International Journal on Applications in Engineering and Technology
Volume 12 : Issue 4 : April 2026, pp 1 – 4 www.aetsjournal.com ISSN (Online) : 2455 - 0523

GestureGuard: AI-Powered Recognition of Emergency Gestures with Integrated Alert Mechanism for Security Enforcement

Mr. K. Ravikumar , Mr. Gowthamraja R, Mr. Hariprasath G, Mr. Indhuprakash P, Mr. Jagadeep K C

***Abstract*—**GestureGuard is an intelligent AI-powered surveillance solution designed to detect emergency hand gestures from CCTV feeds in real time. The system addresses the critical limitation that vulnerable individuals—women, teenagers, elderly, and isolated persons—may be unable to vocally call for help during emergencies. Traditional CCTV systems rely on manual observation, which leads to fatigue, missed incidents, and delayed action. The proposed system integrates YOLOv8 for fast hand detection and Spatio-Temporal Graph Convolutional Network (ST-GCN) to classify gestures based on motion patterns across sequential video frames. When a dangerous gesture is recognized, the system automatically triggers alerts with geolocation details to relevant authorities via SMS or email. The solution operates continuously with 24/7 monitoring, reducing reliance on manual surveillance personnel and ensuring faster emergency response. Experimental validation demonstrates effective gesture classification under diverse environmental and lighting conditions, with high accuracy in distinguishing emergency gestures from normal activity.

***Keywords*—** Emergency Gesture Recognition, YOLOv8, ST-GCN, CCTV Surveillance, Real-Time Alert System, Computer Vision, Keypoint Extraction, Hand Detection..

I. INTRODUCTION

Public safety surveillance has become a growing global concern, particularly for individuals who face threats in isolated or crowded environments. Women, teenagers, elderly people, and individuals with disabilities are among the most vulnerable demographics who may be unable to call for help verbally during emergencies. In such situations, silent gesture-based communication can serve as a critical lifeline to emergency services. However, existing surveillance infrastructure predominantly relies on passive, manual monitoring, which is inherently limited by human attentiveness, fatigue, and cognitive overload.

The rapid advancement of computer vision and deep learning has created an opportunity to automate emergency detection through intelligent systems that analyze video feeds in real time. Gesture recognition, a subfield of human-computer interaction, enables machines to understand and interpret human body language and hand movements. When integrated with modern surveillance systems, gesture recognition can bridge the gap between a distress signal and an effective emergency response.

GestureGuard presents an AI-powered surveillance framework that detects predefined emergency hand gestures from CCTV video streams. The system uses YOLOv8 for accurate and real-time hand localization, followed by keypoint extraction and classification through Spatio-Temporal Graph Convolutional Networks (ST-GCN). The combination of spatial keypoint relationships and temporal motion patterns enables the system to robustly distinguish emergency gestures from normal activities. Upon detection, automated alerts containing geolocation data are dispatched to law enforcement or emergency contacts via SMS and email.

The proposed system is deployable in homes, schools, offices, public transportation stations, and hospitals, providing a scalable and silent emergency communication channel that does not require the victim to speak, shout, or physically reach a device.

II. RELATED WORK

Early investigations into gesture recognition established the significance of combining multiple sensing modalities to accurately detect hand movements [1]. Research on vision-based hand gesture systems demonstrated that deep learning approaches, particularly convolutional neural networks, significantly outperform traditional feature engineering methods in terms of recognition accuracy and generalizability across different users and environments [2].

