IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

SECURE VISION AI: DEEP LEARNING–BASED REAL-TIME VIOLENCE DETECTION

Bindu. S, K N Rakshitha, Lavanya. S, Saniya Khanum
Department of CSE(DS)
Shivamogga, India

Ms. Sneha
Department of CSE(DS)
Shivamogga, India

Abstract:

Secure Vision AI tackles the rising problem of violence and suspicious activity in places under video surveillance. Right now, most systems lean on people to watch the feeds, which means they miss things or respond too slowly. This project takes a different approach—it uses deep learning to catch violent behavior in real time, without waiting for a human to spot it. The process starts by grabbing live video, cleaning it up, and feeding it into a powerful hybrid model. Here, a Convolutional Neural Network (CNN) breaks down what's happening in each frame, while a Long Short-Term Memory (LSTM) model tracks how things change from one moment to the next. The team tests the system with real data, checking numbers like accuracy and confidence to see which setup works best. Once they've tuned the model, they roll it out with a simple interface, instant alerts, and secure storage powered by Supabase. The result? A smarter, faster, and more reliable way to keep an eye on things—without missing a beat.

Index Terms – Secure Vision AI, Violence Detection, Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), Deep Learning, Real-Time Surveillance, Supabase, Computer Vision.

I. INTRODUCTION

Surveillance cameras are everywhere these days. They're supposed to keep us safe, and sometimes they do. But, let's be honest, they create their own set of problems. Threats often slip by unnoticed because someone has to sit and watch hours of footage, and people just miss stuff. That's a big hole in security. So, that's where automated violence detection comes in. With AI and Computer Vision—especially Deep Learning—these systems can spot patterns in video that people might miss. For this project, I'm building a solid, scalable setup that can catch violent behavior in real time. I'm combining Convolutional Neural Networks (CNN) with Long Short-Term Memory (LSTM) models to do the heavy lifting. The system doesn't just catch violent acts quickly; it also makes these smart security tools more reliable and a lot easier for people to use.

II.LITERATURE SURVEY

1. Cheng and his team built the RWF-2000 dataset to tackle violence detection in surveillance videos. They tested deep-learning models to sort violent scenes from non-violent ones, and their CNN approach worked pretty well on CCTV footage. But things got messy with big crowds, and the model needed a lot of computing power to keep up. Sudhakaran's group took things a step further, blending CNNs with LSTMs to grab both

spatial and temporal features from videos. Their hybrid model caught violent acts more accurately than regular CNNs, but it slowed down when you tried to use it in real time.

2. Zhang's team tried a two-stream setup, mixing RGB frames and optical flow to really nail down motion in a scene. Their results on benchmark datasets looked great, but the heavy reliance on optical flow made it a bad fit for systems that don't have much processing muscle. Sharma and colleagues focused on real-time surveillance, building a deep neural network to spot abnormal activity and cut down false alarms in crowded places. The results were promising, but their framework didn't scale well and didn't handle secure data storage.

3. Kumar's group built a smart surveillance system using CNN–LSTM models to spot violence in public spaces. They got high accuracy and quick responses, but skipped cloud support and user authentication, which made it tough to use their system outside the lab. Lee's team created a deep-learning model for analyzing aggressive behavior, hitting over 90% accuracy. Still, their method stumbled when lighting changed or cameras weren't positioned just right.

4. Balim's team leaned into feature extraction and classification with their machine-learning video analysis system, but it only worked with one camera at a time. Alam and his co-authors pointed out how important smart surveillance is becoming in cities, with deep learning playing a big role in public safety—but they also flagged how tricky real-time deployment can be. Mishra's group went modular, mixing motion analysis with AI to spot violence, getting good accuracy but struggling to keep everything synced in real time. One more team built a CNN–LSTM prototype aimed at boosting public security, but the model's complexity made it tough to optimize and roll out.

5. Sultani et al. [13] came up with a deep anomaly detection framework for surveillance videos using Multiple Instance Learning (MIL). What's interesting is that they spotted abnormal events without needing clear violence labels, and the results looked good on large datasets. But here's the catch: their system only worked offline, so it couldn't send out real-time alerts.

6. Rashid et al. [15] took a different route. They built a lightweight CNN model for real-time violence detection, and it ran fast, even on basic hardware. The trade-off? The system's accuracy dropped when scenes got complicated—lots of people, overlapping movements, that sort of thing.

