G-Net: A Deep Learning-Based Real-Time Gesture-Driven Alert System for Enhancing Women's Safety

Bir Fateh Singh, Ashmit Sheoran, and Ansh Agrawal

Bennett University

Abstract. With the increasing demand for proactive public safety mechanisms, particularly for vulnerable groups such as women, there is a pressing need to enhance real-time surveillance systems. Existing solutions often depend on manual activation or post-incident analysis, limiting their effectiveness in critical situations. In this paper, we present *G-Net*, an AI-integrated surveillance framework capable of autonomously detecting distress gestures and issuing emergency alerts in real-time. Leveraging deep learning models like MobileNetV2 and YOLOv5, G-Net performs gesture recognition, facial detection, gender classification, and evidence collection. The system is trained using diverse datasets—including WIDER FACE, Adience, and custom surveillance data—to ensure robustness under varied environmental conditions. Comprehensive evaluation demonstrates high accuracy (over 90%) and latency under 1.5 seconds, suitable for real-time deployment. Compared to recent advancements in smart surveillance, such as the approach in [9], G-Net introduces context-aware automation with minimal human intervention. This framework offers scalable and efficient public safety infrastructure adaptable to smart cities and campus environments.

Keywords: Gesture Recognition · Facial Detection · Deep Learning · Women Safety · Real-Time Surveillance · Public Security

1 Introduction

The increasing frequency of crimes against women in public spaces has necessitated the development of intelligent surveillance systems capable of autonomous intervention. Traditional mechanisms, including CCTV setups and SOS mobile applications, rely heavily on human input or retrospective video review, rendering them ineffective in high-risk or time-critical scenarios. Addressing these shortcomings, we propose *G-Net*, a deep learning-driven real-time surveillance system that proactively recognizes distress through gestures, detects nearby individuals, and alerts authorities without requiring user interaction. With context-aware automation, G-Net significantly reduces response time while improving reliability across diverse public settings.

2 Related Work

Research in Human-Computer Interaction (HCI) has established the viability of gesture-based control systems using deep learning models [1, 2]. However, these approaches often focus on consumer electronics rather than public safety. Surveillance systems enhanced by deep learning have gained traction [3], yet they commonly rely on motion detection without understanding contextual cues. Mobile-based emergency apps, such as bSafe and Shake2Safety, require active user participation, limiting their utility during actual emergencies [4]. Facial recognition methods like FaceNet [5] and YOLOv5 [6] demonstrate impressive accuracy but are seldom integrated with gesture-triggered alerts. A recent system, SafeSight [9], incorporated spatiotemporal features for threat detection, but lacked real-time gesture-reactive modules. G-Net bridges this gap by integrating real-time gesture recognition with face analytics and immediate alert dispatching.

3 Datasets

To ensure high model accuracy across different modules, G-Net utilizes a comprehensive combination of datasets. For gesture recognition, public datasets such as Kaggle's Hand Gesture Recognition—with over 20,000 labeled images under varied lighting and orientations—were employed alongside a custom gesture dataset. WIDER FACE [7] provided robust facial detection training, with its high variability in facial angles and occlusions. Gender classification relied on the Adience dataset [8], containing annotated images for age and gender across unconstrained scenarios. Additionally, a custom surveillance dataset was developed in simulated environments like parking lots and metro stations to fine-tune the models for real-world applicability. All datasets were preprocessed using normalization, data augmentation, and resizing to enhance model generalization.

4 Methodology

G-Net is designed as a robust and modular real-time surveillance architecture that seamlessly integrates multiple deep learning components to detect potential threats and initiate rapid alerts. The system operates through a carefully orchestrated four-stage pipeline, with each module enhancing the accuracy and responsiveness of the overall framework. An illustration of the system workflow is conceptually depicted.

- **Stage 1: Gesture Recognition** The initial module is dedicated to identifying emergency gestures commonly associated with distress, such as an open palm or a two-handed wave. This task is performed using a MobileNetV2-based Convolutional Neural Network (CNN), chosen for its optimal balance between computational efficiency and classification accuracy. To improve the model's generalizability, the training dataset comprises both publicly available

and custom-collected gesture images. A comprehensive set of augmentation techniques—including rotation, flipping, brightness variation, and Gaussian noise—ensures resilience to real-world environmental conditions.

