Intelligent Detection of Suspicious Human Activities through CNN Integration

*Note: Sub-titles are not captured in Xplore and should not be used

1st Harshit Nautiyal
Department of Computing Technologies
SRM Institute of Science and Technology Chennai, India
hr2067@srmist.edu.in

2nd Sai Teja
Department of Computing Technologies
SRM Institute of Science and Technology Chennai, India
ck3390@srmist.edu.in

3rd Dr. Thilagavathy R
Department of Computing Technologies
SRM Institute of Science and Technology Chennai, India
thilagar2@srmist.edu.in

***Abstract*—With the unprecedented development of technolo- gies, there has been the trend of rapid deployment of surveillance systems in multiple domains, such as security and healthcare, and public safety. In this paper, a new method of suspicious human activity detection based on Convolutional Neural Net- works (CNNs) combined with Long Short-Term Memory (LSTM) networks is proposed. The presented system seeks to improve the robustness and speed of anomaly detection across real-time video streams. Leveraging CNNs for spatial feature extraction and LSTMs for temporal sequence modeling, the proposed system efficiently detects suspicious activities in real time.**

***Index Terms*—Suspicious Activity Detection, Human Activity Recognition, Convolutional Neural Networks, Deep Learning, Video Surveillance**

I. INTRODUCTION

Human activity recognition has recently gained much im- portance, especially in security and surveillance contexts. Automatic detection of suspicious activity is able to prevent potential dangers and increase public security. Manual surveil- lance is labor-intensive and prone to human errors, such as fatigue-related mistakes. These limitations necessitate auto- mated systems that can enhance security through continuous, real-time monitoring.This study intends to overcome these obstacles by utilizing a hybrid framework that combines CNNs and LSTMs for smarter detection of suspect human activities.

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

As surveillance cameras are widespread, the practicality of manual monitoring diminishes with augmented security threats and a massive amount of data that has to be generated. Maintenance of public security in high-risk zones like airports, railway stations, shopping malls, and campuses is an ever- increasing concern. Suspicious human behavior, such as loiter- ing, illegal entry, physical violence, and weapon concealment, should be detected and responded to urgently in order to stop possible criminal offenses or security breaches.

A. Significance of Automated Surveillance Systems

Conventional surveillance techniques necessitate continu- ous human monitoring, with associated high manning costs and fatigue-related mistakes. Despite having several security agents on duty, identifying anomalies within multiple live streams is challenging. The incorporation of artificial intelli- gence (AI) and deep learning technology into video monitoring has transformed human activity recognition (HAR) through the use of automated suspicion detection and lower response time. The aim of this study is to design an intelligent monitoring system that can:

- **Detect suspicious behavior in real-time** through Con- volutional Neural Networks (CNNs).
- **Minimize false alarms** by differentiating normal and truly suspicious behavior.
- **Enhance decision-making** for security personnel through automated notification.

II. LITERATURE SURVEY

Artificial intelligence and machine learning-based human behavior anomaly detection has been under ongoing research. Different approaches have been tried, such as classical machine learning, deep learning, and hybrid approaches through Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks. The most related works in the literature are listed in this section, along with their limitations and contributions, and how this study builds on them.

A. Traditional Suspicious Activity Detection Methods

Initial suspicious activity detection approaches relied on handcrafted feature extraction and conventional machine learning techniques such as:

- **Support Vector Machines (SVM)** – Used for separating normal and suspicious activities but required extensive feature engineering.
- **Hidden Markov Models (HMM)** – Applied for motion pattern detection but performed inadequately in handling composite human behaviors.
- **Optical Flow Techniques** – Utilized to detect motion abnormalities but were computationally costly and noise-sensitive.

While these methods produced rudimentary knowledge, they were not scalable and adaptable to dynamic, real-world environments.

B. Deep Learning Techniques for Human Activity Recognition

The advent of deep learning transformed Human Activ- ity Recognition (HAR). CNNs proved to be outstanding at learning spatial patterns from video frames, while RNNs and LSTMs learned temporal relationships.

