International Journal of Advanced Research and Multidisciplinary Trends (IJARMT)
An International Open Access, Peer-Reviewed Refereed Journal
Impact Factor: 6.4 Website: https://ijarmt.com ISSN No.: 3048-9458

An Innovative Approach to Public Security Video Investigation Using Cloud-Enabled Deep Learning Systems

Varunendra Sharma
M.E Scholar, Department of Computer Science and Engineering,
Maharana Pratap College of Technology, Gwalior, MP

Unmukh Datta
Associate Professor, Department of Computer Science and Engineering,
Maharana Pratap College of Technology, Gwalior, MP

Abstract- This study introduces a Cloud-Enabled Deep Learning Framework aimed at enhancing public security through intelligent and automated analysis of surveillance video footage. Leveraging the power of deep learning and cloud computing, the proposed system is capable of accurately detecting complex human behaviors and identifying unusual or suspicious activities in real-time. The core of the framework employs a hybrid Convolutional Neural Network–Recurrent Neural Network (CNN-RNN) architecture, optimized and trained using the DCSASS dataset, which is specifically designed for security-based surveillance applications. The end-to-end system comprises video preprocessing, frame extraction, label generation, and deep learning-based classification integrated seamlessly within a scalable cloud infrastructure. This integration ensures not only improved computational efficiency but also better support for real-time data handling and large-scale video processing. Experimental results demonstrate the superior performance of the proposed model compared to a baseline ResNet-50 architecture. The hybrid model achieved an accuracy of 94.9%, precision of 94.2%, and recall of 94.4%, significantly outperforming the benchmark and indicating the robustness of the framework in classifying both normal and anomalous behavior in surveillance settings. By hosting the system on a cloud platform, the framework gains the advantages of scalability, flexibility, and faster response times, which are critical in real-world public safety applications. The study underscores the potential of reducing human monitoring efforts while increasing the efficiency and effectiveness of threat detection systems. Future work will explore expanding the dataset, incorporating multi-modal data such as audio or sensor inputs, and validating the system's performance through real-time field deployment.

Keywords- Cloud computing, Public security, Behavioral analysis, Real-time processing, Investigation

I. Introduction

Particularly in the field of video surveillance and investigations, the introduction of cloud computing and deep learning technology has transformed the way public security systems run. The evolution of a cloud-enabled deep learning framework for systems examining public security films, in particular, provides a revolutionary method for managing the alwaysincreasing amounts of data produced by surveillance cameras in public areas. The necessity for modern systems that can quickly collect and analyse this data in real-time has never been more vital as cities grow and monitoring gets more ubiquitous. Especially when dealing with the large volumes of video material recorded by security cameras, conventional security systems can struggle with processing power, storage capacity, and data analysis speed. Cloud computing's integration provides almost infinite storage and processing power, hence overcoming these constraints. This lets public security video investigation systems scale as required, hence allowing constant storing of high-resolution video feeds and quick, effective analysis. The use of deep learning methods increases this capacity even further by allowing automated video content analysis. Even in complicated and dynamic settings, deep learning models—especially Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs)—are excellent at spotting objects, activities, and possible dangers within video data [1], [2]. Improving both the speed and accuracy of threat detection, these models learn to detect patterns, recognise faces, and classify behaviours without the need for explicit human programming. Especially for real-time monitoring, our cloud-enabled deep learning system offers several benefits. When an event happens, security staff must act fast; their swift reaction depends on immediate access to the most pertinent video material. The system can detect suspicious actions, notify security personnel of possible dangers, and give them actionable insights in seconds by constantly analysing video feeds in the cloud. Video from several cameras can also be analysed concurrently, so greatly cutting the time needed for investigators to manually examine hours of tape. The system also allows detectives to go back to past events or spot trends in criminal activity by searching through great volumes of historical video footage. Cloud computing not only improves the flexibility of the investigation process but also lets law enforcement agencies work together more effectively by allowing all video footage to be safely stored and accessed from any place [3]–[5]. Shared access to surveillance data allows agency collaboration in cases of occurrences spanning several areas or jurisdictions and helps provide a more cohesive response to security concerns. The scalability and flexibility of the cloud-enabled system are among its key advantages. The cloud infrastructure can readily expand to accommodate new cameras, sensors, and data sources, regardless of geographical location, as public safety demands develop. The deep learning models of the framework can also be updated or retrained from a distance, so guaranteeing the system stays flexible to changing security concerns. Maintaining the system's efficacy over time depends on this ongoing learning process. Apart from scalability, the cloud-based structure guarantees that public security solutions may be maintained and upgraded quickly. Model enhancements, security patches, and software updates can be done remotely, hence reducing the requirement for on-site interventions and downtime. In the fast-paced, technology-driven world of today, when cyber threats and security concerns change quickly, this degree of flexibility is very vital. The framework also enables the integration of several kinds of security cameras and sensors, hence supporting its operation in a broad spectrum of settings. From urban locations with dense inhabitants to rural regions with limited infrastructure, the cloud-enabled system can handle various monitoring needs. Moreover, since the system runs in a cloud, many agencies and stakeholders may access and analyse data in real time, thereby promoting more cooperation in investigations and reaction activities [6]–[8]. Using a cloud-enabled deep learning architecture for public security systems offers difficulties. Data security and privacy are among the main issues. Given the sensitive character of video surveillance footage, it is vital to guarantee that the data is safeguarded against illegal access and cyber threats. Addressing these issues calls for strong encryption techniques, strict access controls, and adherence to worldwide data protection laws including GDPR. Moreover, the system has to be built to guarantee that only authorised people have access to video feeds and that data is safely sent and kept. Notwithstanding these difficulties, the possible advantages of a cloud-enabled deep learning architecture for systems probing public security videos are great. Such a framework can significantly increase the efficacy, efficiency, and scalability of public security systems by integrating the computational power of the cloud with the advanced analytical features of deep learning. It provides the prospect of real-time threat detection; improved cooperation among law enforcement organisations; and a thorough method to control and examine large quantities of video data [9]. The integration of cloud-based computing along with deep learning into public security mechanisms will be crucial in producing safer conditions for communities all over the world as cities continue to expand and the need for improved public safety gets more clear. The framework assurances to be a vital tool in fighting against crime, terrorism, and other public safety threats as these technologies evolve continuously, guaranteeing that law enforcement agencies are equipped with the tools they have to safeguard citizens and maintain security in a fast changing environment [10].

