



INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

HOLISTIC THREAT PREDICTION IN PUBLIC SPACES USING MULTIMODEL DEEP LEARNING

Bhuvaneshwari G

Department of Computer Science and Business Systems, K.Ramakrishnan College of Engineering, Trichy, Tamil Nadu, India

Roshna parveen J

Department of Computer Science and Business Systems, K.Ramakrishnan College of Engineering, Trichy, Tamil Nadu, India

Swathi R

Department of Computer Science and Business Systems, K.Ramakrishnan College of Engineering, Trichy, Tamil Nadu, India

Dr. P. Shanmuga Priya, M.E., Ph.D.,

*Assistant Professor, Computer Science and Business Systems
K Ramakrishnan College of Engineering, Trichy Tamil Nadu, India*

Abstract: Traditional healthcare and supply chain systems operate independently and rely heavily on manual processes, resulting in poor coordination, limited visibility, delayed decision-making, and inefficient resource utilization. In the healthcare sector, this leads to problems such as appointment congestion, lack of remote consultation facilities, delayed access to medicines, and inefficient management of medical records and supplies. Existing systems also lack intelligent support for predicting demand, managing inventory, and monitoring service usage patterns.

To address these challenges, this project proposes an integrated digital platform that combines a telemedicine system with an AI-enabled digital supply chain framework. The telemedicine module allows patients to book appointments, consult doctors remotely, receive digital prescriptions, and access medical reports through an online platform. At the same time, artificial intelligence techniques are applied to analyze healthcare usage data, appointment trends, and medicine demand to support better planning and decision-making. AI-based analysis helps in predicting service demand, identifying resource shortages, and improving the availability of medical supplies and services.

Compared to the existing system, the proposed solution reduces manual workload, improves accessibility to healthcare services, enhances decision accuracy, and enables efficient coordination between healthcare delivery and supply chain operations. The project demonstrates how integrating telemedicine with AI-driven digital supply chain concepts can improve healthcare accessibility, operational efficiency, and overall system reliability.

Keywords- *Multimodal Deep Learning, YOLOv8, MobileNetV2, MFCC, Threat Detection, Public Safety*

I. INTRODUCTION

With the increasing population and rapid urbanization, public spaces such as malls, airports, railway stations, and stadiums are becoming more crowded and vulnerable to security threats. Traditional surveillance systems rely heavily on human monitoring and primarily focus on detecting visible threats like weapons. However, many dangerous situations such as stampedes, riots, and panic behavior occur without the presence of weapons and are difficult to detect using conventional systems.

Advancements in artificial intelligence and deep learning have enabled the development of intelligent surveillance systems capable of analyzing complex scenarios. Multimodal deep learning combines data from different sources such as video and audio to provide a more comprehensive understanding of the environment. This project introduces a holistic threat prediction system that integrates multiple technologies to improve public safety through real-time monitoring and analysis.

I.1. Abbreviations and Acronyms

- **AI - Artificial Intelligence**
- **DL - Deep Learning**
- **CNN - Convolutional Neural networks**
- **CCTV - Closed Circuit Television**
- **YOLO - You Only Look Once**
- **MFCC - Mel Frequency Cepstral Coefficients**

I. Problem Statement

The traditional healthcare system relies heavily on manual and fragmented processes for appointment scheduling, patient record management, consultation, and prescription handling, which often leads to long waiting times, data inconsistency, limited accessibility, and inefficient coordination between patients, doctors, and healthcare administrators. Patients face difficulties in finding suitable doctors, managing medical records, and accessing timely consultations, while doctors struggle with maintaining organized patient data and managing schedules effectively. The absence of a centralized digital platform further restricts remote healthcare services, real-time communication, and administrative monitoring. These challenges highlight the need for an integrated, secure, and scalable digital healthcare management system that streamlines healthcare operations, improves accessibility, and enhances the overall quality of medical services.

II. Background of the Project

Nowadays, public places like malls, railway stations, airports, and stadiums are becoming more crowded due to increasing population and urbanization. These crowded environments are more prone to safety issues such as weapon-based threats as well as non-weapon dangers like stampedes, riots, and panic situations. Traditional surveillance systems mainly depend on CCTV cameras and human monitoring, which are not fully effective in detecting complex situations like abnormal crowd behavior or sudden panic. With the development of Artificial Intelligence and Deep Learning, smarter systems can now analyze both video and audio data to detect threats automatically. This project is based on the need to improve public safety by developing a system that uses multimodal deep learning techniques to identify and predict different types of threats in real time more accurately and efficiently.

