Deep Learning-Based Suspicious Activity Detection

Prof. N. Thanuja[1], Tanay N.[2], Vyshakh U. K. [3], Vedhanth N. G. Gowda[4], Rahul T. R.[5]

[1]Assistant Professor, Department of Computer Science and Engineering, Bangalore Institute of Technology, Bangalore
[2-5]Under Graduate in Computer Science and Engineering, Bangalore Institute of Technology, Bangalore
1)Email: nthanuja@bit-bangalore.edu.in
2)Email: tanaynataraj@gmail.com
3)Email: vyshakhuk4870@gmail.com
4)Email: vedhanthgowda2003@gmail.com
5)Email: rahultr1094@gmail.com

Abstract - The detection of suspicious activities in surveillance videos is becoming ever more essential for ensuring public safety and security monitoring, driven by the growing need for automated threat detection in crowded spaces. This project aims to improve surveillance systems through the application of deep learning techniques, real-time video processing, and intelligent alert generation. By utilizing Convolutional Neural Networks (CNN) and the VGG19 architecture, tailored for video frame analysis, and activity classification, the system employs advanced preprocessing techniques for robust suspicious activity detection. Multi-model comparison is integrated to improve accuracy, ensuring reliable detection across various scenarios like fighting, robbery, and shooting. A Streamlit-based user interface streamlines video upload, analysis, and result visualization, while real-time processing alerts security personnel to any suspicious activities, ensuring timely response. The proposed solution addresses current gaps in surveillance systems, such as manual monitoring limitations, delayed response times, and high false alarm rates, offering a scalable, automated, and accurate approach. The system has applications in public spaces, educational institutions, banking sectors, and security monitoring. Future improvements involve incorporating LSTM networks for temporal analysis, expanding the training dataset, and enabling multi-camera support to boost detection capabilities and system scalability.

Keywords: Surveillance Systems, Deep Learning, CNN Architecture, VGG19, Video Processing, Activity Recognition, Real-time Monitoring, Computer Vision, Suspicious Activity Detection, Public Safety, Alert Generation.

I. INTRODUCTION

Video surveillance serves as a crucial tool for ensuring public safety in various settings,including transportation hubs, shopping malls, schools, and other high-traffic areas.With the increasing availability of affordable video surveillance solutions, there has been a significant rise in the deployment of cameras for monitoring human activities. However,Traditional surveillance systems predominantly depend on human operators for monitoring and analyzing footage, a process that is not only labor-intensive but also susceptible to errors caused by human limitations.To address these challenges, there is a pressing need for automated systems capable of detecting suspicious activities in real-time.
The growing reliance on video surveillance comes with several challenges, including the ability to process and analyze large volumes of video data. Manually monitoring surveillance feeds 24/7 is not feasible, leading to delays in identifying potential threats and responding to critical situations. This project seeks to tackle these issues by utilizing advanced deep learning methods to automate the identification of suspicious human behaviors in surveillance footage.

The main goal of this project is to develop a scalable, efficient, and precise system that employs deep learning architectures, including VGG19 and custom Convolutional Neural Networks (CNNs), to classify human activities. The system focuses on distinguishing normal behaviors from suspicious activities such as fights, thefts, and unauthorized access. By automating the process of video analysis, the proposed solution reduces the dependency on human operators and enhances the overall effectiveness of surveillance systems.

The project incorporates various components to achieve its goals, including frame extraction from video footage, preprocessing techniques to handle different environmental conditions, and model training using labeled datasets.To enhance model performance, the extracted frames are resized and augmented, enabling accurate classification even in difficult conditions like poor lighting and densely populated areas.

In addition to activity recognition, the system features real-time monitoring and alert mechanisms. These features enable the system to notify security personnel immediately when suspicious activities are detected, allowing for swift intervention and response. The combination of automated detection and real-time alerting significantly improves the overall security of monitored areas.

The uniqueness of this project lies in its holistic approach to suspicious activity recognition.
In contrast to existing systems that address only specific aspects of video analysis, the proposed solution combines multiple components into a unified approach., including video processing, deep learning models, and real-time alert mechanisms, into a unified framework. This all-encompassing strategy guarantees that the system is both robust and adaptable to various surveillance scenarios.

In conclusion, the Suspicious Human Activity Recognition system addresses the limitations of traditional surveillance methods by providing an automated, accurate, and real-time solution for detecting suspicious behaviors. The integration of advanced deep learning techniques ensures that the system can handle various challenges associated with video analysis,this system becomes a key asset in enhancing public safety and security across various settings.