The adoption of Graph Convolutional Networks (GCNs) for skeleton-based action recognition was a pivotal development in the field. ST-GCN, introduced by Yan et al., modeled the spatial relationships between body joints as graph edges while learning temporal dynamics across frames, enabling nuanced understanding of complex motion patterns [3]. Subsequent extensions introduced adaptive graph structures and attention

Mr. K. Ravikumar AP/CSE, Department of Computer Science and Engineering, Knowledge Institute of Technology, Salem, Tamilnadu, India. (Email: krkcse@kiot.ac.in)

Mr. Gowthamraja R, Department of Computer Science and Engineering, Knowledge Institute of Technology, Salem, Tamilnadu, India(Email: gowtham2004raja@gmail.com)

Mr. Hariprasath G, Department of Computer Science and Engineering, Knowledge Institute of Technology, Salem, Tamilnadu, India. (Email: harihp932@gmail.com)

Mr. Indhuprakash P, Department of Computer Science and Engineering, Knowledge Institute of Technology, Salem, Tamilnadu, India. (Email: indhuprakashpalanivel2005@gmail.com)

Mr. Jagadeep K C, Department of Computer Science and Engineering, Knowledge Institute of Technology, Salem, Tamilnadu, India. (Email : jagadeep2004@gmail.com)

mechanisms to improve feature discrimination for fine-grained gestures [4].

In parallel, the evolution of object detection algorithms, culminating in the YOLO family of models, enabled real-time detection with high accuracy. YOLOv8 represents the state-of-the-art in single-stage detection, providing an optimal balance between detection speed and precision that is critical for real-time surveillance applications [5]. MediaPipe and OpenPose have been used extensively for hand and body keypoint extraction, and their integration with deep learning classifiers has shown promising results for gesture-based human-machine interaction [6].

Research on silent emergency systems specifically targeting vulnerable populations remains limited. Existing work has explored distress call detection and fall detection, but gesture-based silent emergency signaling integrated with CCTV surveillance and automated alert mechanisms has received comparatively less attention [7]. GestureGuard addresses this gap by providing an end-to-end pipeline from gesture detection to alert dissemination.

III. METHODOLOGY

The GestureGuard methodology adopts an end-to-end integrated framework combining real-time video processing, deep learning-based hand detection, keypoint extraction, spatiotemporal gesture classification, and multi-channel alert generation. The pipeline is designed for continuous operation with low-latency response and high classification accuracy across diverse real-world environments.

A. System Architecture and Hardware Requirements

The system architecture is organized into three processing tiers. The acquisition layer captures continuous video from IP or standard CCTV cameras. The processing layer, hosted on a server with an Intel i5/Ryzen 5 processor, 16 GB RAM, and 512 GB SSD running Python 3.9+, performs frame analysis, hand detection, and gesture classification. The communication layer dispatches alerts and logs incidents to a web-based administrative dashboard. The software stack includes OpenCV for frame handling, MediaPipe/OpenPose for landmark extraction, and TensorFlow/PyTorch for model inference.

B. Hand Detection Using YOLOv8

YOLOv8 is employed as the primary hand detection module. Being a single-stage detector, YOLOv8 processes entire frames in a single forward pass, making it suitable for real-time applications. Each video frame captured from the CCTV feed is preprocessed through resizing and normalization before being passed to the YOLOv8 model. Bounding box predictions localizing each detected hand in the frame are produced with associated confidence scores. Only detections exceeding a confidence threshold are forwarded to the keypoint extraction stage, minimizing spurious detections.

C. Keypoint Extraction and Preprocessing

For each detected hand region, a keypoint extraction module based on MediaPipe or OpenPose identifies 21 anatomical landmarks corresponding to finger joints and palm. The 2D coordinates of these landmarks form the spatial graph nodes. Raw keypoint data is preprocessed through normalization relative to the palm center, scale invariance adjustment, and noise reduction via a moving average filter. This ensures consistent input representation regardless of hand size or distance from the camera. The sequence of keypoint frames over a temporal window constitutes the input to the ST-GCN classifier.

