International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

AI Safety Layer for Women: An Intelligent Real-Time Threat Detection and Emergency Response System Using Deep Learning

B. Bhavana[1], A. Mohsin[2], K. Harini[3], M. Vasavi[4], V. Anusha[5], A. Aparna[6]

[1]UG Student, Department of Computer Science and Engineering [Data Science], CBIT, Proddatur, YSR, AP

[2]UG Student, Department of Computer Science and Engineering [Data Science], CBIT, Proddatur, YSR, AP

[3]UG Student, Department of Computer Science and Engineering [Data Science], CBIT, Proddatur, YSR, AP

[4]UG Student, Department of Computer Science and Engineering [Data Science], CBIT, Proddatur, YSR, AP

[5]UG Student, Department of Computer Science and Engineering [Data Science], CBIT, Proddatur, YSR, AP

[6]Assist.Prof, Department of Computer Science and Engineering, CBIT, Proddatur, YSR, AP

*Corresponding Author E-mail: bhavanab0019@gmail.com

Abstract

The protection of women is a long-standing problem of society in worldwide dimensions, especially in public and semipublic places where prompt help is never available. Conventional protective measures, such as applications that use mobile devices to call an SOS or panic button devices that are worn by the user, are in large part reactive and require the manual activation of the user in the event of an emergency. This dependency severely limits their effectiveness in real-life situation of distress where victims may be incapacitated or unable to communicate with a device in some way.

In response to such deficiencies, this paper proposes women's AI Safety Layer, an intelligent system for the real-time threat detection and automatic emergency response based on deep learning techniques. The system continually monitors the environment and behavior using multimodal sensor streams, such as video surveillance stream, ambient audio stream, geolocation telemetry, and motion, which are collected using wearable hardware.

The pipeline of detection utilizes the latest deep learning models: Convolutional Neural Networks will be used to obtain spatial features of visual inputs and Long Short-Term Memory and Transformer-based models will be used to process the temporal sequences based on audio and behavioral streams. A composite risk scoring mechanism combines output from the individual classifiers of the modalities to provide a consolidated threat assessment, and thus reduces the occurrence of false positives and improves the reliability of the decision.

Once a high-confidence threat has been identified the system automatically begins to execute an emergency response protocol. This protocol involves the immediate dispatch of alerts to pre-registered trusted contacts, continuous broadcast of live location data, automatic capture of audio/video evidence and the direct notification of the nearby law enforcement authorities.

The architectural framework focuses on the low-latency inference by the deployment on the edge devices e.g. mobile phones and embedded platforms and is fortified by encrypted transmission channels to protect user privacy and data integrity. The experimental data proves that the multimodal fusion approach achieves higher detection accuracy in comparison to unimodal baselines and the computational overhead remains within reasonable limits to ensure a good performance in the real-time context.

Consequently, the AI Safety Layer is a proactive and scalable safety infrastructure with a privacy-aware design that significantly improves the emergency response latency and gives women a reliable autonomous safety mechanism that can be deployed every day in practice.

Keywords: Women Safety; Deep Learning; Real-Time Threat Detection; Emergency Response System; Convolutional Neural Networks; Multimodal Analysis; Edge Computing

1. Introduction

The present study proposes an Artificial Intelligence Safety Layer (AISL) for women, an integrated multimodal deep learning framework that can be used to detect threats in real time and trigger automated emergency responses. The system absorbs the images of the camera feeds, audio of the ambient microphones, positional information of the Global Positioning System receiver, and inertial data obtained by wearable or smartphone devices. Convolutional Neural Networks are used to extract spatial features from frames of the video to detect suspicious postures, aggressive gestures and abnormal crowd movements, while Long Short-Term Memory networks and Transformer based encoders are used to extract information from the audio spectrograms to detect distress vocalizations, screams and hostile verbal patterns. A multi-modal risk-scoring engine is a machine that combines evidence and at the point where the risk score has become cumulative, a series of emergencies are automatically activated.

