Violence Detection System using CNN-LSTM Models

1. Introduction

Traditional CCTV monitoring is hampered by delayed responses, human fatigue, and scalability. This project develops an automated violence detection system using deep learning to identify violent activities in real-time video. By combining Convolutional Neural Networks (CNNs) for spatial feature extraction with Long Short-Term Memory (LSTM) networks for temporal pattern recognition, the system provides a 24/7 proactive security solution. It automatically detects violence, immediately notifies authorities with video evidence and location data, and maintains comprehensive incident logs, drastically reducing response times and providing actionable intelligence.

2. Project Objective and Scope

Objectives:

1. To create a predictive model using a hybrid CNN-LSTM architecture that can accurately classify violent behavior in video streams.
2. To improve the speed and objectivity of security monitoring and threat detection.
3. To provide immediate, evidence-backed alerts and comprehensive incident logging.
4. To establish a link between advanced deep learning and practical security applications.

Scope:

1. The project will focus on developing a software application for processing video streams.
2. Data will be processed in real-time from video input sources.
3. The model's performance will be evaluated using standard machine learning metrics.
4. The system will provide instant alerts and logging but will not replace human security authority or decision-making.

3. Methodology

The system processes video through an integrated five-stage pipeline:

1. **Video Input & Preprocessing:** Streams are decomposed into frames, resized to 224x224, and normalized.
2. **Spatial Feature Extraction:** A pre-trained MobileNetV2 CNN extracts key visual features from individual frames.

3. **Temporal Pattern Analysis:** An LSTM network analyzes sequences of CNN features to recognize motion patterns characteristic of violence.
4. **Violence Classification:** Predicts "Violent" or "Non-Violent" with a confidence score, using adaptive thresholding to minimize false positives.
5. **Automated Response**: Upon detection, the system simultaneously saves the video clip, dispatches an alert via Telegram with GPS and video, and logs the incident in a MySQL database.

4. Hardware & Software

1. **Software**: Python, TensorFlow/Keras, OpenCV, MySQL, Telegram Bot API, NumPy, Pandas.
2. **Hardware:** Multi-core CPU (Intel i5/i7, AMD Ryzen 5/7), 16GB+ RAM, NVIDIA GPU (CUDA-supported), SSD storage.

5. Limitations:

1. **Technical Constraints:** Performance depends on video quality, lighting, and camera angle. Requires significant computational resources and network connectivity.
2. **Model Performance:** Accuracy is limited by training data diversity. Potential for false positives from contact sports or staged violence.
3. **Privacy and Ethical Considerations:** Raises privacy concerns regarding video capture and storage. False positives could trigger unnecessary emergency responses.

6. Future Scope:

1. **Advanced Detection**: Integrate audio analysis, weapon detection, 3D pose estimation, and crowd behavior analysis.
2. **Technical Optimization:** Deploy on edge devices, use model compression, and develop cloud-edge hybrid systems.
3. **System Integration:** Connect with IoT ecosystems, emergency service APIs, and develop analytical dashboards.
4. **AI Improvements:** Implement explainable AI, continual learning systems, and enhance model robustness.

7. Conclusion:

This project presents a practical violence detection solution that integrates CNN-LSTM deep learning for proactive security monitoring. Its key achievement is transforming passive surveillance into an active system capable of automated detection, instant evidence-backed alerts, and comprehensive logging. Despite challenges, this framework represents a significant advancement in enhancing public safety through proactive, intelligent threat detection.

8. References:

1. Sudhakaran & Lanz (2017): Foundational paper on the CNN-LSTM model for violence detection.
2. Simonyan & Zisserman (2014): Introduced the classic "Two-Stream" architecture for action recognition.
3. Tran et al. (2015): Proposed the C3D model, pioneering the use of 3D CNNs for video.
4. Carreira & Zisserman (2017): Developed the influential I3D model and the large-scale Kinetics dataset.
5. Wang et al. (2018): Applied self-attention to video with "Non-local Neural Networks."
6. Cheng et al. (2020): Introduced the RWF-2000 dataset for real-world fight scenes.
7. Sultani et al. (2018): Presented the UCF-Crime dataset for anomaly detection in surveillance.
8. Bregu et al. (2021): Provided the UBI-Fights dataset, a modern resource for violence research.
9. Goodfellow et al. (2016): The fundamental textbook on *Deep Learning* theory.
10. Chollet (2021): The practical guide to implementation, *Deep Learning with Python*.