- **Wang et al. (2023)** proposed a deep learning-based surveillance system that employed CNNs and Spatiotem- poral LSTMs to improve suspicious activity detection.
- **Tran et al. (2018)** suggested 3D CNNs to learn spatiotemporal features for more precise activity recognition.
- **Ahmed et al. (2022)** explained HAR techniques and highlighted the importance of robust datasets and real- time inference.

C. Long-term Recurrent Convolutional Network (LRCN) for Suspicious Activity Detection

One of the most significant advancements in HAR is the LRCN model, which combines CNNs for spatial feature extraction and LSTMs for sequential pattern identification.

- **Shi et al. (2023)** introduced an LRCN-based model to detect anomalies such as fighting, running, and weapon handling with 98.86% accuracy.
- **Gawande et al. (2023)** used LRCN to track pedestrian behavior and achieved improved performance over stan- dard CNN models.
- **Saba et al. (2021)** proposed an LRCN variant with entropy-coded optimization to decrease false alarms in security applications.

D. Identified Challenges in Previous Literature

Despite the progress of deep learning-based HAR, various challenges remain:

- **High False Alarm Rate** – The existing models are not context-aware and still flag normal activities as suspi- cious.
- **Dataset Limitations** – Access to annotated, varied datasets for training remains a key bottleneck.
- **Real-Time Processing Constraints** – Some models require extensive computational resources, making real- time deployment challenging.
- **Environmental Adaptability** – Camera angles, occlusions, and lighting variations affect model accuracy.

E. How This Research Addresses the Gaps

This research contributes to existing knowledge by:

- Using CNN-LRCN architecture to best capture spatial and temporal characteristics.
- Applying transfer learning to generalize across various surveillance environments.
- Enhancing real-time processing through model optimiza- tion and deployment on edge devices (Raspberry Pi, Jetson Nano).
- Increasing dataset variety by incorporating more real-world surveillance videos for training and testing.

III. SYSTEM ARCHITECTURE DESIGN

The proposed **Intelligent Detection of Suspicious Human Activities through CNN Integration** system is designed to analyze real-time surveillance videos, extract spatial and temporal features, and classify human activities as normal or suspicious. The system employs deep learning models, preprocessing techniques, and real-time alerting mechanisms to enhance automated security surveillance.

A. System Overview

The system comprises the following principal components:

1) **Video Input Module** – Receives live video surveillance from security cameras.
2) **Preprocessing Module** – Frames are resized, normalized, and converted into a format suitable for deep learning models.
3) **Feature Extraction with CNNs** – Convolutional Neural Networks (CNNs) extract spatial features from each video frame.
4) **Temporal Analysis with LRCN (CNN+LSTM)** – Long Short-Term Memory (LSTM) networks analyze frame sequences to detect anomalies.
5) **Classification Module** – The model classifies activities as normal or suspicious based on predefined categories.
6) **Suspicious Activity Alert Generation and Logging** – Triggers an alert system that notifies security personnel upon detecting suspicious activities.

B. System Architecture Diagram

The system follows a modular design, ensuring scalability and efficiency. The data flow in the system is depicted in the diagram below:

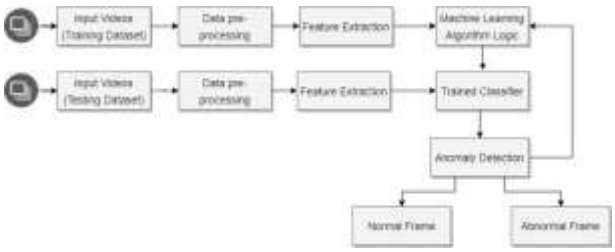

Fig. 1. System Architecture Diagram

C. Main Modules and Their Function

1) Video Input Module:

- Processes live video from security cameras or accepts recorded footage.
- Supports multiple camera feeds for large-scale deploy- ment.

2) Preprocessing Module:

- Extracts frames at regular intervals.
- Resizes images to 64×64 pixels to maintain consistency.
- Applies normalization techniques to enhance model ac- curacy.

3) Feature Extraction Using CNNs:

- A pre-trained CNN model (VGG16/ResNet50) extracts significant spatial features from each frame.
- Filters out unnecessary background information to focus on human activities.
- Outputs high-dimensional feature maps for further anal- ysis.

