2, TimeSformer, I3D, and SlowFast. Results of the four models are presented in this paper based on class-wise and macro performance metrics, inference efficiency, model complexity and computational costs. TimeSformer is proposed for developing the real-time human action recognition model, leveraging its leading macro F1 score (95.33%), recall (95.49%), and precision (95.19%) along with significantly higher inference throughput compared to the others. This research provides insights to enhance safety and health of the elderly and people with chronic illnesses in assisted living environments, fostering sustainable care, smarter communities and industry innovation.

D. Video-based Human Action Recognition using Deep Learning: A Review

Human action recognition is an important application domain in computer vision. Its primary aim is to accurately describe human actions and their interactions from a previously unseen data sequence acquired by sensors. The ability to recognize, understand, and predict complex human actions enables the construction of many important applications such as intelligent surveillance systems, human-computer interfaces, health care, security, and military applications. In

recent years, deep learning has been given particular attention by the computer vision community. This paper presents an overview of the current state-of-the-art in action recognition using video analysis with deep learning techniques. We present the most important deep learning models for recognizing human actions, and analyze them to provide the current progress of deep learning algorithms applied to solve human action recognition problems in realistic videos highlighting their advantages and disadvantages. Based on the quantitative analysis using recognition accuracies reported in the literature, our study identifies state-of-the-art deep architectures in action recognition and then provides current trends and open problems for future works in this field.

III. PROPOSED METHODOLOGY

The proposed system introduces YOLOv8n (You Only Look Once, version 8 – nano variant) as the core model for real-time human action recognition. Unlike traditional machine learning approaches such as SVM and Random Forest, which rely on handcrafted features, YOLOv8n leverages deep learning to automatically learn spatial and temporal features directly from RGB images. Its lightweight architecture makes it highly efficient, enabling real-time detection and classification of human activities while maintaining strong accuracy levels.

YOLOv8n is particularly suited for this task due to its end-to-end object detection and classification capability, which eliminates the need for separate feature extraction and classification stages.

The model can process image frames rapidly, detecting body postures and movements while mapping them to predefined activity categories such as sitting, walking, running, or drinking. This direct integration of detection and classification streamlines the pipeline and enhances robustness in indoor environments.

Moreover, YOLOv8n is designed to perform well under challenging conditions such as low lighting, partial occlusions, and background clutter, which often limit traditional approaches.

Its nano variant is optimized for resource-constrained environments, making it deployable on edge devices and low-power systems without sacrificing much accuracy. As a result, the proposed system ensures scalability, adaptability, and efficiency, making it a practical solution for real-time human action monitoring in healthcare, surveillance, and smart indoor environments.

International Journal of Engineering Research and Science & Technology

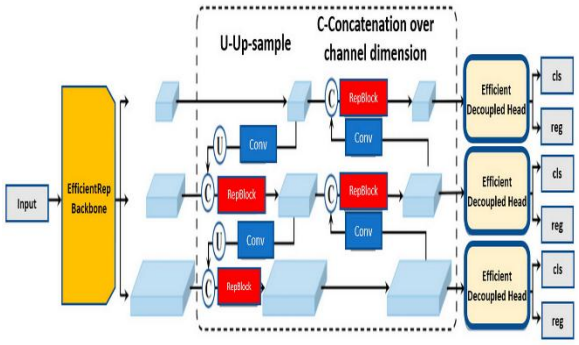


Fig. 1: System Overview

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.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

The experimental results show that the proposed smart human action monitoring system accurately recognizes daily human activities using RGB images and motion features in real time. The deep learning-based feature extraction effectively captures body posture and movement patterns, ensuring robust performance under varying lighting conditions. The model demonstrates high recognition accuracy and low processing latency, making it suitable for real-time deployment.



Fig. 2: Home Page

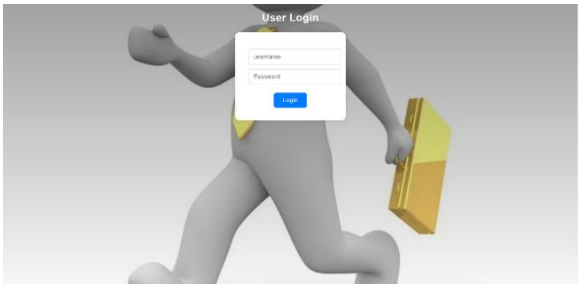


Fig. 3: Login



Fig. 4: Choose File

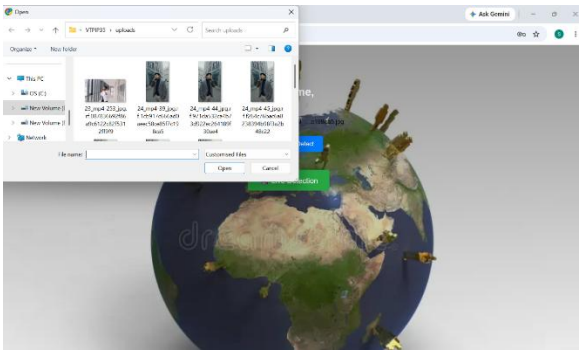


Fig. 5: Select Image

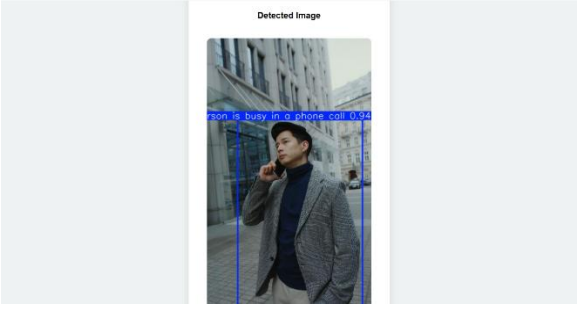


Fig. 6: Result

V. 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.

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.

AUTHORS Profile



Mrs. RAVULA ANITHA has received her B.Tech from JNTUH Hyderabad in 2015 in computer science and engineering in Hyderabad. And M.Tech from JNTUH University, Hyderabad in 2019, in Computer science and Engineering. She has been dedicated to teaching for the last 5.5 years. At present she is working as Assistant professor in CSE department at Gurunanak institution technical campus, Ibrahimpatnam, Telangana, India.



Ms. KOPPELA NIRISHA REDDY pursuing B. Tech degree in Computer Science and Engineering at Guru Nanak Institutions Technical Campus affiliated to JNTUH in 2022-2026.



Ms. LATTUPALLY SWECHA REDDY pursuing B. Tech degree in Computer Science and Engineering at Guru Nanak Institutions Technical Campus affiliated to JNTUH in 2022-2026.



Ms. PAILLA THULASI REDDY pursuing B. Tech degree in Computer Science and Engineering at Guru Nanak Institutions Technical Campus affiliated to JNTUH in 2022-2026.