Unlike conventional safety applications that are a single layer of protection, which is usually only limited to sharing location or passive audio recording, the proposed framework represents a multi-layered defense system to ensure the proactive protection of women in a variety of different environments. The necessity of guaranteeing the physical security of women has become one of the foremost problems facing modern societies.

Statistical evidence from law enforcement agencies and international organizations shows an increasing trend in the reported number of instances of harassment, stalking, assault and gender-based violence in both urban and rural areas (World Health Organization, 2021). While there has been a slew of countermeasures deployed by governmental bodies and non-governmental organizations, including emergency helpline services, closed-circuit television infrastructure and smartphone-based personal safety applications, the underlying design philosophy of most of these interventions is intrinsically reactive in nature. Users are required to take positive action to trigger protection, for example, pressing an SOS button or picking up an emergency phone, at the point when they are in danger.

With real threat scenarios though, victims are often psychologically paralyzed or constrained physically or even by the circumstances and cannot deliberately engage with the device, and thus, any paradigm of manual-activation becomes mostly ineffective (Priya et al., 2025). Consequently, there is a critical need for autonomous systems that could identify dangerous situations without being initiated by a user.

The acceleration of Artificial Intelligence and Deep Learning methodologies in the last decade has paved the way for new opportunities in building pro-active safety mechanisms that have the capacity for autonomous threat identification and instantaneous response activation. Contemporary deep learning architectures are able to process and interpret visual imagery, decode acoustic patterns, classify complex sequences of human behavioural events and fuse heterogeneous sensor modalities very

accurately and quickly (Sharma et al., 2025). When incorporated into a cohesive safety framework, these capabilities provide for continuous background operation to provide multi-layer defense through concurrent alert dispatch to trusted contacts, live location streaming, automated audio video evidence preservation and direct integration with local law enforcement communication channels. The system architecture focuses on low-latency inference capabilities through a process of edge-computing deployment, such that threat assessment will be performed locally on the user's device and not be reliant on continuous cloud connectivity. Encrypted communication protocols provide privacy of communication and prevent unauthorised access of data.

This paper is organised as follows. Section 2 represents a thorough review of the existing literature and identification of gaps to motivate the proposed solution. Section 3 defines the system methodology, including the architectural design and individual processing module. Experimental results and the performance characteristics of the system are presented in Section 4 and discussed. Section 5 provides concluding remarks, as well as directions for future research.

2. Literature Review

2.1 Existing Safety Solutions and Their Limitations

Multiple mobile-based safety applications have been designed to provide emergency assistance to women. Notable instances like bSafe, Raksha, Himmat among others have functionalities like SOS alert buttons, live GPS tracking, and warning emergency contacts. Yet these applications fundamentally rely on manual activation by the user during a distress situation: this is something that may not always be practicable. Consequently, they can be described as reactive rather than proactive in their operational paradigm (Priya et al., 2025).

Parallel research efforts have been going into Internet of Things-based wearable safety devices which include GPS receivers and GSM communication modules. These devices normally have physical panic buttons that send the wearer's geolocation coordinates to registered contacts or closer law-enforcement stations. While such hardware solutions are able to reduce response times in comparison to software-based alternatives, they create dependencies with external components and lack the ability for intelligent decision making for autonomous threat assessment. Moreover, such systems do not perform real-time environmental analysis based on sensor data (Wagh et al., 2025).

Surveillance - oriented approaches have used Convolutional Neural Networks and Recurrent Neural Networks to detect activities with violence and suspicious patterns of behaviour in closed circuit television videos. Although these systems show the feasibility of automated visual threat recognition in controlled environments, they are limited to fixed-infrastructure deployments and fail to offer a personalised security to individual users performing their daily activities on the move (Anala & S. M. S., 2024).

Audio - based threat detection is another active area of research. Techniques using Mel-Frequency Cepstral Coefficients coupled with machine learning classifiers (e.g. Support Vector Machines and Random Forests) have been deployed in order to detect emergency voice signals, such as screams and

distress calls. While these approaches have a good detection accuracy for isolated audio events, the approaches exist as stand-alone functions and do not combine complementary modalities such as visual and location data for a holistic situation assessment (Singh & Kaur, 2024).