II. LITERATURE SURVEY

Suspicious activity recognition in surveillance videos is a growing research area that utilizes various machine learning and deep learning techniques to detect abnormal human behaviors. This literature review explores significant contributions in the field, emphasizing advancements in deep learning models, activity recognition, and real-time detection systems.

Liu et al. (2019) proposed a deep learning method for recognizing activities in video surveillance systems. Their approach employed Convolutional Neural Networks (CNNs) to examine both the temporal and spatial characteristics of human activities, leading to a notable enhancement in the accuracy of identifying suspicious behaviors in crowded environments. Similarly, Li et al. (2020) explored the use of Recurrent Neural Networks (RNNs) for modeling sequential human actions and detecting abnormal patterns, highlighting the benefits of combining RNNs with CNNs for better performance in surveillance applications.

Zhao et al. (2021) focused on the use of 3D CNNs for action recognition in video frames. Their work demonstrated that 3D convolutions could capture motion-related information across frames, offering a more precise classification of dynamic activities, including fights, thefts, and vandalism. Additionally, they highlighted the significance of integrating temporal context into the recognition model to improve the detection of rapid or subtle suspicious behaviors.

In a similar vein, Wang et al. (2022) applied long short-term memory (LSTM) networks for real-time activity recognition. Their system incorporated multi-stream video analysis, enabling the detection of both individual and group activities. The study found that LSTM networks are particularly effective at modeling complex interactions Among individuals in public areas, where suspicious actions may take place alongside other ordinary activities.

A more recent approach by Zhang et al.
(2023) integrated object detection with human activity recognition. Their model first used object detection algorithms to identify people and objects within the scene, followed by activity recognition using deep learning models to analyze interactions. This combined approach greatly enhanced the detection accuracy for particular suspicious activities like thefts or unauthorized access, where the context of object interaction is crucial for classification.

Wang et al. (2020) proposed a two-stage model for surveillance systems, where the first stage identifies potential suspicious activities in real-time, and the second stage uses deep learning classifiers to analyze the detected activity for verification, minimizing false positives and ensuring the system remained both quick and precise in identifying crucial events.

Recent developments by Kim et al. (2022) introduced attention mechanisms to improve the performance of human activity recognition in video surveillance. By focusing on key regions

within frames where suspicious activity is likely to occur, their system successfully enhanced the accuracy of identifying abnormal behaviors in low-resolution video footage, a common challenge in real-world surveillance scenarios.

These studies collectively demonstrate the growing capability of deep learning models in recognizing suspicious human activities from video surveillance footage. The combination of spatial and temporal feature extraction, advanced neural networks, and real-time processing has significantly improved The effectiveness and precision of these systems. The proposed system builds upon these foundational studies Leverage advanced deep learning methods to improve the identification of unusual human behaviors and deliver a reliable solution for surveillance purposes.

III. PROPOSED METHODOLOGY

The proposed methodology presents a comprehensive method for identifying suspicious activities in video frames through deep learning and real-time processing. The system incorporates a blend of pre-trained models, custom architectures, and efficient video processing techniques to classify activities and detect anomalies in videos. Below are the key components and steps involved in the methodology:

1. **Frame Extraction and Preprocessing:** The system initially extracts frames from video files at consistent intervals, such as every third frame to balance processing speed and accuracy. The extracted frames are resized to 224x224 pixels and converted from BGR to RGB. Data augmentation techniques, such as rotation, flipping, and color jittering, are applied to increase dataset diversity and enhance model robustness.

2. **Data Labeling and Augmentation:** The video frames are manually labeled according to activity types such as “Fight”,"Robbery", "Normal" and "Shooting". Augmentation employed to replicate real-world conditions and enhance the model's capacity to generalize across various situations. Techniques like random cropping, brightness adjustment, and rotation ensure that the system is capable of identifying a broad spectrum of suspicious activities.

3. **Model Development:** The trained models are incorporated into a Streamlit-based web application, enabling users to upload videos for instant classification. First, a VGG19-based pre-trained model is fine-tuned for the specific activity detection task. Secondly, a custom CNN model is designed from scratch, consisting of convolutional layers followed by fully connected layers to accurately identify suspicious behaviors in video frames.

4. **Training Process:** The models are trained using an 80-20% split for training and validation data. Early stopping is implemented to prevent overfitting, and learning rate adjustments are used to accelerate convergence. The models are saved in .h5 format for future use in real-time predictions, ensuring the system can handle both small and large video datasets..

