

# Reinforcing Local Safety Ecosystems: Integrating the "Signal for Help" Detection in Urban and Inner Territories

Federico Buccellato(✉), Sarah Azimi, Corrado De Sio, and Luca Sterpone

Dipartimento di Automatica e Informatica, Politecnico di Torino, Turin, Italy
{federico.buccellato,sarah.azimi,corrado.desio,luca.sterpone}@polito.it

**Abstract.** The *Signal for Help* is a discrete hand gesture that allows a person in a situation of risk to ask for help without speaking and without attracting the attention of a potential aggressor. Although the gesture is widely recognized, it produces a concrete effect only if it is detected promptly and if the alert reaches someone able to intervene. This paper addresses the gap between recognition and intervention by examining how a gesture detection system can be integrated into existing territorial security infrastructures.

Building on a previously validated recognition model, the work focuses on its effective integration into existing video surveillance networks, operational workflows, and local security services. The analysis highlights the architectural, organizational, and governance conditions that enable a sustainable and responsible implementation, both in urban contexts with higher institutional density and in inner areas characterized by more limited resources.

From this perspective, the contribution shows how a silent signal can be incorporated into a structured and traceable decision-making chain, strengthening local security ecosystems while maintaining human oversight and compliance with data protection principles.

**Keywords:** Signal for Help · Territorial Security · AI for Social Good · Privacy-by-Design · Territorial Deployment

## 1 Introduction

One in three women worldwide experiences physical or sexual violence during her lifetime [1]. This figure is not merely a statistic; it reflects a structural condition of vulnerability that spans different social, geographic, and cultural contexts. In many cases, a person at risk is not able to ask for help explicitly. She may be under the direct control of the aggressor, in a situation where a verbal request would immediately attract his attention.

During the COVID-19 pandemic, forced isolation and reduced social interaction made this vulnerability even more visible. In this context, the *Signal for Help* was introduced, a simple and silent hand gesture designed to allow a person

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2027
O. Gervasi et al. (Eds.): ICCSA 2026 Workshops, LNCS 16761, pp. 162–176, 2027.
https://doi.org/10.1007/978-3-032-30527-5_11

to indicate that she is in danger without speaking. The gesture is performed by raising one hand with the palm facing outward, folding the thumb into the center of the palm, and then closing the remaining fingers over it, as shown in Fig. 1. The sequence is discreet but intentional and can be recognized by those who are familiar with its meaning. In several documented cases, its correct interpretation enabled timely interventions [2].

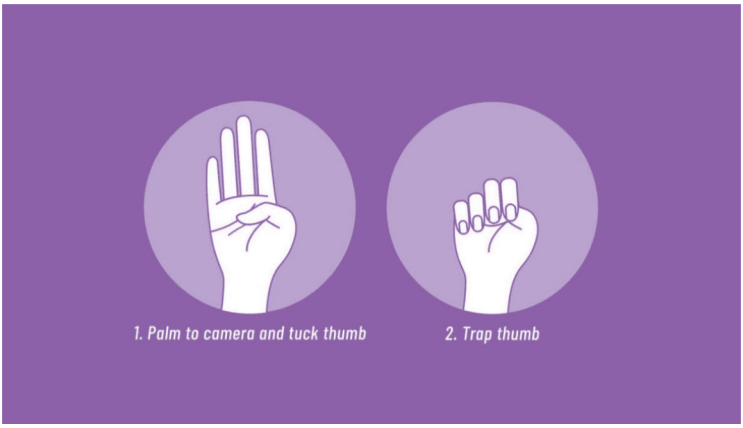


**Fig. 1.** The "Signal for Help" hand gesture.

However, its effective use is contingent on certain circumstances which cannot always be ensured. In situations when the environment features an increased level of visual complexity, overcrowding, or other factors which distract a person's attention, a gesture may either go unnoticed or get misinterpreted [4]. It indicates that there is a structural problem in question. The mere presence of means of communication does not suffice if there is no way to detect it and implement an appropriate operational reaction.

In recent years, advances in computer vision and gesture recognition have made it possible to automatically identify hand movements in real time, even within video streams from large-scale surveillance networks. Methods based on hand landmark detection, spatiotemporal modeling of the sequences, and training a neural network classifier specifically to perform tasks at the edge have shown their ability to recognize complex configurations of gestures in realistic conditions. In this context, a system capable of detecting the *Signal for Help* within video streams and generating structured events for security operators has been developed and validated. The model described in [3] provides the technological foundation for the present work.