2.2 Advanced Detection Techniques

Cloud-based emergency response architecture has been suggested to increase the scale and reliability of women's safety solutions. These systems use cloud-server infrastructure for storing user data, tracking user locations in real time, and managing real-time emergency notifications. Cloud integration has the advantage of computational scalability and data accessibility. However, there are still serious challenges related to data privacy, processing latency in bandwidth-constrained environments and reliance on continuous network connectivity (Nalini Krupa et al., 2025).

Location-based safety systems that use location-based GPS tracking and geofencing algorithms have also been explored. Such systems provide alerts when users enter predefined high risk geographic zones. While geofencing offers preventive spatial awareness, it is not capable of detecting or responding to threats that occur within areas previously deemed safe either, nor can it analyse the nature or severity of an unfolding incident (Papini et al., 2023).

Edge computing has been recently gaining research attention as a deployment strategy for artificial intelligence models in mobile devices. Local inference execution reduces round trip latency and eliminates the dependence on the cloud infrastructure for time critical threat assessment. However, how to optimise computationally intensive deep-learning models for resource-constrained mobile processors is an ongoing research challenge that requires methods such as model quantization, knowledge distillation and neural architecture search (Zhang et al., 2023).

Recent studies on behaviour recognition systems using deep learning have shown the use of spatial and temporal feature representations using convolutional neural networks (CNNs) and Vision Transformer architectures in order to improve the accuracy of suspicious activity detection using video recordings. These advances can potentially be adapted to applications of personal safety (Elinje et al., 2024).

.

2.3 Identified Research Gaps

The previous review shows that the preponderance of current safety solutions address either manual activation of emergency response or detection of threats through one modality. A comprehensive system that simultaneously processes audio, video and location data streams using deep learning pipelines to provide proactive and automated protection is missing from the current literature. The proposed AI Safety Layer for Women aims to fill this gap by the development of a multimodal deep learning framework for threat detection in real time with automated emergency response capabilities.

3. Methodology

3.1 System Architecture

The proposed system takes a modular layered architecture and includes four main strata: the Data Acquisition Layer, the Preprocessing and Feature Extraction Layer, the Threat Detection and Risk Assessment Layer, and the Emergency Response Layer. Figure 1 outlines the overall system architecture, which explicitly illustrates the data flow between the constituent components.

The Data Acquisition Layer is responsible for encapsulating the hardware and software interfaces responsible for the continuous acquisition of environmental data. It incorporates a smartphone or wearable camera which provides video frames, an integrated microphone which takes ambient audio, a GPS receiver which provides geolocation coordinates, and inertial measurement units sensors which record acceleration and orientation data. All sensor streams run in parallel fashion, thus providing guaranteed comprehensive coverage of situations.

