IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940



International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 10, October 2026

DOI: 10.17148/IJARCCE.2026.1510135

AI DRIVEN SUSPICIOUS ACTIVITY RECOGNITION USING DEEP LEARNING

Neelakantappa T T[1], Meghana G S[2], Nandini G[3], Gousiya Fathima A[4], Anusha R H[5]

Assistant Professor, Department of Computer Science & Engineering, S.J.M Institute of Technology, Chitradurga, Karnataka, India[1]

B.E. Student, Department of Computer Science & Engineering, S.J.M Institute of Technology, Chitradurga, Karnataka, India[2-5]

Abstract: The growing deployment of closed-circuit television (CCTV) cameras produces large volumes of video that are difficult to monitor continuously through human observation alone. Manual surveillance is vulnerable to fatigue, delayed response, and missed events, while basic motion detection cannot reliably distinguish ordinary movement from potentially suspicious activity. This paper presents a deep-learning-based framework for suspicious human activity recognition from surveillance videos using a hybrid Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) architecture. Video frames are extracted and preprocessed using OpenCV. A pretrained CNN, such as MobileNet or InceptionV3, is intended to extract spatial features from individual frames, while an LSTM models temporal relationships across frame sequences. A classification layer predicts activity categories and can provide confidence scores to support alert generation. The framework is designed using Python and TensorFlow/Keras, with potential evaluation on public video datasets such as UCF-Crime or HMDB51. Since measured implementation results were not provided in the project report, this paper does not claim a specific accuracy, frame rate, or real-world performance. Experimental validation is required to establish effectiveness, false-alarm rates, and deployment suitability.

Keywords: Suspicious activity recognition, intelligent surveillance, deep learning, CNN-LSTM, computer vision, human activity recognition, video analytics

I. INTRODUCTION

CCTV surveillance is widely used in public places, educational institutions, hospitals, transport facilities, commercial buildings, and residential areas. Although cameras continuously record activity, the task of watching many video feeds remains demanding. Operators may experience fatigue or overlook important events, and reviewing stored footage after an incident can delay response.

Computer vision and deep learning provide a way to analyse video content automatically. CNNs learn spatial representations from image frames, while recurrent models such as LSTM networks capture changes and dependencies across time. Combining these approaches allows a system to consider both the appearance of a frame and the sequence of movements that precedes or follows it.

This paper describes a proposed CNN-LSTM framework for recognising suspicious human activities in surveillance videos. The intended output includes a predicted activity label and confidence score, with alerts for activities configured as suspicious. The system is intended as a decision-support tool; predictions should be reviewed by a human operator and should not be treated as proof of wrongdoing.

II. RELATED WORK

Research on video anomaly detection has explored weakly supervised learning, deep feature extraction, and temporal modelling. Sultani, Chen, and Shah introduced a large-scale real-world surveillance anomaly dataset and a multiple-instance-learning approach for anomaly detection [1]. Their work highlighted the difficulty of identifying abnormal events in untrimmed surveillance footage when detailed frame-level labels are unavailable.CNN-based methods are commonly used to extract visual features, while LSTM and other recurrent architectures model temporal sequences. Studies of human activity recognition also identify dataset imbalance, context variation, occlusion, and computational cost as important challenges [2]. Work focused on examination-hall surveillance has explored person detection and activity classification in a constrained environment [3], although systems developed for one environment may not

 This work is licensed under a Creative Commons Attribution 4.0 International License

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 10, October 2026

DOI: 10.17148/IJARCCE.2026.1510135

generalise to public surveillance settings.The proposed framework combines frame-level CNN feature extraction with LSTM temporal modelling. Its practical contribution is a modular workflow that connects video input, frame preparation, feature extraction, sequence classification, and alert output. Its performance must be measured against suitable baselines before claims of superiority can be made.

III. PROBLEM STATEMENT AND OBJECTIVES

Conventional surveillance relies heavily on human operators and basic motion detection. These approaches may miss events, produce unnecessary alarms, and fail to interpret the temporal sequence of human actions. The problem addressed is how to design an automated video-analysis pipeline that learns spatial and temporal patterns and classifies selected activities as normal or suspicious.

OBJECTIVES:

- Extract spatial features from video frames using a pretrained CNN such as MobileNet or InceptionV3.
- Model temporal relationships between frame features using an LSTM network.
- Classify activities into predefined classes and provide a prediction confidence score.
- Provide an alert/output mechanism and evaluate the system using suitable classification metrics.

IV. PROPOSED SYSTEM ARCHITECTURE

V.