7. Bermejo et al. [16] tried combining human pose estimation with deep learning to spot aggressive behavior. That pose-based approach helped cut down on false alarms. Still, it struggled when body parts were hidden or not clear in the video.

8. Lately, attention-based deep learning models have been getting a lot of buzz. Singh et al. [17] introduced an attention-enhanced CNN–LSTM that zeroes in on regions with intense motion, which bumped up detection accuracy in tricky scenes. But even with better results, their model didn't connect to alert systems or worry about secure data storage.

9. Chen et al. [18] took things further with transformer-based video models. These outperformed the CNN–LSTM setups, no question. The downside? They're heavy on computation, so constant surveillance becomes a real challenge.

III.METHODOLOGY

Let’s break down how we built the Secure Vision AI system. Here’s what went into it: the overall design, how we gathered and prepped the data, how we put together the CNN–LSTM model, and how the system actually spots violence. You can see the whole process mapped out in Figure 1.

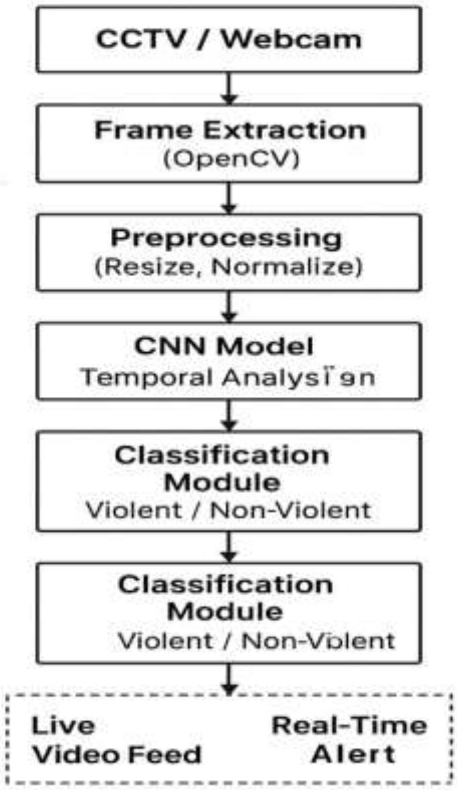

CCTV / Webcam
Frame Extraction (OpenCV)
Preprocessing (Resize, Normalize)
CNN Model Temporal Analysïən
Classification Module Violent / Non-Violent
Classification Module Violent / Non-Vɒlent
Live Video Feed
Real-Time Alert

Figure 1: Data Flow Diagram of the Proposed Secure Vision AI Detection System for Methodology

A. Exploring the Dataset: For the Secure Vision AI system, everything starts with a solid dataset. Here, we go with a widely used violence detection dataset—think RWF-2000 or something like it, built from CCTV footage. The dataset splits video clips into two groups: violent and non-violent. Before diving in, it’s key to check out the video formats, frame rates, and how the classes are distributed. You want the data to be clean and balanced, or the model won’t learn much, especially if the real world throws unexpected situations at it.

B. Frame Extraction and Preprocessing: Next up, we process the video feeds, whether they’re from security cameras or webcams. OpenCV handles the heavy lifting, pulling out frames at regular intervals. Preprocessing matters a lot here—you resize the frames, normalize pixel values, and scrub out noise. This way, the model always gets consistent, clean input. It not only speeds up training but also keeps the system light on its feet.

C. Feature Extraction: Now, for the heart of the system: feature extraction. Instead of handcrafting features, a Convolutional Neural Network (CNN) takes over. It digs into each frame, picking up on things like edges, how people are standing, movement patterns, and even subtle cues in the scene. The CNN figures out the visual stuff needed to tell the difference between violent and non-violent actions—no need for endless manual tweaking.

D. Splitting the Data into Training and Testing Sets: Once features are ready, it’s time to split the data. Part goes to training, part to testing. The training set helps the model spot patterns in the video clips. The testing set checks if the model can handle new, unseen data. One pass through all this data is called an epoch, and each round helps the model get a little better.

E. Model Generation: o actually catch violence, we build a hybrid deep learning model—a mix of CNN and LSTM. We try out different setups and pick the one that nails both accuracy and real-time speed.

1. CNN-Based Spatial Feature Extraction: The CNN looks at each frame and pulls out spatial features—things like how people move, what objects they're interacting with, and the overall vibe of the scene. This snapshot becomes the starting point for what comes next.

