Real-Time Human Action Recognition and Alert System

*Aryan Kumar[1], Dhruv Swami[2] and R Kavitha[3]
[1, 2, 3]SRM Institute of Science and Technology, Kattankulathur, Chennai, 603203, Tamil Nadu, India
[1]ak8598@srmist.edu.in
[2]ds3122@srmist.edu.in
[3]kavithar14@srmist.edu.in

Abstract. Human Action Recognition (HAR) is a major field of study in both smart surveillance systems and computer vision. Human activities such as falls or fights are of great concern due to the need for immediate contact to avert injuries and preserve public safety. This research proposes a real-time HAR and emergency alerting system using IoT hardware and Machine Learning technology. The proposed HAR system consists of a camera and MediaPipe Pose for recognizing human body landmarks in real time. The detection method is hybrid and uses rule-based logic to recognize basic activities (e.g., standing, sitting, walking, reading, clapping, and falling) and employs a Machine Learning model to recognize fighting activity only. A Random Forest Classifier is used to model pose landmark features across a variety of actions, providing multidimensional information about the detected activity. When a fall or fight is detected, an alert signal is sent from the detection system (camera/media pipe) to the Arduino Uno via a Serial Communication interface. The Arduino Uno activates the buzzer, displays warning messages on an LCD screen, and sends out an emergency SMS via a GSM module. Based on the experimental results, the proposed system achieves 99.7% accuracy and operates effectively in real time. The accuracy was calculated by testing on a collected dataset consisting of 35,000 pose landmark samples, split into an 80:20 train-test ratio. Finally, the proposed HAR and notification system is a practical and computationally efficient solution for safety-based surveillance applications.

Keywords: Human Action Recognition, MediaPipe, Random Forest Classifier, Arduino Uno, GSM Module, Smart Surveillance

1 Introduction

Human Action Recognition (HAR) aims to provide an accurate means of classifying & recognizing citizens' activities through sensors and/or video feeds. The technology developed as a result of this research can also help identify falls & other dangerous behaviors commonly encountered in healthcare monitoring, Aged Care, Smart Home & Security applications. Such a system can help with the early detection of harmful activities, such as falls or violence, thus enabling a quick response and reducing injury risk.

In traditional surveillance systems, human operators had to manually monitor video feeds to detect dangerous or harmful behavior, which is inefficient for long-duration monitoring. In addition, although wearable sensor-based video cameras may be used to

B. Singh et al. (eds.), *Proceedings of the International Conference on Advances in Computing Technology and Artificial Intelligence (COMPUTATIA 2026)*, Atlantis Highlights in Intelligent Systems 18,
https://doi.org/10.2991/978-94-6239-713-2_47

monitor individuals on an ongoing basis, they create additional inconvenience for the wearer. As pose detection technologies evolve, human movement detection based on joint coordinate data rather than image data will reduce computational time and increase the reliability of human activity detection. In this study, an HAR system will be introduced, comprising Pose Detection, Machine Learning Techniques, and an IoT Alert Mechanism to enable real-time detection of critical events.

2 Previous Works

Previously, HAR systems primarily relied on wearable devices (accelerometers/gyroscopes). These systems have good accuracy but require users to keep sensors on at all times by the authors [1]. Vision-based approaches using deep learning models (CNNs/LSTMs) have also been researched extensively [4][5]. These approaches provide good performance; however, they often require large datasets and very powerful hardware, which limits their usefulness in real-time and low-cost applications. Recent studies have demonstrated that pose-based activity recognition using skeletal landmarks is a highly efficient way to recognize and perform activities in real time. Some of these studies have focused on detecting falls by recognizing changes in pose. Others have focused on detecting assaults by recognizing the patterns associated with these motions [2][3]. Some existing human action recognition systems provide false alerts when distinguishing between similar human motions. This study proposes an alternative means for addressing these problems.

Table 1. A Comparative Study of Previous Models

Technologies	Description	References
Support Vector Machine (SVM)	A supervised machine learning algorithm used for classifying static human poses or simple activities based on fixed features. Performs well on small datasets but struggles with sequential video data.	[1]
Random Forest (RF)	An ensemble learning technique combining multiple decision trees to classify human actions. Provides good accuracy but is limited to time-series or motion-based recognition.	[2]
Long Short-Term Memory (LSTM)	Recurrent neural network architecture designed to capture long-term temporal dependencies across video frame sequences.	[3]
Convolutional Neural Network (CNN)	A deep learning model is used to extract spatial features from individual video frames.	[4]

CNN–LSTM Hybrid Model	Combines CNN for feature extraction and LSTM for temporal sequence learning. Well-suited for video-based human action recognition, providing higher accuracy for complex activities.	[5]

3 Model Proposed

The model proposed has three main modules:

3.1 Pose Extraction Module

The Pose Extraction Module consists of live video streaming from a webcam and capture of pose landmarks using the MediaPipe Pose library, which identifies and stores 33 body landmarks, including the coordinates of the shoulder joints, both hip joints, both knee joints, both wrist joints, and both ankle joints, that are provided as input for the action recognition process.