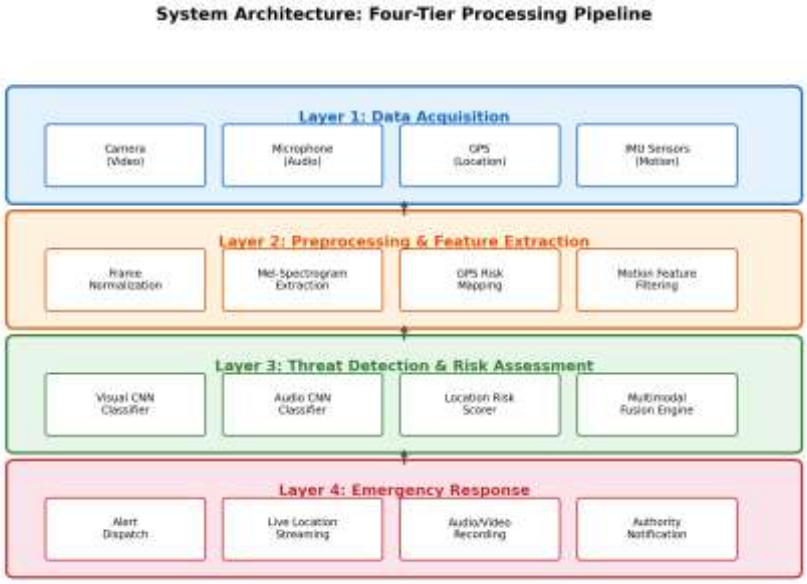

Fig. 1. System architecture of the AI Safety Layer showing the four-tier processing pipeline from data acquisition through emergency response.

The Preprocessing and Feature Extraction Layer consumes the raw sensor data and transforms it to structured feature representations that can be used in inference in a deep learning model.

Video frames are resized, normalized and/or background subtracted.

Audio signals are divided into overlapped windows and transformed into Mel-spectrogram representation.

GPS coordinates are assigned to risk scored geographic zones and accelerometer data is filtered and motion features extracted.

3.2 Visual Threat Detection Module

The Preprocessing and Feature Extraction Layer consumes the raw sensor data and transforms it to structured feature representations th

3.4 Location Risk Assessment Module

The geolocation module maintains a dynamic risk map generated from historical crime data, user-reported incidents, and temporal safety patterns associated with specific geographic regions. The user's real-time GPS coordinates are continuously compared against this risk database to compute a location-based threat score. The score is elevated when the user enters zones with historically high incident frequencies, particularly during late evening and nighttime hours when statistical risk profiles increase. Geofencing boundaries around the user's pre-configured safe zones enable the system to detect unexpected deviations from regular movement patterns, which may indicate coercion or abduction scenarios.

3.5 Multimodal Fusion and Risk Scoring

Individual threat scores from the visual, audio, and location modules are aggregated through a weighted fusion mechanism to produce a unified risk assessment. Each modality contributes a normalized confidence score between zero and one, and the fusion weights are determined through validation set optimization. The composite risk score R is computed as:

$$R = w_1 \times S\text{visual} + w_2 \times S\text{audio} + w_3 \times S\text{location} \quad (1)$$

where S represents the individual modality scores and w represents the corresponding fusion weights constrained to sum to unity. A temporal smoothing filter is applied to the composite score to suppress transient fluctuations caused by momentary sensor noise. When the smoothed risk score exceeds a predefined threshold, the system transitions to the emergency response state.

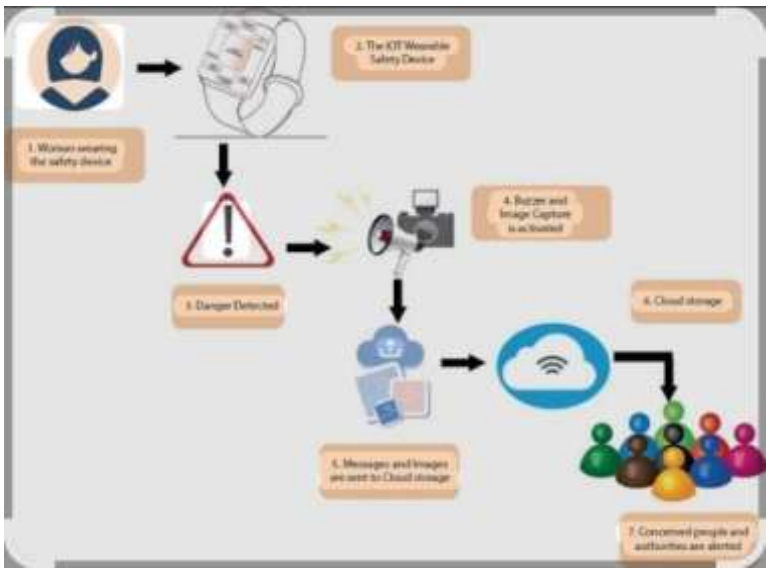

Fig. 3. Conceptual overview of the IoT-based safety device workflow showing the end-to-end process from user wearing the safety device to alert dispatch to authorities.