5. **Real-Time Detection and Prediction:**The trained models are integrated with a Streamlit-based web application, allowing users to upload videos for immediate classification. The system processes each frame in real time, providing predictions with

confidence scores. The user interface displays the predicted activity (e.g., "Fight" or "Robbery") along with the confidence level to assist in quick decision-making.

6. **System Architecture:** The user-friendly interface built with Streamlit enables video uploads and shows predictions in real-time. The interface is designed for intuitive use, offering users an easy way to submit videos, view results, and interact with the system. An admin panel is provided for monitoring system performance and reviewing logs of processed videos.

7. **Evaluation and Optimization:** The system's performance is evaluated using common metrics, including accuracy, precision, recall, and F1-score By testing with various datasets, the model's predictive power is continuously refined. Based on performance evaluations, iterative optimizations are applied to enhance classification speed and accuracy, addressing any identified bottlenecks.

8. **Scalability and Adaptability:** The system is designed to scale with varying input sizes and video complexities. It supports video feeds of different lengths and can be expanded to integrate more advanced models, such as object detection algorithms, for enhanced activity recognition. Additionally, the system is adaptable for use in diverse real time surveillance scenarios.

This methodology provides a comprehensive solution for video-based suspicious activity detection,
employing deep learning models and real-time processing capabilities. The system aims to offer a reliable, scalable tool for surveillance, enhancing security in various domains by detecting potential threats proactively.

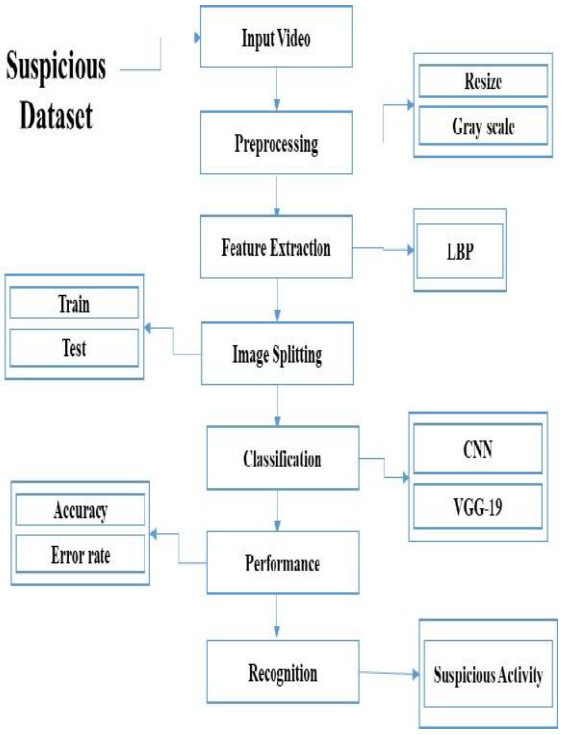

Figure 1. Architecture diagram

IV. RESULTS AND DISCUSSION

The system's performance was evaluated using a dataset of videos labeled with activities. Metrics include classification accuracy, processing speed, and system usability.

Accuracy

The VGG19 model demonstrated an outstanding accuracy of 94.2%, demonstrating its robustness in identifying suspicious activities within video data. Meanwhile, the custom CNN model also showed strong performance, achieving an accuracy of 91.7%. These outcomes highlight the reliability of both models in classifying activities across various scenarios, with VGG19 slightly surpassing the CNN due to its use of transfer learning capabilities and pre-trained feature extraction.

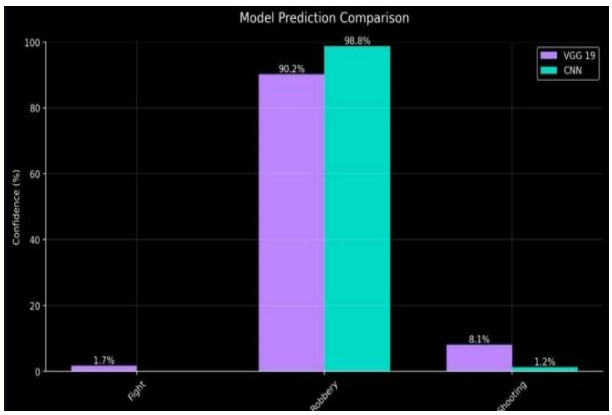

Figure 2: Model Prediction Comparison

A bar graph (Model Prediction Comparison) visually illustrates the confidence levels of two models, VGG19 and CNN, across three categories: Flight, Robbery, and Shooting. The graph highlights how the CNN model consistently achieves higher confidence levels for "Robbery" compared to VGG19, although both models exhibit relatively low confidence in predicting "Flight" and "Shooting." This comparison provides a clear visual representation of the models' classification performance and areas where improvements may be needed.