Automatic recognition, however, constitutes only one side of the coin. Gesture detection attains practical significance when it is placed within a broader operational framework that encompasses alert triggering, operator confirmation, and possibly the initiation of action plans.

The purpose of this paper is to explore how the technology can be integrated into different territorial contexts, with specific emphasis on the contrast between urban settings and rural regions. Variations in connectivity, access to computing facilities, institutional presence, and arrangement of security forces have an immediate impact on the selection of deployment strategies. The research highlights the architectural, organizational, and administrative factors that allow for the sustainable application of the innovation amid diverse circumstances.

The objective is to demonstrate how a silent signal can be incorporated into a traceable decision-making structure that is proportionate to the resources available, thereby strengthening local security ecosystems while maintaining human oversight and compliance with data protection requirements.

## 2 Related Works

In the last few years, there have been several developments regarding the utilization of technological systems for providing assistance in ensuring safety, especially when it comes to automatic detection of situations involving violence or danger. The currently available systems try to identify aggression by recognizing various signs like physical fights, abrupt body movement, facial expression of fear or sound of distress.

In this regard, various video-based techniques have been devised. Some of these techniques depend upon identifying the visual cues of physical attacks that may include the variation in colors indicating the presence of blood or specific body postures for attacks [5,6]. Some other techniques make use of combining visual and audio clues like screaming and unusual sounds to make detection more effective [7]. Such techniques prove effective when there are some distinct signs of violence that can be easily observed. However, they show limitations in situations where a person at risk cannot openly express her condition and must rely on more discreet forms of communication.

Signal for Help constitutes such form of communication. Signal for Help is an intended and silent signal that one sends out requesting help without alarming any potential aggressor. Since it does not involve abrupt movements, audible signals, or visible physical evidence, it cannot be effectively detected by traditional violence detection systems, which are designed to identify explicit and visually marked events.

This specificity requires a shift in perspective. Instead of focusing on the recognition of violent actions, attention moves toward gesture recognition, that is, the identification of intentional hand movements within video sequences. Automatic gesture recognition is widely used in several application domains, including humanâĂŞcomputer interaction, device control, and automatic sign language translation. Techniques developed in these contexts, based on convolutional neural networks, temporal modeling, and hand landmark analysis [8–11], have demonstrated the ability to describe and distinguish complex movements in video sequences by capturing both their spatial structure and temporal evolution.

On this basis, such methodologies have progressively been applied to the recognition of the *Signal for Help*. Some contributions have focused on training models capable of distinguishing the gesture from similar hand movements [12, 13], achieving relevant technical results. However, in many cases the analysis has remained limited to the classification phase, with primary attention given to model performance and limited consideration of integration into real operational contexts.

One element that has not yet been extensively examined within the academic literature is the systematic element of gesture recognition in a critical context. While most studies concentrate on detection reliability and performance assessment, research on how such systems may be incorporated into the present security framework and linked to formalized warning handling protocols is still limited. In the absence of incorporation within the decision-making process, gesture detection technology runs the risk of being merely a technical achievement without an impact on intervention procedures.

Our research group has actively worked to overcome these technical barriers through a consolidated line of work initiated in 2023 [14]. By developing an extensive dataset of 4,500 video sequences featuring hundreds of participants, we validated a system based on a Random Forest classifier capable of 94% accuracy [1], specifically designed to recognize the gesture in realistic and difficult positions, such as when the hand is held low or behind the back. However, having established this robust technological foundation, the challenge now shifts from the reliability of the detection to its actual territorial deployment.

The present work addresses this research gap. Moving beyond the algorithmic validation achieved in our previous studies, the focus is now on analyzing the conditions that enable the adoption of this technology in real contexts. In particular, the research is aimed at analyzing the ways of configuring and incorporating the technology in urban contexts and inner territories in view of infrastructure, organizational, and institutional peculiarities of different territories. The goal is to conceptualize the introduction of the signal as part of a security system of territories, not just a technical innovation.

## 3 System Architecture

This section includes all information needed about the operation of the system to understand its performance, architecture, and integration into the existing surveillance system. Such a description builds a bridge between the technological side of the system and its further evaluation from the architecture and territorial perspectives.

The system is a technological component of a security system that will detect silent alarms in existing video surveillance systems. As mentioned above, the technology used in the proposed solution is based on the technology presented in [3]. The technology allows detecting a *Signal for Help* in a video stream and notifying security officers about it.