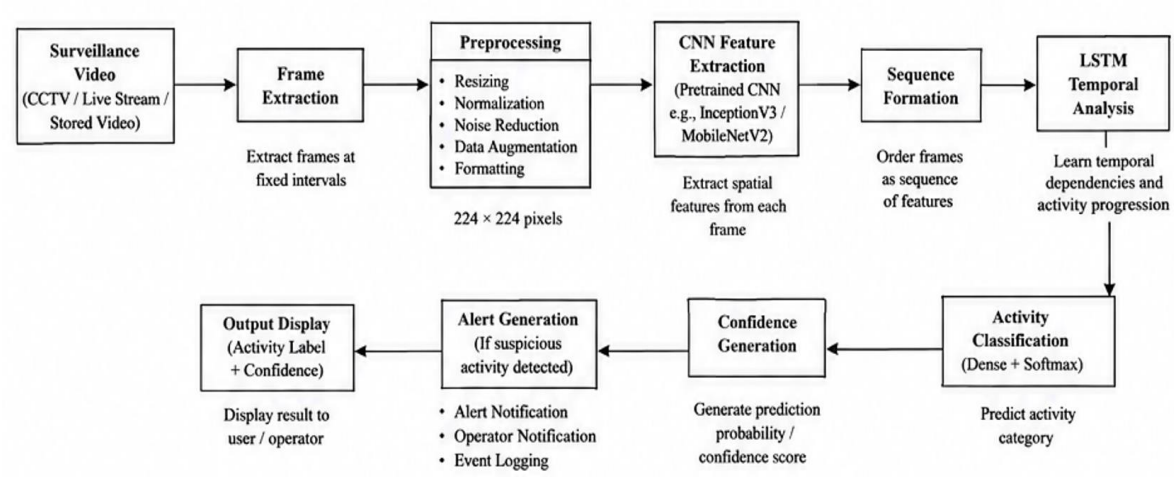


Fig. 2. Proposed System Architecture for Suspicious Human Activity Recognition Using Deep Learning

A. Video Input and Preprocessing

The input layer receives recorded video files or a supported camera stream. Frames are extracted in chronological order, resized to the input dimensions required by the selected CNN (for example, 224 × 224 pixels), and normalized. A fixed number of consecutive frames is grouped into a sequence so that temporal information can be modelled.

B. CNN Feature Extraction

A pretrained CNN such as MobileNet or InceptionV3 processes each frame and produces a feature vector representing visual information. Transfer learning can reduce the amount of training required. The final implementation should document the chosen backbone, preprocessing method, input dimensions, and whether the CNN is frozen or fine-tuned.

C. LSTM Temporal Modelling and Classification

The ordered feature vectors are passed to an LSTM network, which learns dependencies across consecutive frames. A dense classification layer with an appropriate output activation predicts one of the defined activity classes. Candidate classes in the project include normal activity, walking, running, fighting, theft-related activity, falling, and other suspicious behaviour. The classes must be clearly defined and supported by the training labels.

D. Output and Alerts

The output module displays the predicted class and confidence score. When the prediction meets a predefined alert rule, the system can record the event and notify an operator. Confidence scores should be calibrated and alert

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 10, October 2026

DOI: 10.17148/IJARCCE.2026.1510135

thresholds selected through validation. A model prediction alone should not be interpreted as confirmation that a crime has occurred.

VI. METHODOLOGY

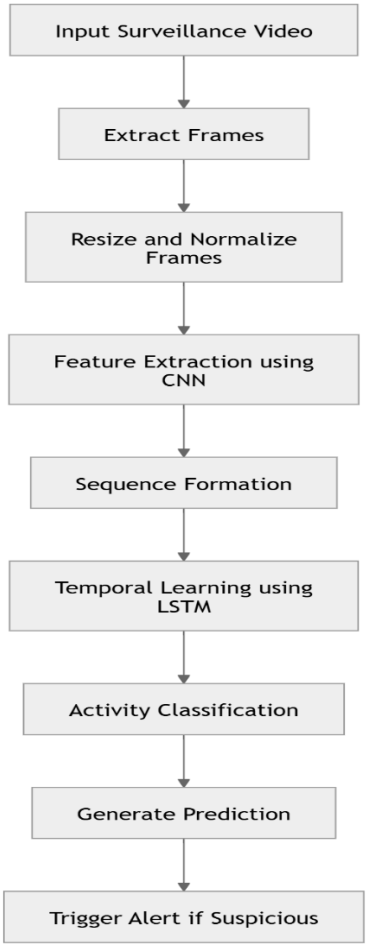


Fig.1. Proposed System Architecture for Suspicious Human Activity Recognition Using Deep Learning

A. Dataset Preparation

The project report identifies public video datasets such as UCF-Crime and HMDB51 as possible sources, with the option of collecting a suitable custom dataset. The final study must state the exact dataset version, selected classes, number of videos, class distribution, and annotation method. Training, validation, and test sets should be separated by source video to reduce data leakage.