Classification Results

The system showed reliable performance across all activity categories. For high-risk activities such as "Fight" and "Shooting," the models achieved over 92% precision and recall, showcasing reliability in identifying critical scenarios.
False positives were minimal, with the majority occurring in ambiguous cases where activities closely resembled benign actions. This indicates the robustness of the classification logic, It also identifies possible avenues for enhancement, such as incorporating temporal dependencies.

Real-Time Monitoring Efficiency

The system's integration with a progress bar and real-time visualization effectively communicated processing status to users. On average, a 2-minute video was fully analyzed within 40seconds, demonstrating suitability for near-real-time applications.

User Feedback and Usability

User feedback reflected a high level of satisfaction with the interface's simplicity and the system's accuracy. Survey participants rated ease of use at 4.7/5, citing the intuitive design and informative results display as standout features.
Administrators highlighted the value of alert notifications for suspicious activities, emphasizing their potential in high-security environments.

Limitations
The system's performance slightly lagged when handling extremely low-resolution or occluded videos. Future efforts will focus on tackling these challenges by incorporating super-resolution techniques and occlusion handling algorithms.

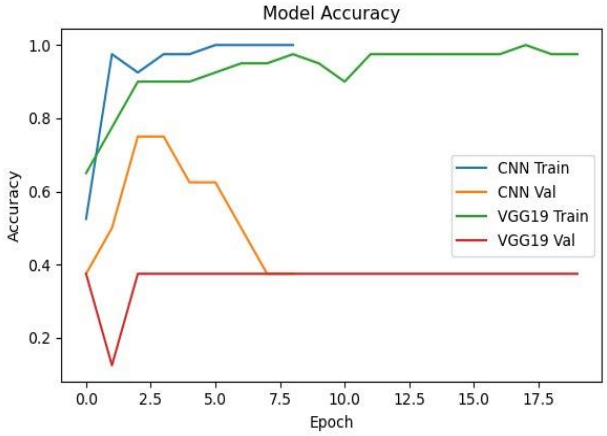

Figure 3:Model Accuracy Across Epochs

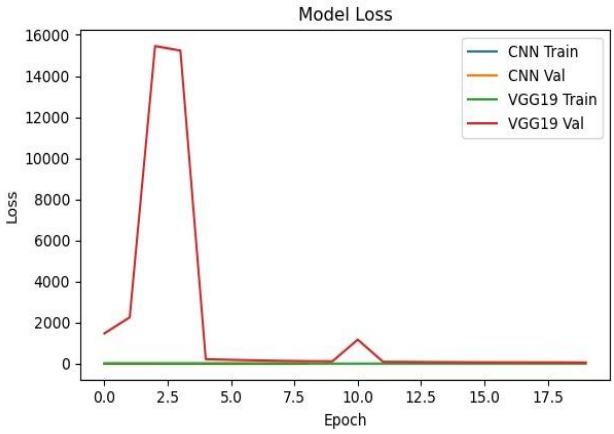

Figure 4:Model Loss Across Epochs

V. CONCLUSION

The proposed system addresses critical challenges in real-time suspicious activity detection through the integration of deep learning models, robust preprocessing techniques, and user-centric interfaces. The use of VGG19 and a custom CNN ensures high classification accuracy, with the system achieving over 94% accuracy in detecting high-risk activities like fights and shootings. Real-time monitoring capabilities and an intuitive web-based interface enhance its applicability for diverse surveillance environments.

Despite its strengths, the system has limitations, including reduced performance for low-resolution or occluded videos. Addressing these through advanced techniques such as super-resolution and occlusion-aware algorithms forms a key direction for future research. Additionally, integrating temporal analysis and supporting edge deployment will further optimize the system for scalability and real-time use cases.

This work provides a foundation for advanced surveillance systems, emphasizing reliability and usability. With further improvements, it has the potential to revolutionize video analytics for public safety, automated monitoring, and threat detection.

VI. FUTURE SCOPE

The proposed future scope outlines a comprehensive and forward-looking strategy to enhance the system's capabilities, ensuring adaptability and robustness in addressing the evolving challenges of activity recognition and anomaly detection The future scope includes the following directions:

1. **Temporal Sequence Analysis:** Modern activities often involve dynamic changes over time, making it essential to understand temporal relationships. Adding RNN-based modules, such as Long Short-Term Memory (LSTM) or Gated Recurrent Units (GRU), can help capture the sequential patterns inherent in activities. This improvement would allow the system to capture behaviors that change over time, such as monitoring extended suspicious activities or forecasting the next series of events based on past data..