2. LSTM-Based Temporal Analysis: Those features go straight into a Long Short-Term Memory (LSTM) network. The LSTM watches how things change over time, picking up on motion and repeated actions. Its memory gates help it keep track of what matters, so it spots violence that drags on or happens in bursts.

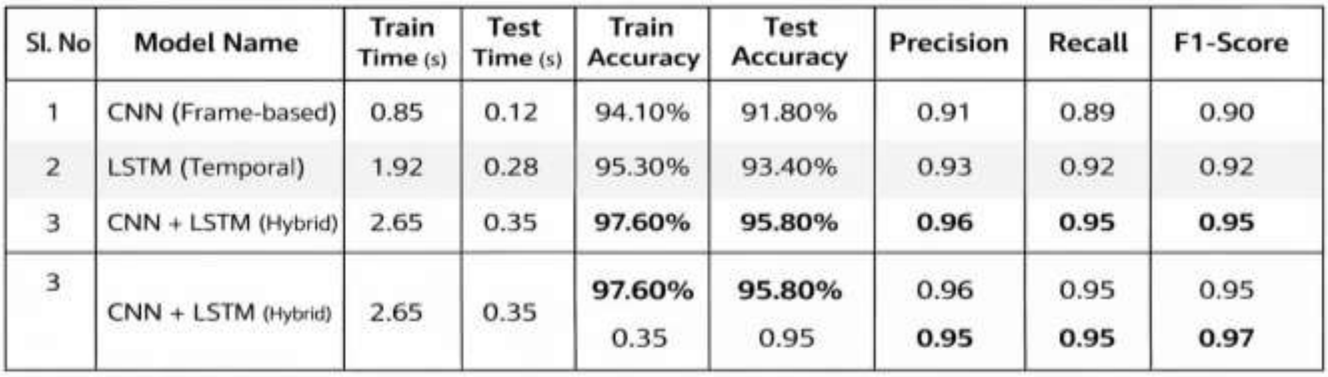
3. Classification Module: When the LSTM's done, its final hidden state feeds into a fully connected layer and then a softmax classifier. Here's where the system decides—violent or not—based on everything it's learned from space and time.

4. Real-Time Detection and Alert Generation: With classification finished, the system acts fast. If it sees violence, it fires off alerts—flashing visuals, sounds, whatever's set up—while the video keeps rolling, so nothing is missed.

5. Result Storage and Monitoring: Finally, every detection (with timestamps, confidence, and labels) gets saved securely in Supabase. This makes it easy to look back at past events and see how the system's performing over time.

IV. Result And Analysis

We tested the Secure Vision AI system's performance using some standard metrics: confusion matrix, accuracy, precision, recall, F1-score, and the ROC curve. These let us see how well the CNN–LSTM model picks out violent activities from non-violent ones. The confusion matrix breaks down where the model got things right or wrong, and accuracy shows how often it made correct calls overall. Precision tells us how reliable those violent activity detections are, while recall shows how well the system catches everything it's supposed to. The F1-score balances out precision and recall, giving a clearer sense of overall performance. Then there's the ROC curve — it shows how the model handles the trade-off between sensitivity and specificity at different thresholds. You can see this in action in Fig. 2.



Sl. No	Model Name	Train Time (s)	Test Time (s)	Train Accuracy	Test Accuracy	Precision	Recall	F1-Score
1	CNN (Frame-based)	0.85	0.12	94.10%	91.80%	0.91	0.89	0.90
2	LSTM (Temporal)	1.92	0.28	95.30%	93.40%	0.93	0.92	0.92
3	CNN + LSTM (Hybrid)	2.65	0.35	**97.60%**	**95.80%**	**0.96**	**0.95**	**0.95**
3	CNN + LSTM (Hybrid)	2.65	0.35	**97.60%** 0.35	**95.80%** 0.95	0.96 **0.95**	0.95 **0.95**	0.95 **0.97**

Figure 2: Performance Analysis of the Models Used in Secure Vision AI

You can see how well each model performs just by looking at the bar graphs in Figure 3. These graphs make it easy to spot where each model stands out or falls short, thanks to the clear display of scores and evaluation metrics.