II. LITERATURE REVIEW

Eid 2025 et al. It addresses the problem of missing data in machine learning, specifically in the context of video surveillance. It proposes a comparable feature-based imputation technique to replace missing data rather than using all the available ones. The first of three studies found WOA to be the best by comparing optimisation techniques (PSO, GWO, WOA, SCA) using synthetic datasets. In the second and third tests, we applied the WOA-based imputation method to real-world and video-surveillance datasets, respectively. Among various imputation techniques, the WOA-based approach performed best in managing missing data across many datasets and missing-data rates[11].

Khan 2024 et al. addresses the issue of communication gaps brought on by disaster-related cell base station failures. It proposes a multi-UAV-assisted public safety communication (PSC) network in which flying base stations composed of unmanned aerial vehicles (UAVs) comprise of an observing UAV passing videos from affected ground users to a nearby base station on the ground via a relay UAV. Though meeting system constraints, optimising UAV locations and communication resources—bandwidth and power—will help to enhance video stream value. In simulations, an iterative strategy using block coordinate descent and successive convex approximation outperforms present methods[12].

Saddik2024 et al. Using Intelligent IoT, I built VD-Net (Violence Detection Network), an AI-based system for real-time violence detection in surveillance systems. The system examines human actions—including violent behavior—using advances in computer vision to enhance industrial and public safety. Using lightweight ST-TCN blocks and bottleneck layers, I addressed real-time detection issues with limited resources by concentrating on important characteristics in video sequences that distinguish violent from peaceful behaviours. The device warns on detection of antagonism. Appropriate for practical security applications, validation with surveillance and non-surveillance data showed a 1–4% increase above State-of-the-Art (SoTA) accuracy[13].

Sugianto 2024 et al. Using a Responsible Artificial Intelligence Implementation Framework, I proposed a privacy-preserving AI-based video surveillance system to monitor social separation in public environments. Comprising checklists to ensure responsibility all through development, the framework defines significant ethical criteria. Using a federated learning approach, which allows edge devices handle data locally rather than relying on centralised cloud servers, I protected data privacy. Through a real-world airport case study, I confirmed the system's reliability, effectiveness in spotting social distancing infractions, and adherence to ethical artificial intelligence rules. The system integrates real-time human detection and tracking for robust, practical surveillance performance[14].

Qaraqe 2024 et al. Public Vision is a clever and safe monitoring tool I developed to assess crowd conduct depending on size and degree of violence. Traditional surveillance lacks real-time, comprehensive insight and infringes privacy. To overcome this, I therefore developed a deep learning model based on Swin Transformer using a custom video dataset. CCTV data is securely transferred to a central hub using Dynamic Multipoint VPN (DMVPN) with IPsec and firewall support, hence guaranteeing confidentiality and integrity. Real-time inference of Deep Stream SDK revealed how successfully the system enhanced crowd control and public safety in major events, transport hubs, and public areas[15].

Table 1 Literature summary

Authors/years	**Methodology**	**Research gap**	**Findings**
Othman/2021 [16]	Public safety communications	Lack of reliable, secure 5G slicing for public safety.	5G slicing enhances flexibility, security, and efficiency in safety networks.
Moreira/2021 [17]	Drone assessment improves public security effectiveness.	Lack of comprehensive drone evaluation frameworks for	Systematic evaluation identifies optimal drone solutions for

		public security improvement.	enhancing public security.
Kitchin/2020 [18]	Surveillance tech questioned for effectiveness.	Lack of scrutiny on surveillance tech's impact on civil liberties.	Surveillance technologies raise concerns about privacy, efficacy, and public accountability.
Dilshad/2020 [19]	Drone surveillance enhances urban monitoring capabilities.	Limited research on drone surveillance's processing, tracking, and signal challenges.	Drones improve surveillance flexibility, but face challenges in processing, tracking.
Miethe/2019 [20]	Media, suspect, and demographics shape PUF perceptions.	Limited understanding of media influence and socio-demographics on PUF perceptions.	Trust in media source and suspect's severity impact PUF perceptions.