3.2 Hybrid Activity Recognition Module

The hybrid activity detection system incorporates a mixture of rule-based detection and machine learning detection:
• Rule-Based Detection: Using joint angle measurements and/or distance measurements and/or motion history measurements to detect activities such as standing, sitting, walking, reading, clapping, and falling.
• Machine Learning Detection: Fighting detection is accomplished by using a Random Forest Classifier that has been trained on the pose landmark data. It was trained using pose landmark coordinates extracted from MediaPipe Pose for seven activities, with around 5000 samples per class. To reduce the number of false-positive detections, machine learning is typically applied only to complex activities.

3.3 Web App

The web application is built with the help of a Flask server, and provides:

- User authentication
- Live video streaming
- Real-time display of detected activities

3.4 Hardware Alert System

When the fall event or fighting event occurs, the following actions will take place:

1. An alert will be sent to the Arduino Uno for detection via serial communication.
2. The Arduino will activate the following devices:
   - Buzzer to provide an audible alert
   - LCD to display a visual alert

- GSM module to send an emergency SMS message to the admin.

We have used a step-down transformer for a regulated power supply and a voltage regulator to provide stable power to the Arduino, GSM module, LCD, and buzzer.

4 Evaluating Performance

Seven custom datasets (standing, sitting, walking, reading, clapping, falling, having a fight) with 5000 samples each were developed using MediaPipe Pose landmarks. The dataset was created by recording video sequences of volunteers performing each activity and extracting pose landmarks using MediaPipe Pose. A Random Forest classifier with 300 trees and a maximum depth of 25 was trained on the seven datasets with an 80:20 train-test split. The model was evaluated using an 80:20 train-test split, with no overlap between training and testing samples. Results:

- Accuracy: 99.7%
- Low false alert
- Stable real-time performance
- Rapid emergency alerts

Using a hybrid approach that combines rule-based logic with machine learning only where required increases the application's reliability.

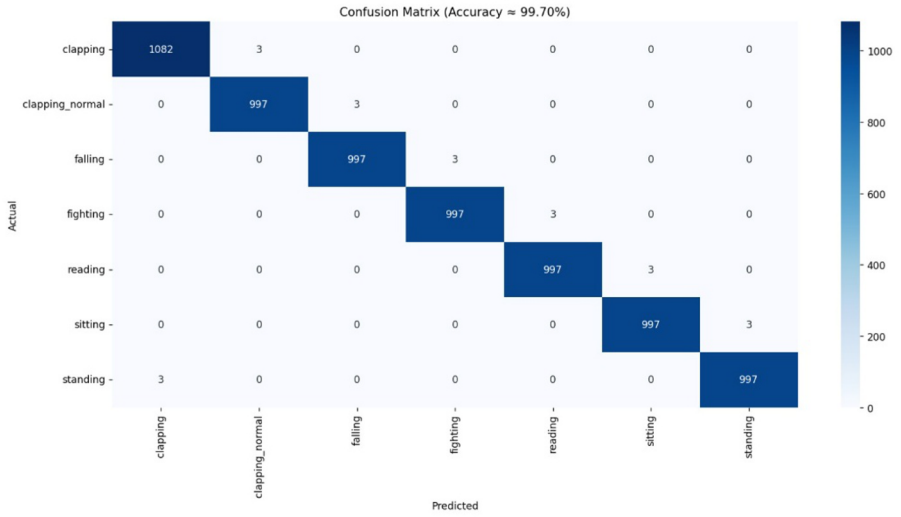

Fig. 1. Confusion Matrix

5 Results

The proposed hybrid Human Action Recognition (HAR) system was tested in real time using a live camera feed, in conjunction with MediaPipe Pose and a Random Forest classifier. The system accurately detected several human activities, including standing, sitting, walking, clapping, reading, and falling (left/right), as well as fighting. The rule-based module consistently detected posture-based actions, such as sitting, standing, walking, clapping, and reading, with minimal latency during testing. The system's fall detection rules identified when a person fell and the direction of the fall by analyzing shoulder tilt. The Random Forest machine learning model was trained on pose landmark data and achieved good accuracy during offline testing. It was used to detect complex actions, such as fighting, in real-time data. The hybrid approach helped reduce incorrect detections by applying machine learning only to complex activities, such as fighting, while simple actions were identified using predefined rules. When the system detects a dangerous action, such as a fall or a fight, it sends a signal to the Arduino, which then activates the buzzer. It also activates alerts such as an LCD warning display and an SMS Notification. The system is designed to activate the alert only once to avoid repeated notifications. Overall, the system worked smoothly in real time, maintained a stable frame processing speed, and showed potential for continuous indoor monitoring and safety applications.