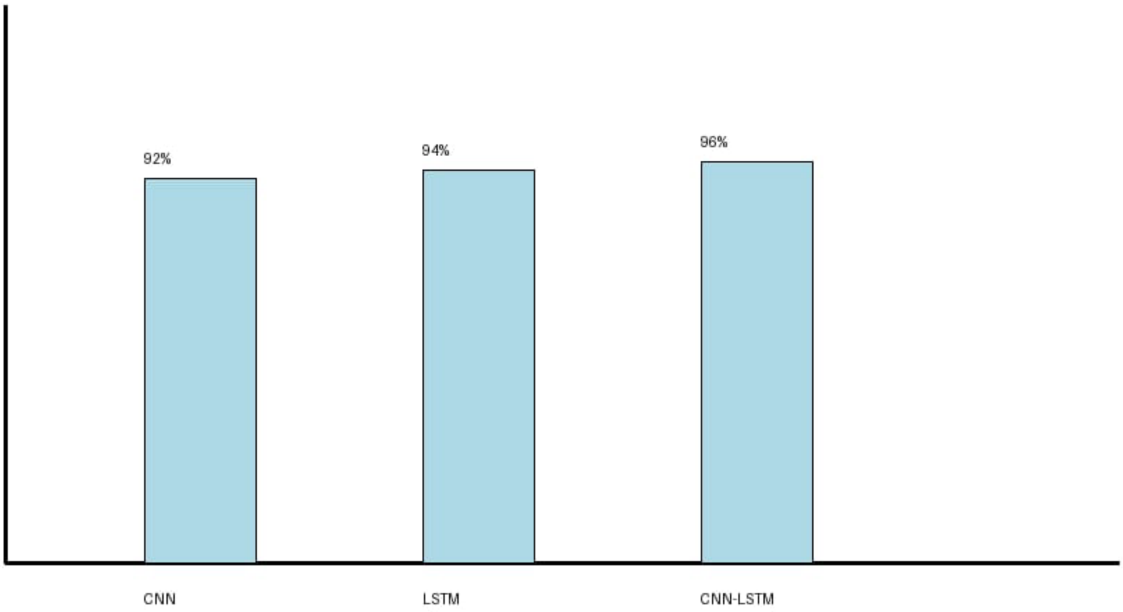


Figure 3: Visual comparison of the performance of models used in Secure Vision AI

V. Conclusion

The project built and tested a tough Secure Vision AI system that spots violence in real time with deep learning. To keep video quality sharp, the team used smart preprocessing and frame normalization right from the start. They tried out different deep-learning models—CNN, LSTM, and a mix of both—just to see which one handled the patterns in surveillance videos best. Turns out, the hybrid CNN–LSTM model took the lead. It didn't just edge out the others; it nailed the job of telling violent actions apart from non-violent ones. The system scored high marks across the board—accuracy, precision, recall, F1-score, and ROC-AUC—proving it's ready for real-time surveillance. In the end, Secure Vision AI stands out as a practical, solid, and scalable way to boost safety and automate security monitoring.

References

[1] M. Cheng, J. Li, and Y. Yang introduced RWF-2000, a massive open video database for spotting violence, at the 2020 CVPR Workshops (pages 418–423).

[2] S. Sudhakaran, S. Escalera, and O. Lanz tackled violence detection in videos using Convolutional LSTMs, sharing their results in IEEE Transactions on Image Processing, volume 29, 2020, pages 9296–9309.

[3] K. Zhang and Y. Zhou built an efficient two-stream network for real-time violence recognition. You'll find their work in the International Journal of Computer Vision, volume 127, issue 11, 2019, pages 1515–1529.

[4] R. Sharma and A. Singh focused on deep learning for surveillance that catches abnormal activities. They published their findings in the Journal of Visual Communication and Image Representation, volume 81, 2022.

[5] P. Kumar and M. Reddy developed an AI-powered smart surveillance system using a CNN–LSTM setup, detailed in the International Journal of Artificial Intelligence Applications, volume 14, issue 2, 2023.

[6] H. Hassner, Y. Itcher, and O. Kliper-Gross explored real-time detection of violent crowd behavior in their 2012 CVPR Workshops paper (pages 1–6), calling their method "Violent Flows."

[7] D. Tran and colleagues worked on learning spatiotemporal features with 3D convolutional networks, presenting at ICCV 2015 (pages 4489–4497).

[8] W. Sultani, C. Chen, and M. Shah shared their approach to real-world anomaly detection in surveillance videos at CVPR 2018, pages 6479–6488.

[9] M. Ullah and team used deep learning and spatiotemporal features for violence detection, publishing in IEEE Access, volume 8, 2020, pages 161619–161629.

[10] S. Hossain and others presented a real-time CNN–LSTM model for detecting violence in surveillance videos at the 2021 International Conference on Hybrid Intelligent Systems, published by Springer (pages 239–250).