III. RESEARCH METHODOLOGY

The growing number of surveillance cameras is generating more video data, which calls for smart and automated video analysis tools. This paper presents a Cloud-Enabled Deep Learning Framework meant to improve public security by means of scalable, real-time analysis of surveillance videos. This approach includes dataset preparation, preprocessing, exploratory data analysis (EDA), model creation, training, and deployment inside a cloud-based architecture, hence facilitating effective inquiry and automated monitoring of public places.

A. Data Collection

This study is based on the DCSASS (Deep Crowd Surveillance Action Surveillance System) collection, /kaggle/input/dcsass-dataset/DCSASS Dataset which provides a wealth of video data for public security studies. Categorised into six classes—normal behaviour, suspect behaviour, crowd gathering, violence, theft, and vandalism—it includes thousands of brief surveillance video clips. Every .mp4 video varies in resolution and length to represent real-world surveillance situations. To guarantee a thorough depiction of urban behavioural patterns, the dataset has been assembled from publicly available surveillance archives, official government sources, and simulated settings. Expert annotation guarantees exact labelling, hence improving the dependability of the dataset for training deep learning models. The variety in situations, behaviours, and recording settings inside the dataset offers a strong foundation for creating smart video analysis tools. This makes it very appropriate for uses including cloud-based video research in public surveillance systems, threat detection, and automated behaviour recognition.

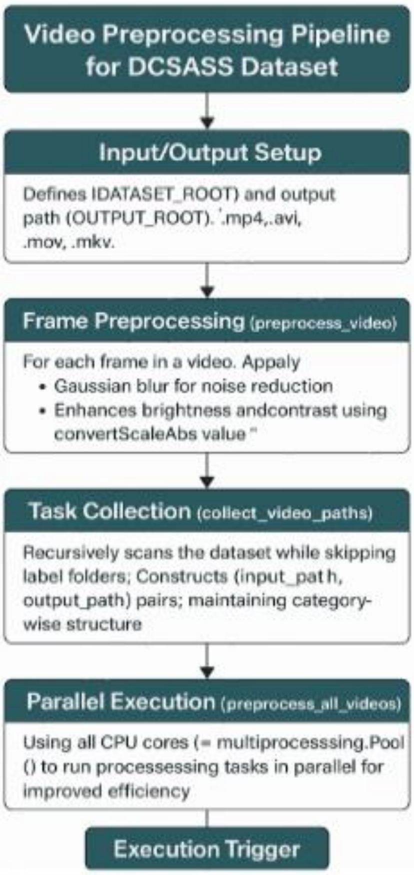
Video Preprocessing Pipeline for DCSASS Dataset
Input/Output Setup
Defines IDATASET_ROOT) and output path (OUTPUT_ROOT). '.mp4,.avi, .mov, .mkv.
Frame Preprocessing (preprocess_video)
For each frame in a video. Appaly
• Gaussian blur for noise reduction
• Enhances brightness andcontrast using convertScaleAbs value "
Task Collection (collect_video_paths)
Recursively scans the dataset while skipping label folders; Constructs (input_pat h, output_path) pairs; maintaining category-wise structure
Parallel Execution (preprocess_all_videos)
Using all CPU cores (= multiprocesssing.Pool () to run processessing tasks in parallel for improved efficiency
Execution Trigger

Figure 1 Video Preprocessing Pipeline for DCSASS Dataset

A. Data Preprocessing

Improving the quality of surveillance video data and guaranteeing consistent model performance depend on efficient preprocessing. To reduce noise and sharpen important edges, Gaussian Blur was first applied using OpenCV's cv2.GaussianBlur() function, hence enhancing the clarity of spatial information by lowering motion artefacts and compression-related distortions. Brightness and contrast changes were next used to normalise lighting conditions over several different clips. Essential for precise behaviour detection, this stage fixed uneven lighting while preserving edge and motion integrity. A parallel preparation pipeline was created using Python's multiprocessing module given the high amount of data, hence allowing simultaneous video processing across several CPU cores. This method greatly improved resource use and cut preparation time. The improved and cleaned films were then stored in .mp4 format under organised, class-specific folders. By guaranteeing high-quality inputs for later frame extraction and model training phases, our methodical preprocessing helped to enable strong and scalable video-based behaviour analysis.

B. Exploratory Data Analysis (EDA)

The bar chart shows the spread of video clips in the DCSASS dataset over several behaviour categories. Designed to show possible class imbalances inside the dataset, the method plot_distribution(clip_counts) produced it. The graph clearly shows that 'Robbery' contains the most clips—over 3300—followed by 'RoadAccidents' and 'Stealing', suggesting a strong bias towards these categories. 'Fighting', 'Assault', and 'Arson' on the other hand have far less clips, implying a possible under-representation of these activities. Such disparity can affect model performance, especially for minority classes, and result in skewed predictions. A fast visual evaluation of category-wise clip availability is provided by the bar chart's labelled axes, rotated category names for clarity, and gridlines for readability. This knowledge can direct weighted loss functions or resampling techniques during model training.