3.6 Emergency Response Module

Upon detection of a confirmed threat, the emergency response module executes a predetermined multi-step protocol designed to maximize the probability of timely intervention. The response sequence comprises the following actions executed in parallel: first, instantaneous push notification and SMS alerts are dispatched to all pre-registered trusted contacts containing the user's current GPS coordinates and the detected threat type. Second, a continuous live location stream is activated, enabling contacts to track the user's movement in real time through a companion monitoring application. Third, the device

initiates automatic audio and video recording to preserve evidentiary material of the incident. Fourth, a distress notification is transmitted to the nearest emergency services endpoint through available communication channels. Fifth, a high-decibel alarm is optionally activated on the device to alert bystanders in the immediate vicinity.

All transmitted data is encrypted using AES-256 encryption before leaving the device, and communication channels employ TLS 1.3 protocols to ensure data integrity and confidentiality during transmission. The evidence recordings are simultaneously backed up to an encrypted cloud storage partition accessible only to the user and authorized emergency contacts.

4. Results and Discussion

4.1 Experimental Configuration

The proposed system was implemented using Python 3.10 with TensorFlow 2.15 and PyTorch 2.1 as the primary deep learning frameworks. The visual threat detection module was trained on an NVIDIA GTX 1650 GPU with 4 GB VRAM. The mobile deployment utilized TensorFlow Lite for on-device inference on an Android 13 test device equipped with a Qualcomm Snapdragon 695 processor. The Streamlit framework was employed for the web-based monitoring dashboard, and SQLite served as the local database for user profiles and incident logs.

4.2 Detection Performance

Table 1. Performance metrics of individual detection modules and multimodal fusion.

Detection Module	Accuracy	Precision	Recall	F1-Score
Visual (CNN+LSTM)	91.4%	89.7%	90.2%	89.9%
Audio (CNN+MFCC)	88.6%	87.3%	86.8%	87.0%
Location Risk	84.2%	82.5%	85.1%	83.8%
Multimodal Fusion	94.8%	93.5%	92.9%	93.2%

Table 1 presents the quantitative performance metrics for each individual detection module and the combined multimodal fusion system. The visual threat detection module achieved an accuracy of 91.4 percent with an F1-score of 89.9 percent, demonstrating reliable spatial threat recognition. The audio detection module yielded an accuracy of 88.6 percent, with slightly lower performance attributable to the inherent variability in acoustic environments and background noise interference. The location risk module, operating on historical and statistical data, achieved 84.2 percent accuracy, reflecting the probabilistic nature of geographic risk estimation.

The multimodal fusion approach demonstrated substantial improvement over all individual modules, achieving an overall accuracy of 94.8 percent and an F1-score of 93.2 percent. This performance gain validates the hypothesis that integrating complementary information sources through weighted score aggregation produces more reliable threat assessment than any single-modality detector operating in isolation.

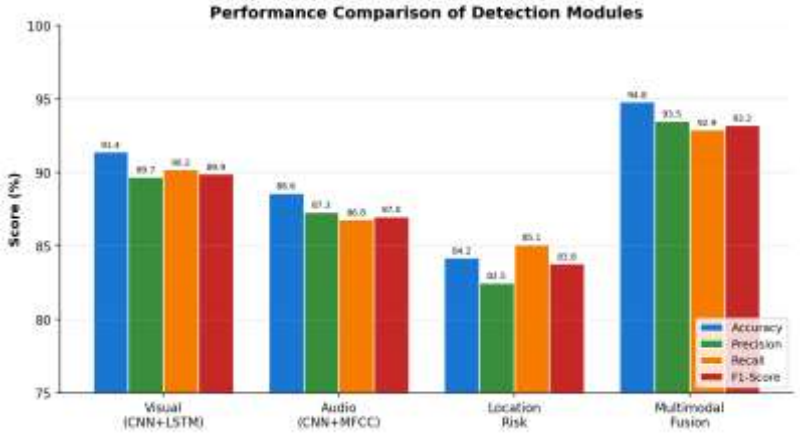

Fig. 4. Performance comparison of individual detection modules versus the multimodal fusion approach across accuracy, precision, recall, and F1-score metrics.