III. Objective

The main objective of this project is to develop a Holistic Threat Prediction System that can improve public safety by detecting both weapon-based and non-weapon-based threats in real time. The system aims to use multimodal deep learning techniques by combining video and audio analysis to identify weapons, abnormal activities like fights, crowd behavior, and panic sounds. It also focuses on classifying threats into different levels such as Normal, Suspicious, and High Risk to help authorities take quick action. Overall, the project aims to increase accuracy, reduce false alarms, and provide a reliable and efficient solution for monitoring crowded public spaces.

IV. Scope of Project

The scope of this project focuses on improving public safety by developing an intelligent system that can monitor crowded environments such as malls, airports, railway stations, and public events. The system is designed to detect both weapon-based and non-weapon threats using multimodal deep learning techniques by analyzing video and audio data in real time. It can identify weapons, abnormal activities like fights, crowd behavior, and panic sounds, and classify them into different risk levels for quick response. The project can be extended for use in smart city surveillance, large-scale event monitoring, and emergency management systems. In the future, the system can be enhanced with higher accuracy models, integration with IoT devices, and deployment on cloud platforms for wider scalability and real-time performance.

V. Proposed Systems

The proposed system is a multimodal deep learning framework that combines video and audio analysis for comprehensive threat detection. It uses YOLOv8 for real-time detection of weapons and crowd density from CCTV footage. MobileNetV2 is used to detect abnormal human activities such as fights by analyzing movement patterns in video frames. In addition to visual data, the system processes audio signals using MFCC (Mel-Frequency Cepstral Coefficients) to detect panic sounds such as screams, shouting, and distress signals. All these modules work in parallel, and their outputs are

combined using a multimodal fusion approach. The system includes a Risk Assessment Module that classifies situations into three categories: Normal, Suspicious, and High Risk. Based on this classification, alerts are generated for security personnel to take immediate action. The system is designed to work efficiently in real-time and perform reliably even in challenging conditions such as crowded environments, low lighting, and high noise levels.

VI. Architecture Diagram

The system architecture of the proposed Holistic Threat Prediction Framework is designed to analyze both video and audio data for effective real-time threat detection. The process begins with input collection from CCTV cameras and microphones installed in public spaces. The video data is processed using deep learning models where YOLOv8 is used to detect weapons and crowd density, and MobileNetV2 is used to identify abnormal activities such as fights and unusual human behavior.

At the same time, the audio data is analyzed using MFCC (Mel-Frequency Cepstral Coefficients) to extract sound features and detect panic-related sounds like screams, shouting, and distress signals. Both video and audio processing modules work simultaneously, and their outputs are combined in a multimodal fusion module. This integration helps in improving the accuracy and reliability of threat detection by considering multiple sources of information.

After fusion, the combined data is passed to the risk assessment module, which classifies the situation into three levels:

Normal, Suspicious, and High Risk. Based on this classification, alerts are generated and sent to security personnel for immediate action. The final output is displayed on a monitoring dashboard, enabling real-time tracking and quick decision-making to ensure public safety. The system architecture of the proposed Holistic Threat Prediction Framework is designed

to handle real-time monitoring by integrating both video and audio data from public spaces. The input layer consists of CCTV cameras and microphones that continuously capture visual and sound data. The video stream is processed using advanced deep learning models, where YOLOv8 is used for detecting weapons and estimating crowd density, while MobileNetV2 analyzes human movements to identify abnormal activities such as fights, sudden running, or unusual behavior. This ensures that both visible threats and behavioral patterns are effectively captured.