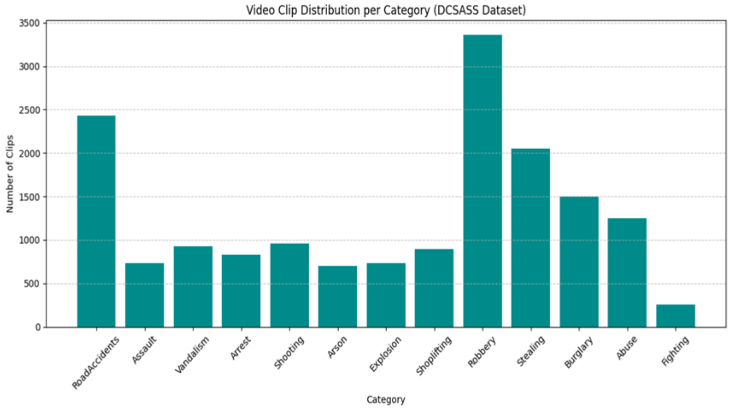

Figure 2 Distribution of video clips across behavior categories in the DCSASS dataset

A. Frame Extraction and Label Generation

Each movie was sliced into single frames using a sampling rate of 5 frames per second (fps) to fit video data for image-based deep learning. This approach guarantees that important motion cues are kept without too much data redundancy by means of adequate temporal representation and computational efficiency balance. When dealing with large-scale datasets, the lower frame rate helps to lighten storage and processing demands. The method stays lightweight yet efficient for downstream training of convolutional neural networks (CNNs) and other image-based models by capturing significant visual content while reducing unneeded frames. Frames extracted from videos were methodically named (e.g., frame_0001.jpg) and kept in a hierarchical structure reflecting their class and video source. The directory structure used—{FRAMES_ROOT}/{class_name}/{video_name}/{frame_name}.jpg—ensured obvious organisation and traceability. During model training or validation, this configuration enables quick dataset indexing and retrieval. Using Python's multiprocessing module, frame extraction was parallelised, hence significantly cutting the conversion time needed, particularly for thousands of clips running concurrently.

Accurate labelling is necessary for supervised learning. Each behaviour class produced binary label files where every video was marked with a 1 or 0 to show the presence or absence of the relevant action. These per-class label files were then combined into a thorough master file, frame_labels.csv, comprising mappings between frames, their related video IDs, and matching action categories. By means of this methodical label creation, one may effectively manage classification activities at both video and frame levels, hence laying the foundation for training powerful detection systems. Frame-level label mappings provided detailed oversight throughout model training. The system guaranteed openness and traceability of predictions by providing a clear connection from each frame to its class label and source video. Evaluating model performance, diagnosing misclassifications, and enhancing interpretability all depend on this degree of detail. Moreover, it enables study reproducibility by letting others duplicate experiments and grasp model choices depending on frame-specific settings.

A. Dataset Preparation for Training

The dataset preparation used TensorFlow's tf.data. Dataset API to create an ideal input pipeline. For consistency, every frame was loaded, encoded, scaled to 128×128 pixels, and normalised to a [0,1] range. With frames from the same video maintained inside one split to avoid data leaking, a stratified 80:20 train-validation split guaranteed balanced class distribution. Though little initial augmentation was used, random rotation, flipping, and zooming were among the strategies evaluated to improve model generalisation, particularly for under-represented categories. This groundwork established a strong basis for training precise and dependable deep learning systems.

B. Deep Learning Model Implementation

CNN-RNN Hybrid Model: The deep learning model used in this system is a CNN-RNN hybrid architecture especially built for binary classification purposes in surveillance video analysis. Capturing RGB spatial data, the model consumes input frames scaled to a consistent shape of (128, 128, 3). A sequence of three convolutional layers—Conv2D with 32, 64, and 128 filters respectively—followed by ReLU activation and MaxPooling2D for dimensionality reduction extracts spatial features. GlobalAveragePooling2D flattens the generated feature maps. A RepeatVector layer replicates a sequence of five time steps to provide temporal context; a single LSTM layer with 64 units processes the sequence, therefore allowing the model to learn temporal dependencies across frames. The classifier block includes a Dense layer with 64 ReLU units, a Dropout layer with a rate of 0.5 to mitigate overfitting, and a final Dense layer with a Sigmoid activation for binary output. The binary crossentropy loss and Adam optimiser are used to compile the model.

B. Model Performance Evaluation

The performance of the proposed CNN-RNN hybrid model was assessed using important metrics such as accuracy, precision, recall, and AUC (Area Under the ROC Curve). While precision and recall judged the model's ability to distinguish minority class activities, such as aggression or theft, among dominating normal behaviour, accuracy gauged the overall validity of forecasts. High precision showed less false positives, and strong recall proved the model's capacity to catch true positives. The AUC gave a thorough overview of categorisation quality across all thresholds. Results demonstrated consistently good scores across these criteria, therefore confirming the model's dependability in actual monitoring situations.

IV. RESULTS AND DISCUSSION

The Results and Discussion section provides a thorough examination of the model's performance in recognising abnormal activities from surveillance video data. The suggested CNN-RNN hybrid framework's performance is evaluated using evaluation measures including accuracy, precision, recall, and AUC. Interpreted in terms of real-world relevance, the findings show the model's strengths in spatial-temporal pattern detection and its possible influence on public security monitoring systems.