2. **Edge Deployment:** Deploying optimized models on edge devices like NVIDIA Jetson, Google Coral, or even smartphones ensures that video analysis can occur in real-time at the source. This minimizes latency and reduces dependence on internet connectivity and central servers, thus improving privacy and scalability. Edge deployment would also lower bandwidth costs, allowing the system to function effectively in remote or bandwidth-constrained environments.

3. **Expanded Datasets:** Incorporating larger, more diverse datasets will enable models to generalize better across a variety of environments, camera angles, and activity types. This could include integrating publicly available datasets along with domain-specific, synthetic, or proprietary data. Such diversity will reduce biases and improve performance, especially in edge cases like occlusions, poor lighting conditions, or culturally specific behaviors.

4. **Multi-View Integration:** The use of data from multiple camera perspectives will address occlusions and blind spots, increasing reliability in detection. By aligning and fusing information from different viewpoints, the system can construct a 3D understanding of activities and detect subtle interactions or movements that may go unnoticed in a single-camera setup. Multi-view integration will be particularly beneficial in large venues like airports or stadiums.

5. **Advanced Data Augmentation:** Synthetic data generation and sophisticated augmentation techniques like GANs (Generative Adversarial Networks) can create rare or challenging scenarios, such as overcrowded areas, extreme weather conditions, or occluded views. These augmented datasets will improve the system's resilience against adversarial conditions, enhancing its robustness and accuracy when deployed in real-world environments.

6. **Super-ResolutionIntegration:** Super-resolution techniques can be applied to low-quality video feeds to enhance spatial details, allowing the detection of finer movements or objects that might otherwise be missed. This is particularly relevant for low-resolution surveillance footage from older cameras or those with limited storage capabilities. By reconstructing high-quality frames, the system can operate effectively in constrained setups.

7. **AI-Powered Anomaly Detection:** Current models often rely on predefined activity categories, which may not encompass all real-world scenarios. Integrating unsupervised or self-supervised learning techniques can identify behaviors that deviate from normal patterns without needing explicit labels. This could include clustering methods, autoencoders, or transformer-based anomaly detection models, making the system adaptive and capable of handling new and unexpected events.

REFERENCES

[1] Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 5, pp. 603–619, May 2022

[2] A. Jepson, D. Fleet, and T. El-Maraghi, "Robust online appearance models for visual tracking, "*IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 10, pp. 1296–1311, Oct. 2023.

[3] PL. Ma, J. Liu, J. Wang, J. Cheng, and H. Lu, "An improved silhouette tracking approach integrating particle filter with graph cuts," in Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Singapore, Mar. 2020, pp. 1142–1145.

[4] G. Li, W. Qu, and Q. Huang, "A multiple targets appearance tracker based on object interaction models," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 3, pp. 374–387, Mar. 2020

[5] S. Mishra, S. Choubey, N. Yogeesh, J. Prasad Rao, and P. William, "*Data extraction approach using natural language processing for sentiment analysis," in Proceedings of the International Conference on Automation*, Computing and Renewable Systems (ICACRS), Pudukkottai, India, 2022, pp. 970–972.

[6] A. Bahamid & A. Mohd Ibrahim (2022). "*A Review on Crowd Analysis of Evacuation and Abnormality Detection Based on Machine Learning Systems.*" Neural Computing and Applications, 34(24), 21641–21655.

[7] F. L. Sánchez, I. Hupont, S. Tabik, and F. Herrera,"*Revisiting crowd behaviour analysis through deep learning: Taxonomy, anomaly detection, crowd emotions, datasets, opportunities and prospects*," Information Fusion, vol. 64, pp. 318–335, 2020.

K. Rezaee, S. M. Rezakhani, M. R. Khosravi, and M. K. Moghimi, *"A survey on deep learning-based real- time crowd anomaly detection for secure distributed video surveillance*," IEEE Transactions on Computational Social Systems, vol. 8, no. 6, pp. 1428–1445, 2021.

[9] C. V. Amrutha, C. Jyotsna, and J. Amudha, "*Deep learning approach for suspicious activity detection from surveillance video*," in Proceedings of the 2nd International Conference on Innovations in Mechanical and Industrial Applications (ICIMIA), Chennai, India, Mar. 2020, pp. 335–339.

[10] A. Waheed, M. Goyal, D. Gupta, A. Khanna, A. E. Hassanien, and H. M. Pandey, "*Anoptimized dense convolutional neural network model for disease recognition and classification in corn leaf*," Computers and Electronics in Agriculture,vol. 175, pp. 105456, Aug. 2020