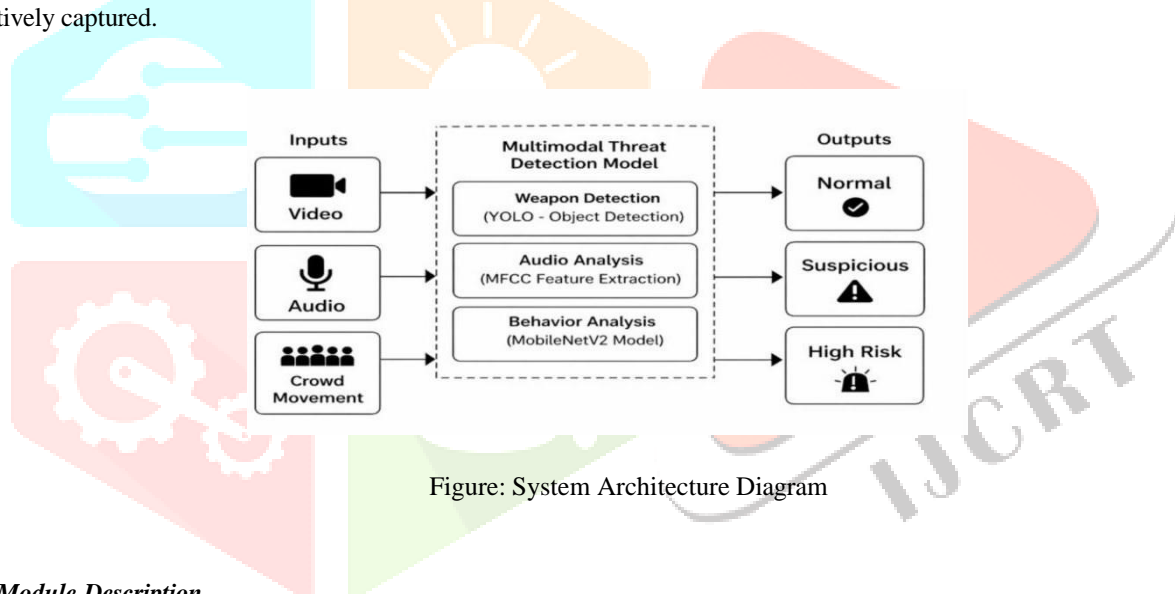


Figure: System Architecture Diagram

VIII Module Description

I.1. Crowd Detection Module:

The Crowd Detection Module uses YOLOv8 to detect and count people from CCTV footage to estimate crowd density. It also analyzes movement patterns like sudden gathering or rapid motion to identify risks such as overcrowding or panic. The results are sent to the risk assessment module for classification and alert generation.

I.2. Fight Detection Module:

The Detects violent activities like fights using MobileNetV2 by analyzing human movements and behavior in video frames, and sends results for risk classification.

I.3. Weapon Detection Module :

Uses YOLOv8 to detect weapons like guns and knives in real-time video and identifies visible threats for immediate action. It detects and localizes weapons within frames with high accuracy, even in crowded environments. This helps in quickly recognizing visible threats that can cause harm. The output is sent to the system for immediate risk evaluation and alert generation.

I.4. Audio Panic Detection Module :

Analyzes sounds using MFCC to detect panic signals like screams and shouting, helping identify emergencies not visible in video. The system provides clear visual outputs such as bounding boxes and risk levels. This helps in quick understanding and effective response to critical situations.

I.5. Multimodel Fusion Module :

Combines video and audio outputs to improve accuracy, reduce false alarms, and support better threat classification. It integrates information from crowd detection, fight detection, weapon detection, and audio analysis. By using multiple data sources, the system reduces false alarms and enhances reliability. The fused data is then used for final threat classification and decision-making.

I.6. Alert and Visualization Module :

Displays results on a dashboard and generates real-time alerts to notify security personnel for quick response. The system provides clear visual outputs such as bounding boxes and risk levels. This helps in quick understanding and effective response to critical situations.

IX Methodology / Working Process

The proposed system follows a structured approach by collecting video and audio data from CCTV cameras and microphones in public spaces. The video data is processed using YOLOv8 for weapon and crowd detection and MobileNetV2 for identifying abnormal activities like fights, while audio data is analyzed using MFCC to detect panic sounds. The outputs from these modules are combined using multimodal fusion to improve accuracy and reduce false alarms. Finally, the system classifies the situation into Normal, Suspicious, and High Risk levels and generates real-time alerts for quick response.

X Tools & Technologies

The proposed system uses various tools and technologies to ensure efficient threat detection. Programming is done using Python, and deep learning models such as YOLOv8 and MobileNetV2 are used for video analysis. MFCC (Mel-Frequency Cepstral Coefficients) is used for audio feature extraction and panic sound detection. Libraries like OpenCV are used for image and video processing, while frameworks such as TensorFlow or PyTorch are used for building and training models. The system can be implemented on standard computers and integrated with CCTV cameras and microphones for real-time data collection and processing.