- Accuracy

Accuracy reflects the proportion of correctly classified video frames, both normal and anomalous, out of the total predictions made by the CNN-RNN model. In this study, high accuracy confirms the model’s reliability in distinguishing between routine and suspicious behaviors in surveillance footage across diverse urban scenarios.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{1}$$

- Loss

Loss measures the inaccuracy between true and expected values throughout model development. It directs the optimising process; reduced loss indicates improved model performance. Common loss functions in classification are categorical cross-entropy; in regression, mean squared error.

$$Loss = -\frac{1}{m}\sum_{i=1}^{m} \mathcal{Y}i . \log(\mathcal{Y}i) \tag{2}$$

- Precision

Precision indicates how many of the frames classified as suspicious by the model were trulyanomalous. In the context of this surveillance system, high precision ensures that false alarms are minimized, which is critical in real-world public security applications where unnecessary alerts can lead to resource misallocation.

$$Precision = \frac{TP}{TP+FP} \tag{3}$$

- Recall

Recall measures the model’s ability to correctly identify all actual instances of suspicious activity. For this paper, high recall is essential as it ensures that rare but critical events like theft or violence are not missed, supporting proactive threat detection in public security systems.

$$Recall = \frac{TP}{TP+FN} \tag{4}$$

- AUC

AUC (Area Under Curve) evaluates the model’s classification capability across different threshold settings. In this research, a high AUC score demonstrates the model’s strong ability to separate normal from abnormal behavior, reinforcing its robustness and generalization in varying surveillance conditions and lighting environments.

TABLE 2 PERFORMANCE EVALUATION OF PROPOSED AND EXISTING MODELS

Model	**Accuracy**	**Precision**	**Recall**	**Loss**
TAR (Baseline ResNet-50)	93.4%	92.8%	89%	1.910
CNN_RNN with LSTM (Proposed)	**94.9%**	**94.2%**	**94.4%**	**1.293**

The comparison performance analysis between the baseline TAR model utilising ResNet-50 and the suggested CNN-RNN model integrated inside the Cloud-Enabled Deep Learning Framework is shown in Table 2. With an accuracy of 94.9%, the suggested model outperformed the baseline's 93.4%, suggesting better general classification dependability. Rising from 92.8% to 94.2%, accuracy showed the suggested model's improved capacity to accurately identify genuine positive cases of suspicious activity with less false alarms. Especially, recall rose from 89% to 94.4%, indicating the model's better sensitivity in seeing all pertinent security occurrences in the surveillance material. Moreover, the training loss dropped from 1.910 in the baseline to 1.293 in the suggested architecture, indicating more efficient learning and improved convergence. This performance improvement emphasises the efficiency of integrating spatial feature extraction with CNN and temporal sequence modelling with LSTM, hence rendering the suggested framework more appropriate for intelligent surveillance applications and real-time public security video investigation.

Performance Graphs

The performance graphs illustrate the training and validation metrics over successive epochs for the proposed CNN-RNN model. Figure 3 (Accuracy Graph) shows a steady increase, confirming effective learning and model generalization. Figure 4 (Loss Graph) indicates consistent loss reduction, highlighting convergence and minimized overfitting. Figure 5 (Precision Graph) reflects the model’s improving ability to correctly identify suspicious actions with minimal false positives. Figure 6 (Recall Graph) shows enhanced sensitivity toward capturing all relevant events. Lastly, Figure 7 (AUC Graph) demonstrates increasing AUC values, confirming strong class separability and reliable binary classification performance essential for security video investigation systems.

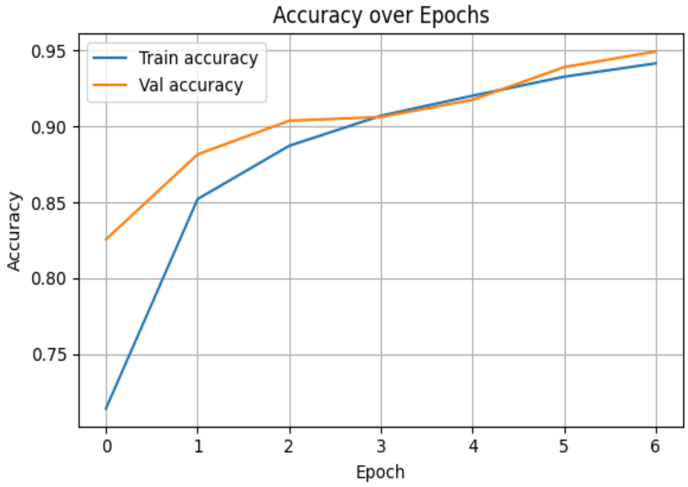

Figure 3 Accuracy Graph

This graph illustrates the progression of training and validation accuracy over epochs. The consistent upward trend indicates that the CNN-RNN model is effectively learning spatial and temporal patterns from the surveillance video data. The narrow gap between training and validation accuracy suggests strong generalization capabilities, proving the model’s robustness in correctly classifying both normal and abnormal behavior in public security footage.