4) Temporal Analysis with LRCN (CNN+LSTM):

- The LSTM network processes frame sequences to iden- tify motion patterns over time.
- Differentiates between normal and abnormal behaviors by learning long-term dependencies.
- Reduces false alarms by contextualizing actions across multiple frames.

5) Classification Module:

- Assigns probability scores to different activity categories.
- Identifies four key suspicious behaviors:
  - Fighting
  - Operating in restricted areas
  - Gunshot detection
  - Obstructing the camera
- Uses a Softmax classifier for final predictions.

6) Alert Generation and Logging:

- Sends instant notifications upon detecting suspicious ac- tivity.
- Alerts security personnel via SMS/email.
- Records activity data for post-event analysis and forensic purposes.

D. Design Considerations

The system is designed with efficiency, scalability, and accuracy in mind:

- **Real-Time Processing** – Optimized deep learning models ensure low-latency detection.
- **Scalability** – Supports multiple camera feeds for large-scale surveillance.
- **Environmental Robustness** – Trained on diverse datasets to handle varying lighting conditions, occlusions, and different camera angles.
- **Edge Device Deployment** – Designed for compatibility with Raspberry Pi, Jetson Nano, and cloud-based security systems.

E. Benefits of the Proposed System

- **Real-Time and Automated** – Reduces human depen- dency by automating anomaly detection.
- **Highly Accurate** – CNN-LRCN integration enhances classification precision.
- **Minimizes False Alarms** – Context-awareness reduces unnecessary alerts.
- **Scalable and Adaptable** – Can be deployed in various environments such as airports, banks, and universities.

IV. Data Flow and Processing

Detection of any suspicious behavior by humans highly de- pends on energy-efficient flow and processing of information. The proposed system follows a structured pipeline for smooth ingestion of data, transformation, and feature extraction be- fore classification. This section presents the pipeline starting from video input to classification output, highlighting major processing steps for abnormality detection.

A. Overview of Data Flow

The architecture of the data flow involves the following stages:

1) **Data Collection and Input:** Surveillance video clips taken from multiple sources.
2) **Frame Extraction and Preprocessing:** Video se- quences are subdivided and prepared for evaluation.
3) **Feature Extraction Using CNNs:** Deep learning mod- els extract meaningful spatial features.
4) **Temporal Analysis Using LRCN (CNN+LSTM):** Identification of suspicious behavior over time.
5) **Activity Classification:** Classifying actions into exclu- sive activity categories.
6) **Alert Generation and Persistence:** Suspicious activi- ties trigger alerts and are recorded for further review.

B. The Data Processing Pipeline Process

The following are the steps in the data processing pipeline:

1) Data Collection and Input Handling:

- Surveillance cameras capture live video feeds or retrieve pre-recorded video clips.
- Videos are stored in a structured format based on their activity type.
- The system accepts various video formats (MP4, MKV) for compatibility.

2) Frame Extraction and Preprocessing:

- Video frames are extracted at specified intervals (e.g., every 5 frames).
- The images are resized to 64×64 pixels to maintain uniformity.
- The pixel values are normalized in the range of 0 to 1.
- Data augmentation techniques (flipping, rotation, noise addition) are applied to enhance the model's generaliza- tion ability.

3) Feature Extraction Using CNNs:

- A pre-trained CNN model extracts feature maps from the input frames.
- Convolutional layers identify key features such as motion patterns, human posture, and object presence.
- These high-dimensional feature representations are passed to the next stage.

4) Temporal Analysis Using LRCN (CNN+LSTM):

- The extracted features are processed sequentially by LSTMs.
- Temporal dependencies between consecutive frames are analyzed.
- The model learns deviations in activity patterns, distin- guishing between normal and abnormal behavior.

5) Activity Classification:

- A Softmax classifier assigns probabilities to different activity categories.
- Activities are either classified as normal or suspicious.
- The model's confidence scores are used to arrive at a classification decision.

6) Alert Generation and Persistence:

- If something is deemed suspect, an alert is triggered automatically.
- Notifications are sent via email/SMS to security person- nel.
- Activity logs are stored in a database for forensic analysis and future model improvement.