XI Advantages

The proposed system offers several advantages by improving public safety through real-time threat detection. It can detect both visible threats like weapons and hidden threats such as abnormal crowd behavior and panic situations. The use of multimodal deep learning increases accuracy and reduces false alarms by combining video and audio analysis. It minimizes the need for continuous human monitoring and enables faster response through automatic alerts. Additionally, the system works effectively in challenging conditions and can be applied in various public places such as malls, airports, and railway stations.

XII Limitations

The proposed system has some limitations despite its advantages. It requires high computational power for real-time processing of video and audio data, which may increase cost. The accuracy of the system can be affected by poor lighting conditions, heavy crowd occlusion, and excessive background noise. It also depends on the quality of input from cameras and microphones. In some cases, false positives or missed detections may occur, especially in highly complex environments.

XIII Future Enhancement

The system can be further improved by using more advanced deep learning models to increase accuracy and efficiency. Integration with IoT devices and cloud computing can help in large-scale deployment and real-time monitoring across multiple locations. The system can also be enhanced by adding facial recognition and behavior prediction for better threat analysis. Improving data training with larger and diverse datasets can further reduce errors and improve performance.

XIV CONCLUSION

The proposed Holistic Threat Prediction Framework significantly improves public safety by integrating advanced technologies such as computer vision, audio analysis, and deep learning. Unlike traditional systems, it is capable of detecting both visible threats like weapons and hidden threats such as abnormal crowd behavior and panic situations. The system enhances accuracy, reduces human dependency, and enables faster response through real-time threat classification. Its ability to operate effectively in complex environments makes it suitable for deployment in various public spaces. Overall, this project provides an intelligent and scalable solution for modern surveillance systems and contributes to safer public environments.

XV ACKNOWLEDGMENT

We would like to express our sincere gratitude to K. Ramakrishnan College of Engineering, Samayapuram, Tiruchirappalli, for providing us with the opportunity, infrastructure, and academic support to carry out this research work on “HOLISTIC THREAT PREDICTION IN PUBLIC SPACES USING MULTIMODEL DEEP LEARNING”. The institution’s supportive environment and access to essential resources played a significant role in enabling us to successfully complete this study. We extend our heartfelt thanks to our faculty members and mentors for their continuous guidance, valuable suggestions, and constructive feedback throughout the development of this project. Their expertise, encouragement, and timely support greatly helped us refine our approach and improve the quality of our research work.

We are also thankful to our department for creating a motivating academic atmosphere that encouraged us to explore emerging technologies in artificial intelligence and multimedia analysis. Our sincere appreciation goes to the management and administrative staff for their cooperation during the course of this work. We also acknowledge the support of our peers and classmates for their insightful discussions and collaborative spirit. Finally, we express our deep gratitude to our friends and family members for their constant encouragement, patience, and motivation, which helped us successfully complete and present this research.

XVI REFERENCE

- [1] Akhtar, Naveed, and Ajmal Mian. (2021). *Threat Detection in Surveillance Videos Using Deep Learning: A Comprehensive Survey*. IEEE Transactions on Pattern Analysis and Machine Intelligence.
- [2] Bao Wenxia et al. (2021). *Fire and Gun Detection Based on Semantic Embeddings*, IEEE International Conference.
- [3] Chen, Yanfeng, et al. (2021). *Weapon Detection in Real-Time Surveillance Using YOLOv5 and Enhanced Feature Aggregation*. IEEE International Conference on Computer Vision Workshops (ICCVW).
- [4] Gade, Rikke, and Thomas B. Moeslund. (2018). *Crowd Monitoring and Analysis Using Computer Vision: A Review*. ACM Computing Surveys.
- [5] Jesus Ruiz-Santaquiteria et al. (2021). *Handgun Detection Using Combined Human Pose and Weapon Appearance*, IEEE International Conference.
- [6] Khan, Salman, et al. (2022). *Multimodal Fusion for Video-Based Threat Recognition in Public Environments*. IEEE Transactions on Information Forensics and Security.
- [7] Li, Wei, et al. (2021). *Multimodal Deep Neural Networks for Audio–Visual Event Recognition*. IEEE Transactions on Neural Networks and Learning Systems.
- [8] Muhammad Tahir Bhatti et al. (2021). *Weapon Detection in Real-Time CCTV Videos Using Deep Learning*, IEEE International Conference.