6 Conclusion and Future Scope

This project makes a system that can see what people are doing and send alerts when something dangerous happens. It uses a hybrid method, which means it uses both rule-based analysis and machine learning. MediaPipe Pose is a tool that helps find important body points (landmarks) in camera footage of the human skeleton. The system uses set rules for simple tasks, but for more complicated tasks like fighting, it uses a Random Forest machine learning model. This system is more efficient than traditional HAR systems that only use deep learning because it cuts down on heavy computation. Using rules for simple tasks also makes the system easier to understand and use. There is also a web-based monitoring interface and hardware alerts built into the system. The system can automatically send alerts like a buzzer, display messages, or send notifications when it detects a dangerous event like a fall or a fight. The successful real-time detection of falls and violent activities suggests that the system could be useful in indoor environments, such as old-age homes, hospitals, hostels, schools, and smart surveillance setups. The proposed system can be extended and enhanced in several ways:

- Using multiple cameras can reduce any unwanted obstruction and can cover larger areas more accurately.
- Models like LSTM or Transformers can help detect long and complex activities.
- Integrating cloud and IoT can allow for remote monitoring and instant mobile alerts.

- Expanding and personalizing the dataset can improve performance for different people and environments.
- Deploying on devices like Raspberry Pi or Jetson can make the system efficient and easy to carry.

References

1. Z. Yu and W. Q. Yan, "Human Action Recognition Using Deep Learning Methods," 2020 35th International Conference on Image and Vision Computing New Zealand (IVCNZ), Wellington, New Zealand (2020)
2. Y. -H. Wu, W. -J. Tsai and H. -T. Chen, "Temporal Action Detection Based on Hierarchical Object Detection Networks," 2019 Twelfth International Conference on Ubi-Media Computing (Ubi-Media), Bali, Indonesia (2019)
3. L. Wang, S. Zhu, W. Qi, and J. Yang, "Sparse Method Towards Temporal Action Detection," 2022 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS), Penang, Malaysia (2022)
4. C. -K. Lu, M. -W. Mak, R. Li, Z. Chi and H. Fu, "Action Progression Networks for Temporal Action Detection in Videos," in IEEE Access, vol. 12, pp. 126829-126844, (2024)
5. T. Su, H. Wang, and L. Wang, "Multi-Level Content-Aware Boundary Detection for Temporal Action Proposal Generation," in IEEE Transactions on Image Processing, vol. 32, pp. 6090-6101 (2023)
6. S. Su and Y. Zhang, "Online Hierarchical Linking of Action Tubes for Spatio-Temporal Action Detection Based on Multiple Clues," in IEEE Access, vol. 12, pp. 54661-54672 (2024)
7. K. Xia, L. Wang, Y. Shen, S. Zhou, G. Hua, and W. Tang, "Exploring Action Centers for Temporal Action Localization," in IEEE Transactions on Multimedia, vol. 25, pp. 9425-9436 (2023)
8. F. S. Khan, J. Xu, J. van de Weijer, A. D. Bagdanov, R. M. Anwer, and A. M. Lopez, "Recognizing Actions Through Action-Specific Person Detection," in IEEE Transactions on Image Processing, vol. 24, no. 11, pp. 4422-4432 (2015)
9. J. Huang, N. Li, T. Li, S. Liu, and G. Li, "Spatial–Temporal Context-Aware Online Action Detection and Prediction," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 30, no. 8, pp. 2650-2662 (2020)
10. Y. Li, Z. Wang, Z. Li, and L. Wang, "Sparse Action Tube Detection," in IEEE Transactions on Image Processing, vol. 33, pp. 1740-1752 (2024)
11. Y. Wu et al., "Temporal Action Detection Based on Action Temporal Semantic Continuity," in IEEE Access, vol. 6, pp. 31677-31684 2018)
12. C. Lin, T. Ma, F. Wu, J. Qian, F. Liao, and J. Huang, "Application of Temporal Action Detection Technology in Abnormal Event Detection of Surveillance Video," in IEEE Access, vol. 13, pp. 26958- 26972 (2025)
13. R. Benitez and Z. Nenadic, "Robust Unsupervised Detection of Action Potentials With Probabilistic Models," in IEEE Transactions on Biomedical Engineering, vol. 55, no. 4, pp. 1344-1354 (2008)
14. S. Lee et al., "Learning to Discriminate Information for Online Action Detection: Analysis and Application," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 45, no. 5, pp. 5918-5934 (2023)