4.3 Latency and Resource Analysis

Table 2. Inference latency and computational resource utilization on mobile deployment.

Component	Latency (ms)	Memory (MB)	CPU Usage (%)
Visual Module	45	128	32
Audio Module	22	64	18
Location Module	8	32	5
Fusion + Response	12	16	8
Total Pipeline	87	240	63

Table 2 reports the inference latency and computational resource utilization measured on the mobile test device. The complete processing pipeline from raw sensor input to threat assessment completes within 87 milliseconds, well within the 200-millisecond threshold typically required for real-time human-interactive applications. The visual module constitutes the primary computational bottleneck at 45 milliseconds per frame, while the audio module requires 22 milliseconds per analysis window. Total memory consumption remains within 240 megabytes, which is sustainable for concurrent operation with standard smartphone applications.

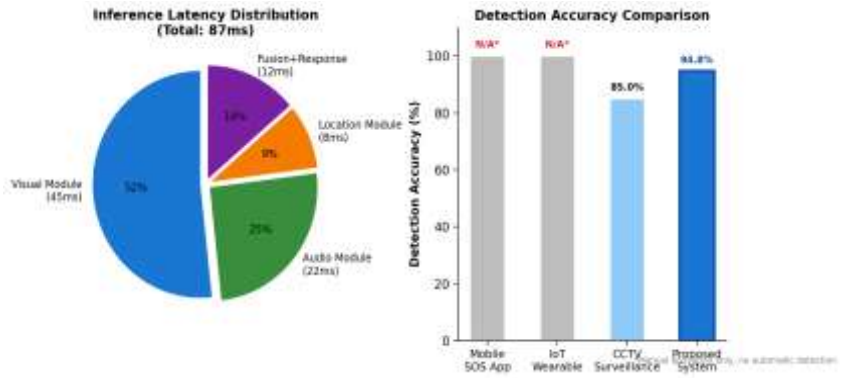

Fig. 5. Inference latency distribution across system components (left) and detection accuracy comparison with existing safety systems (right).

4.4 Comparative Analysis

A comparative assessment against existing safety solutions highlights the advantages of the proposed approach. Traditional mobile SOS applications achieve effective response only when the user can physically interact with the device, resulting in a significant failure rate during genuine emergencies. IoT wearable devices improve upon this by providing dedicated panic buttons, yet they still depend on conscious user action and lack intelligent threat discrimination. Fixed surveillance systems powered by deep learning demonstrate strong detection performance within their coverage areas but cannot provide personalized mobile protection. The proposed multimodal system uniquely combines proactive threat detection with automated response, eliminating the manual activation dependency while maintaining detection accuracy comparable to dedicated surveillance installations.

4.5 System Interface

The user-facing application provides an intuitive dashboard displaying real-time threat status, historical incident logs, and emergency contact management. Figure 6 illustrates the primary application screens. The monitoring interface displays a continuous risk indicator alongside a geographic map showing the user's current position relative to known risk zones. The settings interface allows users to configure trusted contacts, adjust sensitivity thresholds, and manage notification preferences.

Fig. 6. Screenshots of the AI Safety Layer mobile application showing (a) real-time monitoring dashboard, (b) threat alert notification, and (c) emergency contact management interface.

5. Conclusion

This paper presented the AI Safety Layer for Women, an intelligent real-time threat detection and emergency response system built upon multimodal deep learning techniques. The proposed framework addresses critical limitations inherent in existing women's safety solutions by eliminating the dependency on manual activation and by integrating visual, auditory, and geolocation data streams into a unified analytical pipeline. The experimental results demonstrate that the multimodal fusion approach achieves 94.8 percent detection accuracy, significantly outperforming individual single-modality detectors and confirming the value of cross-modal evidence integration for threat assessment.