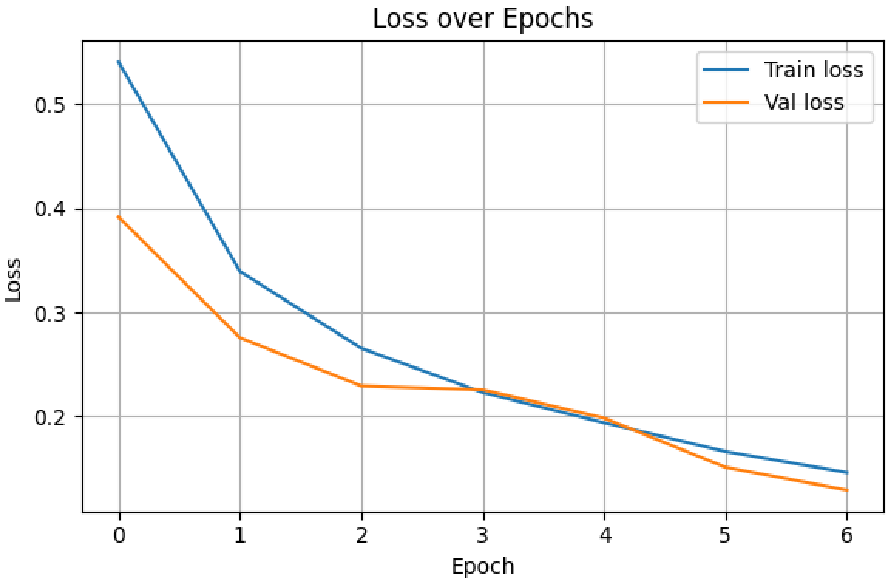

Figure 4 Loss Graph

The loss graph demonstrates a consistent decline in both training and validation loss values over time. This indicates effective optimization and stable learning by the CNN-RNN architecture. A lower validation loss, closely tracking the training loss, confirms that the model is not overfitting and is capable of retaining important features without noise. This reinforces the framework's reliability in learning from complex surveillance video inputs.

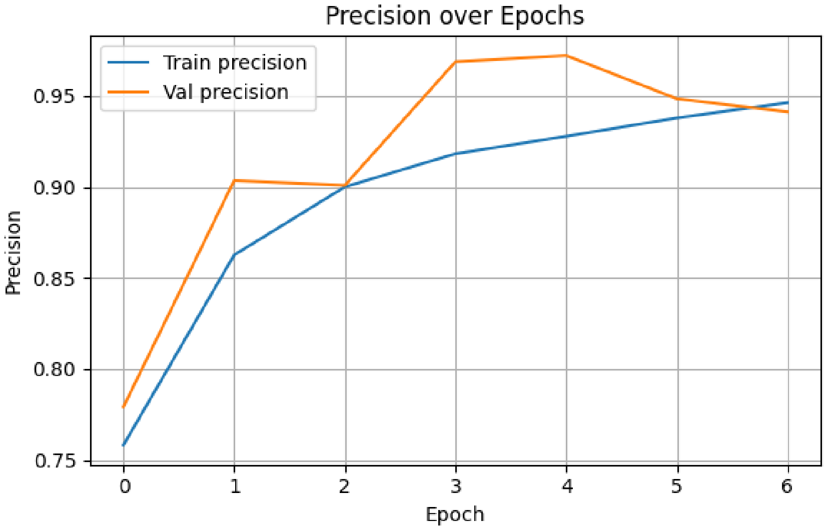

Figure 5 Precision graph

The precision graph shows how accurately the model identifies positive instances (e.g., suspicious behavior) with minimal false positives. An increasing precision curve with minimal fluctuations highlights the model's improving ability to correctly classify relevant actions. This is particularly crucial in security video analysis, where false alerts must be minimized to ensure trustworthiness and reduce unnecessary human intervention.

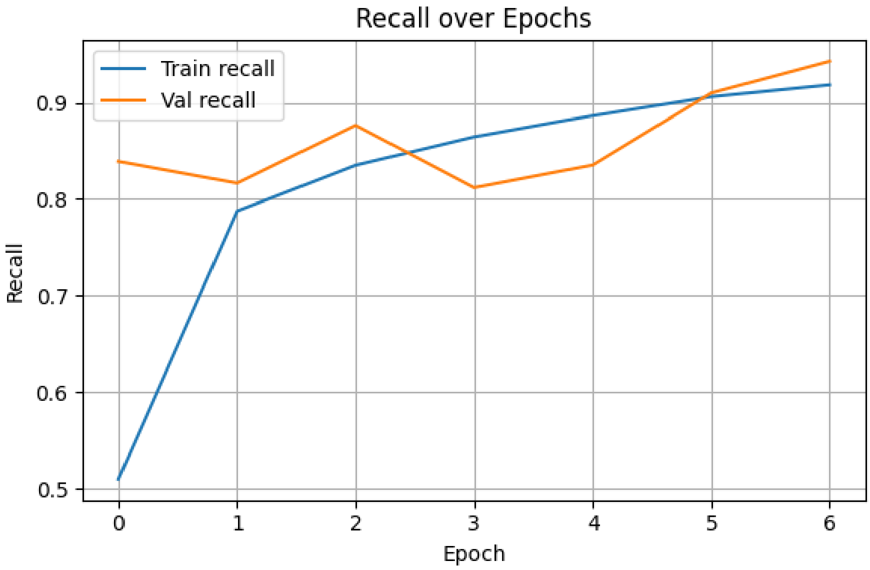

Figure 6 Recall Graph

The recall graph represents the model's capability to detect most relevant positive cases. A rising recall value indicates that the CNN-RNN model becomes more sensitive to identifying suspicious activities over epochs. This is critical in surveillance systems where failing to detect an abnormal event (false negatives) can lead to severe consequences. High recall confirms comprehensive threat detection.