D. Gesture Classification Using ST-GCN

The Spatio-Temporal Graph Convolutional Network receives a sequence of keypoint graphs representing hand pose over multiple consecutive frames. In the spatial domain, graph edges model anatomical connections between finger joints, while ST-GCN layers learn the co-movement patterns across connected joints. In the temporal domain, the network captures motion trajectories and velocity patterns across the frame sequence window. The output layer produces a probability distribution over gesture classes including defined emergency gestures (such as the international distress signal and variations) and normal/non-emergency hand activities. A gesture is flagged as an emergency when the predicted probability for an emergency class exceeds a predetermined classification threshold validated on the training corpus.

E. Dataset Collection and Model Training

Gesture datasets are collected under varied lighting conditions, backgrounds, and camera angles to ensure model generalizability. Data augmentation techniques including rotation, scaling, horizontal flipping, and temporal jittering are applied during training to improve robustness. YOLOv8 is fine-tuned on a custom hand detection dataset, while ST-GCN is trained on gesture sequence data with annotated emergency and non-emergency labels. Validation is performed on held-out test sets to prevent overfitting and evaluate real-world performance.

F. Alert Generation and Notification

Upon confirmed detection of an emergency gesture, the alert module is activated. The system retrieves the geolocation associated with the CCTV camera, timestamps the event, and captures a reference frame. This information is formatted into an alert payload dispatched via SMS and email to pre-configured authority contacts. All events are logged in a MySQL database accessible through the web dashboard for auditing and post-incident review.

IV. SYSTEM MODULES

The GestureGuard system is organized into five functional modules that operate in a coordinated pipeline:

•Web Application Dashboard: Provides a real-time monitoring interface for administrators to observe live feeds, review alerts, and manage system configuration.

•GestureNet Model Training: A dedicated module for dataset management, model training, hyperparameter tuning,

and performance evaluation of the YOLOv8 and ST-GCN models.

•Live Gesture Detection: The core inference pipeline that processes incoming camera frames, detects hands, extracts keypoints, and classifies gestures in real time.

•Alert Generation and Notification: Manages the dispatching of SMS and email alerts to authorities upon confirmed emergency gesture detection, including geolocation embedding.

•Admin/User Access Control: Authentication and role-based access control ensuring that only authorized users can access the dashboard, modify configurations, or review incident logs.

V.RESULTS AND PERFORMANCE EVALUATION

The GestureGuard system is evaluated through controlled experiments covering varied gesture conditions, lighting environments, and camera distances. The performance metrics assessed include hand detection accuracy, gesture classification accuracy, false positive rate, and end-to-end alert latency.

A. Hand Detection Performance

YOLOv8 demonstrates high hand detection performance across diverse conditions. Table 1 summarizes detection accuracy under different test scenarios.

Table 1: YOLOv8 Hand Detection Performance

Test Condition	Total Frames	Correct Detections	Detection Accuracy (%)	Avg. Inference Time (ms)
Normal Lighting	500	492	98.4	23
Low Light	500	471	94.2	25
Partial Occlusion	500	463	92.6	23
High Background Clutter	500	477	95.4	24
Multiple Hands	500	481	96.2	26

B. Gesture Classification Accuracy

The ST-GCN classifier is evaluated on a test set of gesture sequences. Table 2 presents classification accuracy across emergency and non-emergency gesture classes.

Table 2: ST-GCN Gesture Classification Results

Gesture Class	Total Samples	Correct Classifications	Accuracy (%)	False Positive Rate (%)
Normal Activity	200	196	98.0	2.0
Emergency Signal 1	200	191	95.5	0
Emergency Signal 2	200	193	96.5	0
Distress Wave	200	188	94.0	1.5
Overall	800	768	96.0	1.2

C. System Response Time

The end-to-end response time from gesture occurrence to alert delivery is evaluated across multiple network conditions. Table 3 summarizes the system latency performance.