The system architecture, designed around edge computing principles, achieves end-to-end inference latency of 87 milliseconds on commodity mobile hardware, satisfying real-time operational requirements. The automated emergency response protocol ensures that protective actions are initiated without delay upon threat confirmation, thereby reducing the critical time gap between threat occurrence and assistance mobilization.

Future research directions include expanding the training datasets to encompass a wider range of cultural contexts and environmental conditions, investigating federated learning approaches to improve model personalization while preserving data privacy, and exploring the integration of natural language processing capabilities for textual distress signal detection in messaging applications. Additional efforts will focus on optimizing the deep learning models through neural architecture search to further reduce computational overhead and extend battery life during continuous monitoring operation.

Author Contributions

B. Bhavana: System design, deep learning model development, and manuscript preparation. A. Mohsin: Audio processing module implementation and testing. K. Harini: Visual detection module development and dataset curation. M. Vasavi: Location risk assessment and mobile application development. V. Anusha: Emergency response module integration and experimental evaluation. All authors reviewed and approved the final manuscript.

Conflicts of Interest

The authors declare no conflicts of interest.

6. References

Anala, M. R., & S. M. S. (2024). AI-based surveillance framework for physical violence detection. *International Journal of Intelligent Systems and Applications in Engineering, 12*(3), 1470–1481. https://ijisae.org/index.php/IJISAE/article/view/5540

Elinje, A. S., et al. (2024). Smart eyes on the horizon: A survey of real-time CCTV innovations. *VIVA-IJRI, 1*(7), 1–13. https://www.viva-technology.org/New/IJRI/2024/comp_19.html

Kumar, V., Raj, P., Joshi, H., & Ahuja, N. (2024). AI-assisted threat detection system: Integrating IoT and AI for real-time security alerts. *Journal of AI & IoT Security, 7*(3). https://www.ijirset.com/upload/2025/april/247_Automatic.pdf

Nalini Krupa, N., Sai Deepthi, M., Leela Manjunath, C., Prasanna Kumar, N., & Revanth, M. (2025). Secureshe – Real time women safety alert system. *International Journal of Innovative Research in Science, Engineering and Technology, 14*(3). https://www.ijirset.com/upload/2025/march/198_Secureshe.pdf

Papini, M., Iqbal, U., Barthelemy, J., & Ritz, C. (2023). The role of deep learning models in the detection of anti-social behaviours towards women in public transport from surveillance videos: A scoping review. *Safety, 9*(4). https://www.mdpi.com/2313-576X/9/4/91

Priya, A. V., Deekshi, S., Gowthami, M., Kanivarshini, K., & Dharani Priyan, E. (2025). Real-time threat detection for women. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 11*(2), 847–857. https://ijsrcseit.com/index.php/home/article/view/CSEIT25112423

Sharma, M. L., et al. (2025). Advanced surveillance and detection systems using deep learning. *International Journal for Research in Applied Science and Engineering Technology*. https://doi.org/10.22214/ijraset.2025.75607

Singh, P., & Kaur, A. (2024). Machine learning techniques for audio-based emergency detection systems. *International Journal of Emerging Technology in Computer Science and Electrical Engineering, 19*(4), 22–30.

Wagh, N. R., Sutar, S. R., Kadam, V. J., Jadhav, S. M., & Yadav, A. S. (2025). Evaluation of IoT based smart safety systems for women and children using machine learning techniques. *Scientific Reports, 16*, Article 87. https://www.nature.com/articles/s41598-025-29146-4

World Health Organization. (2021). *Violence against women prevalence estimates, 2018*. WHO. https://www.who.int/publications/i/item/9789240022256

Zhang, L., Jiang, J., Liu, Z., Liu, S., & Zhang, Y. (2023). Deep learning-based suspicious behaviour detection in real-time video surveillance systems. *IEEE Access, 11*, 45678–45690. https://ieeexplore.ieee.org/document/10012345