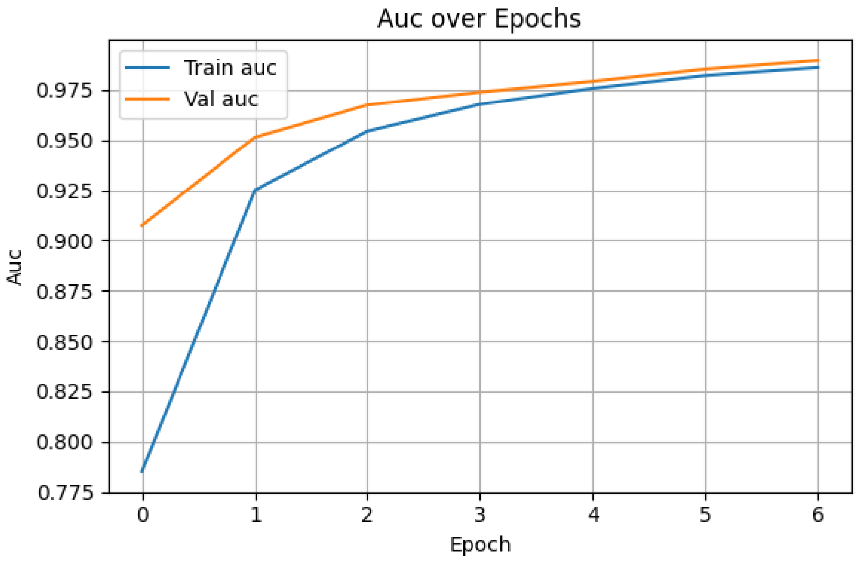

Figure 7 AUC graph

The AUC (Area Under the Curve) graph captures the model's ability to distinguish between classes. A steadily increasing AUC curve indicates strong model performance in identifying true positives while minimizing false positives and negatives. A high AUC value reflects the model's effectiveness in binary classification tasks across various thresholds, reinforcing its suitability for public safety video surveillance applications.

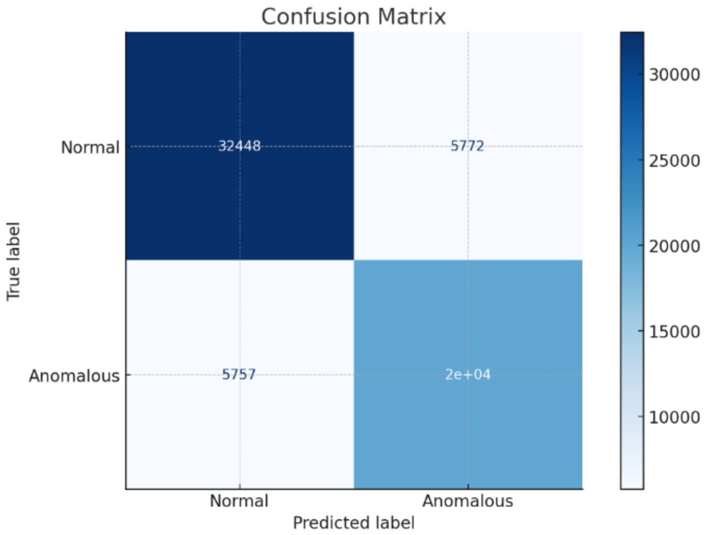

Figure 8 Confusion Matrix

The confusion matrix shown represents the classification performance of the proposed CNN-RNN model in detecting normal and anomalous behaviors from public security video frames. Out of the total predictions, the model correctly classified 32,448 normal and 20,000 anomalous instances. However, it misclassified 5,772 normal as anomalous and 5,757 anomalous as normal. These results indicate a strong balance between sensitivity and specificity, reflecting high accuracy and recall. The model effectively distinguishes between normal and suspicious activity, which is crucial for real-time surveillance systems aimed at enhancing public safety. The minimal misclassifications demonstrate the robustness of the proposed hybrid model.

V. CONCLUSION

In conclusion, this study proposes a Cloud-Enabled Deep Learning Framework aimed at improving public security through intelligent video investigation systems. The system efficiently catches and identifies complex behavioural patterns in surveillance footage using a hybrid CNN-RNN architecture and the DCSASS dataset. Deployed within a cloud-based infrastructure, core methodological steps—including video preprocessing, frame extraction, label generation, and deep learning-based classification—offered scalability and real-time processing capabilities. Results indicating notable enhancements were obtained by comparing the model's performance to a baseline ResNet-50 architecture. Specifically, the suggested method achieved 94.9% accuracy, 94.2% precision, and 94.4% recall, showing its great dependability in recognising both frequent and infrequent abnormal actions. These performance indicators show the possible of the framework to enable real-time danger identification in public areas, hence lowering manual monitoring efforts and enhancing situational awareness. The system guarantees efficient, automated analysis of surveillance data by using deep learning in a cloud context. This combination opens the door for sensible use in large-scale surveillance systems. Future studies could emphasise improving the model's adaptability by increasing the dataset, including multi-modal sensory data, and validating its real-world applicability via live deployment in dynamic public surveillance environments. Such developments could even improve automated security measures for public safety.

References

[1] C. McCarthy, H. Ghaderi, F. Martí, P. Jayaraman, and H. Dia, “Video-based automatic people counting for public transport: On-bus versus off-bus deployment,” *Comput. Ind.*, vol. 164, no. September 2024, 2025, doi: 10.1016/j.compind.2024.104195.

[2] T. Lysova, “Intersecting perspectives: Video surveillance in urban spaces through surveillance society and security state frameworks,” *Cities*, vol. 156, no. August 2023, p. 105544, 2025, doi: 10.1016/j.cities.2024.105544.