Table 3: System Response Time Analysis

Stage	Average Time (ms)	Minimum (ms)	Maximum (ms)	Remarks
Frame Capture & Preprocessing	18	14	25	Stable across conditions
Hand Detection (YOLOv8)	22	18	30	Real-time capable
Keypoint Extraction	15	12	20	MediaPipe optimized
ST-GCN Classification	35	28	45	Temporal window processing
Alert Generation & Dispatch	1800	1200	2800	Network dependent

D.Overall System Performance Summary

Table 4 consolidates the key performance metrics of the GestureGuard system evaluated across all test conditions.

Table 4: Overall GestureGuard System Performance

Performance Metric	Measured Value	Target Range	Performance Level	Remarks
Hand Detection Accuracy (%)	95.4	≥ 90	High	Robust across conditions
Gesture Classification Accuracy (%)	96.0	≥ 92	High	Low false positive rate
False Positive Rate (%)	1.2	≤ 5	Low	Reliable emergency detection
Avg. Inference Latency (ms)	90	≤ 150	Fast	Suitable for real-time use
Alert Delivery Time (s)	2.1	≤ 3	Efficient	Multi-channel dispatch
System Uptime (%)	99.1	≥ 98	High	Continuous 24/7 operation

VI. CONCLUSION

GestureGuard presents a comprehensive AI-powered surveillance system for detecting emergency hand gestures in real time from CCTV video feeds. By combining YOLOv8 for fast and accurate hand localization with ST-GCN for spatiotemporal gesture classification, the system effectively

identifies distress gestures and triggers automated multi-channel alerts to authorities with geolocation information. Experimental results demonstrate high detection and classification accuracy with low false positive rates and end-to-end response times within practical emergency thresholds.

The system directly addresses the safety needs of vulnerable individuals who may be unable to call for help verbally, providing a silent yet powerful emergency communication channel. Its modular architecture enables deployment across diverse environments including homes, schools, public transportation, and hospitals. Future work will focus on full-body emergency activity recognition, edge AI deployment near camera hardware for reduced latency, integration into smart city surveillance networks, and development of a dedicated mobile application for first responders.

REFERENCES

[1] R. Rautaray and A. Agrawal, "Vision Based Hand Gesture Recognition for Human Computer Interaction: A Survey," Artificial Intelligence Review, vol. 43, no. 1, pp. 1–54, 2015.

[2] S. Kopuklu, N. Kopuklu, and G. Rigoll, "Real-time Hand Gesture Detection and Classification Using Convolutional Neural Networks," in Proc. IEEE FG, 2019, pp. 1–8.

[3] S. Yan, Y. Xiong, and D. Lin, "Spatial Temporal Graph Convolutional Networks for Skeleton-Based Action Recognition," in Proc. AAAI Conf. Artificial Intelligence, 2018, pp. 7444–7452.

[4] L. Shi et al., "Skeleton-Based Action Recognition with Directed Graph Neural Networks," in Proc. CVPR, 2019, pp. 7912–7921.

[5] G. Jocher et al., "Ultralytics YOLOv8," Ultralytics, 2023. [Online]. Available: https://github.com/ultralytics/ultralytics.

[6] C. Lugaresi et al., "MediaPipe: A Framework for Perceiving and Processing Reality," in Workshop on Perception for Autonomous Systems, CVPR, 2019.

[7] M. A. Hossain and G. Muhammad, "Emotion Recognition Using Deep Learning Approach from Audio–Visual Emotional Big Data," Information Fusion, vol. 49, pp. 69–78, 2019.

[8] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," in Proc. CVPR, 2005, pp. 886–893.

[9] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in Proc. NeurIPS, 2012, pp. 1097–1105.

[10] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," in Proc. ICLR, 2015.

[11] Z. Cao et al., "OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields," IEEE Trans. Pattern Anal. Mach. Intell., vol. 43, no. 1, pp. 172–186, 2021.

[12] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement," arXiv:1804.02767, 2018.

[13] A. Dosovitskiy et al., "An Image is Worth 16×16 Words: Transformers for Image Recognition at Scale," in Proc. ICLR, 2021.