B. Frame Sampling and Feature Generation

Video frames are sampled at a fixed rate and grouped into sequences of a defined length. Frames are resized and normalized to match the CNN requirements. CNN feature vectors are generated for each sampled frame and arranged in temporal order to form the input sequence for the LSTM.

C. Model Training

The CNN-LSTM model is trained on labelled sequences using a suitable classification loss and optimizer. Hyperparameters such as sequence length, learning rate, batch size, and number of epochs should be selected using the validation set. Data augmentation may be applied carefully to improve robustness, while ensuring that transformations do not change the activity label.

D. Evaluation

The model should be evaluated on a held-out test set using accuracy, precision, recall, F1-score, a confusion matrix, and false-positive rate. For imbalanced or anomaly-detection datasets, class-wise results and ROC-AUC or PR-AUC



International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 10, October 2026
DOI: 10.17148/IJARCCE.2026.1510135

may also be appropriate. Runtime, frames per second, hardware configuration, and alert latency should be recorded if real-time operation is claimed.

VII. RESULTS AND DISCUSSION

The available project report describes the proposed architecture and intended workflow but does not provide verified experimental measurements for the implementation. Therefore, numerical accuracy, precision, recall, F1-score, inference speed, and comparisons with baseline models cannot be reported as established results in this paper.

The expected behaviour is that the system processes frames, generates a sequence-level activity prediction, and issues an alert when a selected suspicious class meets the configured threshold. This expected workflow must be confirmed through experiments. Evaluation should include normal activities that resemble suspicious motion, changes in lighting, occlusion, crowded scenes, camera-angle variation, and videos from environments not represented in training. False positives can interrupt operators, while false negatives may cause important events to be missed.

A complete experimental version should include a results table containing the dataset split, class-wise metrics, overall metrics, inference speed, and comparison with at least one baseline. All reported values should be calculated from the implemented model and documented test protocol.

VIII. CONCLUSION

This paper presents a proposed CNN-LSTM framework for suspicious human activity recognition from surveillance videos. The pipeline combines OpenCV-based frame extraction and preprocessing, CNN-based spatial feature extraction, LSTM temporal modelling, and classification with confidence-based alerting. The design addresses limitations of manual surveillance by providing a structured route toward automated video analysis. However, the available project material does not establish measured model performance. Future work should implement the complete pipeline, conduct controlled experiments on documented datasets, compare alternative architectures, quantify false alarms and missed detections, and assess real-time performance before deployment.

ACKNOWLEDGMENT

The authors acknowledge the management of S.J.M. Vidyapeetha, the Principal, the Department of Computer Science and Engineering, and the faculty of S.J.M. Institute of Technology, Chitradurga, for their academic guidance and support. Please confirm the project guide and author names before submission.

REFERENCES

[1]. W. Sultani, C. Chen, and M. Shah, “Real-World Anomaly Detection in Surveillance Videos,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018.
[2]. J. Wang, Y. Chen, S. Hao, X. Peng, and L. Hu, “Deep Learning for Sensor-Based Activity Recognition: A Survey,” Pattern Recognition Letters, vol. 119, pp. 3–11, 2019. Verify bibliographic details against the original publication.
[3]. S. Auti, P. Darade, A. Munje, N. Patil, and A. Khushwaha, “Deep Learning Approach for Suspicious Activity Detection from Surveillance Video in Examination Hall,” 2025. Verify venue, volume, pages, and DOI from the original publication.
[4]. OpenCV, “OpenCV Documentation.” https://docs.opencv.org/ .
[5]. TensorFlow, “TensorFlow Documentation.” https://www.tensorflow.org/ .
[6]. “Digital Image Watermarking Using Deep Learning: A Survey,” 2024. Author a Details should be verified from the original source.
[7]. Python Software Foundation,“hashlib—Secure has hes and message digests,”Python documentation.
[8]. OpenCV, “OpenCV Documentation,” official documentation.

 This work is licensed under a Creative Commons Attribution 4.0 International License