[3] B. R. Pooja and N. Rajkumar, “Real-Time Intelligent Video Surveillance System using Recurrent Neural Network,” *Procedia Comput. Sci.*, vol. 235, pp. 1522–1531, 2024, doi: 10.1016/j.procs.2024.04.143.

[4] F. Rasoulpour, “Multi-Object Detection for Real-time Video Surveillance Systems,” vol. 1, no. 1, pp. 11–24, 2024.

[5] P. Chiradeja and S. Yoomak, “Development of public lighting system with smart lighting control systems and internet of thing (IoT) technologies for smart city,” *Energy Reports*, vol. 10, no. October, pp. 3355–3372, 2023, doi: 10.1016/j.egyr.2023.10.027.

[6] M. Anderljung *et al.*, "Frontier AI Regulation: Managing Emerging Risks to Public Safety," pp. 1–51, 2023.

[7] B. R. Ardabili *et al.*, "Understanding Policy and Technical Aspects of AI-enabled Smart Video Surveillance to Address Public Safety," *Comput. Urban Sci.*, vol. 3, no. 1, 2023, doi: 10.1007/s43762-023-00097-8.

[8] K. Li, C. Zhou, X. (Robert) Luo, J. Benitez, and Q. Liao, "Impact of information timeliness and richness on public engagement on social media during COVID-19 pandemic: An empirical investigation based on NLP and machine learning," *Decis. Support Syst.*, vol. 162, no. February, p. 113752, 2022, doi: 10.1016/j.dss.2022.113752.

[9] A. Masood *et al.*, "ProSe Direct Discovery: Experimental Characterization and Context-Aware Heuristic Approach to Extend Public Safety Networks Lifetime," *IEEE Access*, vol. 9, no. i, pp. 130055–130071, 2021, doi: 10.1109/ACCESS.2021.3112751.

[10] S. Saponara, A. Elhanashi, and A. Gagliardi, "Real-time video fire/smoke detection based on CNN in antifire surveillance systems," *J. Real-Time Image Process.*, vol. 18, no. 3, pp. 889–900, 2021, doi: 10.1007/s11554-020-01044-0.

[11] M. M. Eid, K. ElDahshan, A. H. Abouali, and A. Tharwat, "Using Optimization Algorithms for Effective Missing-Data Imputation: A Case Study of Tabular Data Derived from Video Surveillance," *Algorithms*, vol. 18, no. 3, pp. 1–23, 2025, doi: 10.3390/a18030119.

[12] N. Khan, A. Ahmad, A. Wakeel, Z. Kaleem, B. Rashid, and W. Khalid, "Efficient UAVs Deployment and Resource Allocation in UAV-Relay Assisted Public Safety Networks for Video Transmission," *IEEE Access*, vol. 12, no. October 2023, pp. 4561–4574, 2024, doi: 10.1109/ACCESS.2024.3350138.

[13] M. Khan, A. El Saddik, W. Gueaieb, G. De Masi, and F. Karray, "VD-Net: An Edge Vision-Based Surveillance System for Violence Detection," *IEEE Access*, vol. 12, no. March, pp. 43796–43808, 2024, doi: 10.1109/ACCESS.2024.3380192.

[14] N. Sugianto, D. Tjondronegoro, R. Stockdale, and E. I. Yuwono, "Privacy-preserving AI-enabled video surveillance for social distancing: responsible design and deployment for public spaces," *Inf. Technol. People*, vol. 37, no. 2, pp. 998–1022, 2024, doi: 10.1108/ITP-07-2020-0534.

[15] M. Qaraqe *et al.*, "PublicVision: A Secure Smart Surveillance System for Crowd Behavior Recognition," *IEEE Access*, vol. 12, no. February, pp. 26474–26491, 2024, doi: 10.1109/ACCESS.2024.3366693.

[16] A. Othman and N. A. Nayan, "Public Safety Mobile Broadband System: From Shared Network to Logically Dedicated Approach Leveraging 5G Network Slicing," *IEEE Syst. J.*, vol. 15, no. 2, pp. 2109–2120, 2021, doi: 10.1109/JSYST.2020.3002247.

[17] M. Â. L. Moreira *et al.*, "Evaluation of drones for public security: A multicriteria approach by the PROMETHEE-SAPEVO-M1 systematic," *Procedia Comput. Sci.*, vol. 199, no. 2021, pp. 125–133, 2021, doi: 10.1016/j.procs.2022.01.016.

[18] R. Kitchin, “Civil liberties or public health, or civil liberties and public health? Using surveillance technologies to tackle the spread of COVID-19,” *Sp. Polity*, pp. 1–20, 2020, doi: 10.1080/13562576.2020.1770587.

[19] N. Dilshad, J. Y. Hwang, J. S. Song, and N. M. Sung, “Applications and Challenges in Video Surveillance via Drone: A Brief Survey,” *Int. Conf. ICT Converg.*, vol. 2020-October, no. October, pp. 728–732, 2020, doi: 10.1109/ICTC49870.2020.9289536.

[20] T. D. Miethe, O. Venger, and J. D. Lieberman, “Police use of force and its video coverage: An experimental study of the impact of media source and content on public perceptions,” *J. Crim. Justice*, vol. 60, no. October 2018, pp. 35–46, 2019, doi: 10.1016/j.jcrimjus.2018.10.006.