- **Stage 2: Facial Detection and Enhancement** Upon successful detection of a distress gesture, G-Net transitions to its facial detection module. YOLOv5, a real-time object detection model, is employed to locate all visible faces within the video frame. Detected faces are then cropped and processed using enhancement techniques such as Histogram Equalization and Contrast Limited Adaptive Histogram Equalization (CLAHE) to maintain image clarity under suboptimal lighting conditions. Each processed face is time-stamped and embedded with contextual metadata for later use in identity analysis and evidence generation.

- **Stage 3: Demographic and Identity Classification** The third stage involves demographic profiling and identity recognition. Cropped facial regions are input into a shallow CNN model trained on the Adience dataset to predict gender. For identity verification, G-Net leverages FaceNet or Dlib to generate facial embeddings, which are then compared against a reference database of known individuals. This module enhances situational interpretation by distinguishing between likely victims, suspects, and uninvolved individuals in the scene.

- **Stage 4: Alert Generation and Dispatch** Following successful classification, G-Net compiles all relevant data into a structured alert packet. This packet includes facial snapshots, gesture classification results, timestamps, inferred gender, and geolocation data if available. The information is securely transmitted through a POST request to a predefined emergency response endpoint. The recipient may be campus security, law enforcement, or a public safety network. The total execution time from gesture recognition to alert dispatch is consistently maintained under 1.5 seconds, ensuring the system's responsiveness in real-time scenarios.

This modular and scalable design ensures G-Net is not only effective but also adaptable for integration with future enhancements such as multilingual audio distress detection, facial emotion recognition, or lightweight deployment on edge computing devices.

5 Results and Analysis

The G-Net framework was tested across multiple controlled and semi-structured public environments. The following table presents the average performance metrics:

Table 1. Performance Metrics of G-Net Modules

Module	**Accuracy**	**Latency (ms)**
Gesture Recognition	94.2%	200
Face Detection	89.1%	180
Gender Classification	91.3%	150
Alert Dispatch	-	<1000

Across all components, the system demonstrated high reliability. The end-to-end alert pipeline—from gesture recognition to emergency alert transmission—operated within 1.5 seconds. Real-world scenarios involving 5–6 individuals in frame maintained over 93% detection success, even with partial occlusions or low lighting.

6 Conclusion and Future Work

G-Net addresses a critical gap in intelligent public safety frameworks by automating the detection and response process through deep learning techniques. Future enhancements include:

- Multilingual voice-based emergency command recognition
- Integration with low-cost edge devices like Jetson Nano or Coral TPU
- Privacy preservation via on-device inference and federated learning
- Deployment in large-scale public venues for further stress testing

By combining gesture interpretation with intelligent alerting and minimal user interaction, G-Net sets a foundation for future-ready safety infrastructure in smart cities and institutional settings.

References

1. P. Molchanov et al., "Hand gesture recognition with 3D convolutional neural networks," in *CVPR*, 2015.
2. S. Mitra and T. Acharya, "Gesture recognition: A survey," *IEEE Transactions on Systems, Man, and Cybernetics*, 2007.
3. W. Wang et al., "Deep learning-based intelligent surveillance system: A survey," *IEEE Access*, 2019.
4. K. Patel and R. Patel, "Emergency Response via Mobile Apps: A Comparative Study," *IJCSIT*, 2020.
5. F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," in *CVPR*, 2015.
6. G. Jocher, "YOLOv5," 2020. [Online]. Available: https://github.com/ultralytics/yolov5
7. S. Yang et al., "WIDER FACE: A face detection benchmark," in *CVPR*, 2016.
8. E. Eidinger et al., "Age and Gender Estimation of Unfiltered Faces," *IEEE TIFS*, 2014.
9. R. Gupta et al., "Context-Aware Real-Time Surveillance System for Public Spaces Using Deep Learning," in *IEEE Internet of Things Journal*, vol. 10, no. 3, pp. 2101–2113, 2023.