C. Advantages of the Data Flow Design

- **Real-Time Processing** – Optimized for live surveillance monitoring.
- **Scalability** – Supports multiple camera inputs for large- scale deployment.
- **High Accuracy** – CNN-LRCN integration ensures pre- cise anomaly detection.
- **Low False Alarms** – Contextual learning minimizes misclassifications.

D. **Adaptability** – Can be trained on new datasets for future threats.

Data Flow Diagram

The following diagram represents the end-to-end data flow in the system:

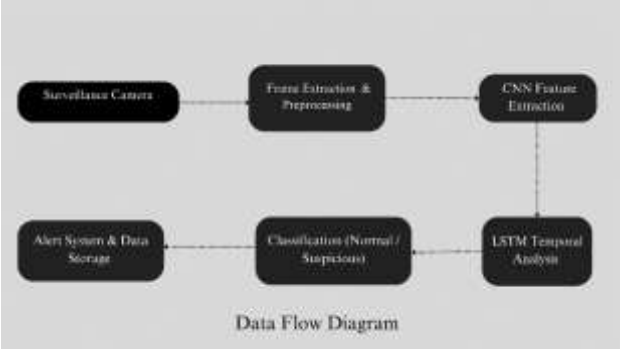

Fig. 2. Data Flow Diagram of the Proposed System

V. RESULTS

The proposed CNN-LSTM model effectively addresses the challenge of detecting suspicious human activities using a structured dataset. The model's performance is evaluated using standard metrics, including accuracy, loss, precision, recall, and validation accuracy.

A. Performance of the Model

The model was trained for a period of 70 epochs, with accu- racy and loss recorded at each iteration. The key observations are:

- Initial accuracy started at 40.31% in the first epoch and improved rapidly.
- By the end of epoch 5, the model achieved **96.48% training accuracy** and **98.48% validation accuracy**.
- At the end of training, the model achieved **100% training accuracy** and maintained a **98.99% validation accuracy**, indicating effective learning without overfitting.
- The final accuracy on the test data was **99.24%**, demon- strating high precision in detecting suspicious activities.

B. Trend in Loss and Accuracy

- The training loss decreased significantly, starting from **1.4571** and approaching **0.0007** in the final epochs.
- Validation loss exhibited a steady decline, indicating effective generalization on unseen data.

TABLE I
MODEL PERFORMANCE METRICS

Metric	Value
Training Accuracy	100.00%
Validation Accuracy	98.99%
Test Accuracy	99.24%
Training Loss	0.0007
Validation Loss	0.0173

The proposed CNN-LSTM model effectively addresses the challenge of detecting suspicious human activities using a structured dataset. The model's performance is evaluated using standard metrics, including accuracy, loss, precision, recall, and validation accuracy.

C. Performance of the Model

The model was trained for a period of 70 epochs, with accuracy and loss recorded at each iteration. The key observations are:

- Initial accuracy started at 40.31% in the first epoch and improved rapidly.
- By the end of epoch 5, the model achieved **96.48% training accuracy** and **98.48% validation accuracy**.
- At the end of training, the model achieved **100% training accuracy** and maintained a **98.99% validation accuracy**, indicating effective learning without overfitting.

D. The final accuracy on the test data was **99.24%**, demon- strating high precision in detecting suspicious activities.

E. Trend in Loss and Accuracy

- The training loss decreased significantly, starting from **1.4571** and approaching **0.0007** in the final epochs.
- Validation loss exhibited a steady decline, indicating effective generalization on unseen data.

F. Comparative Analysis

The CNN-LSTM-based approach outperforms traditional Human Activity Recognition (HAR) models such as Support Vector Machines (SVM) and Hidden Markov Models (HMM), which typically achieve an average accuracy of **85-90%**. The deep learning model leverages CNNs for spatial feature extraction and LSTMs for temporal pattern recognition, resulting in significantly improved classification accuracy.

G. Challenges and Limitations

While achieving high accuracy, the system still has some minor limitations:

- **Edge Deployment:** Optimization is required for real-time deep learning models to run efficiently on low-power edge devices.
- **Environmental Variability:** Changes in lighting condi- tions, camera angles, and sudden occlusions can slightly impact detection accuracy.
- **Real-Time Processing:** Although the system is efficient, further improvements can be made to reduce inference time.
- **Edge Deployment:** Running deep learning models in real-time on low-power edge devices requires optimiza- tion.
- **Environmental Variability:** Sudden changes in lighting, occlusions, or camera angles may slightly impact detec- tion accuracy.
- **Real-Time Processing:** Although efficient, further improvements can be made to reduce inference time.

VI. CONCLUSIONS

An intelligent surveillance system based on CNN-LSTM was developed for detecting suspicious human activities with an accuracy of **99.24%**. The system integrates **CNNs** for spa- tial feature extraction, **LSTMs** for sequential pattern analysis, and **automated alert mechanisms** for real-time security.

A. Key Contributions

- Achieved **99.24% accuracy** in identifying suspicious activities.
- Implemented a robust surveillance system framework using an efficient deep learning approach.
- Reduced false alarms through contextual sequence learn- ing.

B. Future Works

Further improvements will focus on enhancing the system's performance by:

- Deploying the model on **edge/AI-based devices** (Rasp- berry Pi, Jetson Nano) for real-time inference.
- Expanding the dataset to cover **diverse environments** for improved robustness.
- Integrating **multi-modal inputs** (thermal imaging, depth sensors) for more effective anomaly detection.

The proposed system demonstrates that deep learning has significant potential in real-world surveillance applications, enabling proactive security monitoring with minimal human intervention.

REFERENCES

[1] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
[2] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
[3] K. Simonyan and A. Zisserman, "Two-Stream Convolutional Networks for Action Recognition in Videos," *NeurIPS*, 2014.
[4] A. Graves, A.-R. Mohamed, and G. Hinton, "Speech Recognition with Deep Recurrent Neural Networks," in *IEEE ICASSP*, 2013.
[5] A. Singh and T. K. Marks, "A Comprehensive Review on Human Action Recognition in Video," *Pattern Recognition Letters, Elsevier*, 2017.
[6] W. Sultani, C. Chen, and M. Shah, "Real-World Anomaly Detection in Surveillance Videos," in *IEEE CVPR*, 2018.
[7] J. Wang, Y. Chen, S. Hao, X. Peng, and L. Hu, "Deep Learning for Sensor-Based Activity Recognition: A Survey," *IEEE TKDE*, 2019.
[8] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *NeurIPS*, 2012.
[9] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement,"
arXiv preprint arXiv:1804.02767, 2018.
[10] M. Brundage, S. Avin, J. Clark, et al., "The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation," *Future of Humanity Institute, University of Oxford*, 2018.
[11] O. Tene and J. Polonetsky, "Big Data for All: Privacy and User Control in the Age of Analytics," *Northwestern Journal of Technology and Intellectual Property*, 2013.
[12] S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," in *ICML*, 2015.
[13] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *IEEE CVPR*, 2016.
[14] M. Abadi, P. Barham, J. Chen, et al., "TensorFlow: A System for Large-Scale Machine Learning," in *OSDI'16: USENIX Symposium on Operating Systems Design and Implementation*, 2016.
[15] X. Shi, Z. Chen, H. Wang, et al., "Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting," *NeurIPS*, 2015.
[16] J. Brownlee, "A Gentle Introduction to Long Short-Term Memory Networks by Examples," *Machine Learning Mastery*, 2019.
[17] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhut- dinov, "Dropout: A Simple Way to Prevent Neural Networks from Overfitting," *Journal of Machine Learning Research*, 2014.
[18] S. Xie, R. Girshick, P. Dolla´r, Z. Tu, and K. He, "Aggregated Residual Transformations for Deep Neural Networks," in *IEEE CVPR*, 2017.
[19] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A Large-Scale Hierarchical Image Database," in *IEEE CVPR*, 2009.
[20] S. Hwang and J. Park, "Anomaly Detection in Surveillance Systems Using Deep Learning Approaches," *Springer AI Review Journal*, 2021.