The system consists of several layers, including gesture detection and alarm notification. The system was created in such a way to ensure its compatibility with existing surveillance systems.

### Video Analysis and Gesture Recognition Layer

The first level of the architecture concerns the analysis of videos and the recognition of gestures. As shown in Fig. 2, the procedure starts by capturing video sequences using the surveillance cameras. For each frame that is captured in real time, a determination is made whether a hand appears and triggers the tracking procedure.

In case a hand is detected, keypoints are extracted in order to characterize the structure of the hand. Keypoints are then assembled in temporal sequences such that the shape of the hand in a given instant, as well as the evolution of the gesture in time, are taken into account. The spatiotemporal sequence is then fed into the classification module to assess the probability that the observed gesture is the *Signal for Help*.

In order to avoid generating events that have low probabilities, temporal consistency must be ensured. An event is generated if the detected movement is consistent throughout time. It signals the end of the recognition step and the start of the alarms handling step.

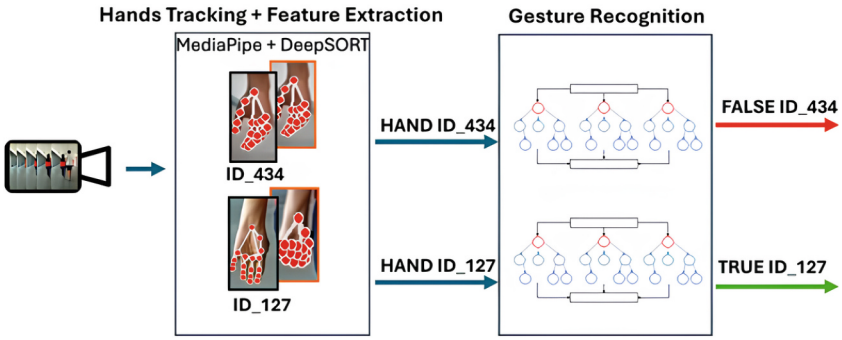


**Fig. 2.** Overall processing pipeline of the proposed system, from video acquisition to operator notification.

### Event Management and Alarm Processing Layer

The second layer processes the event produced by the recognition module and structures it into an operational alert. This means that at this stage, the event is assigned appropriate metadata such as the camera ID, the location of the detection, and the time. Additionally, a video segment from the detection moment is included in order to allow the verification of the event context.

At this level, the technical result produced by the classifier algorithm is translated into an alert suitable for inclusion into the existing workflow of the organization. Structured in such way, the alert is transferred via the system backend to appropriate operators.

### Interaction with the Operator Layer

The last layer involves interaction with the operator. The alert, together with the video segment, is transferred to the operator via the mobile app interface, which is presented on Fig. 3.

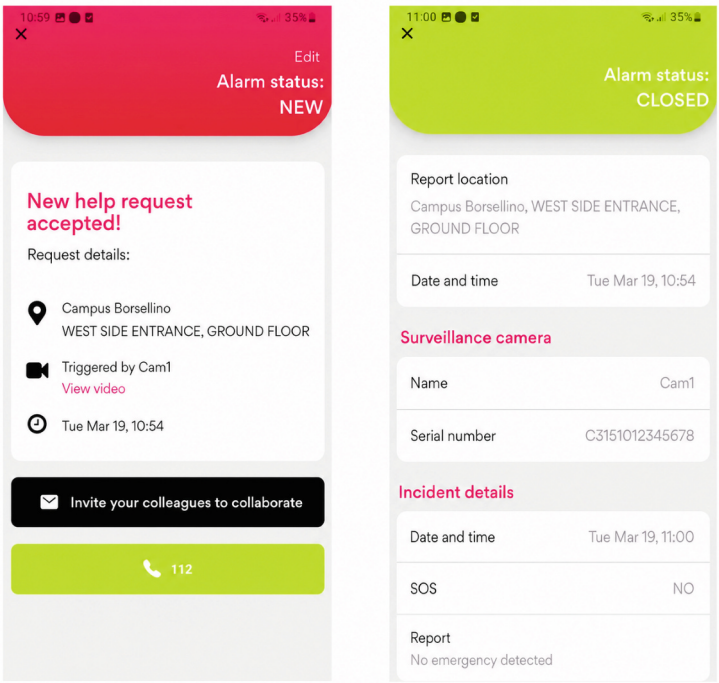


**Fig. 3.** Mobile application interface displaying alert information and related video preview.

The interface is meant to allow for rapid and informed analysis of the situation. The operator analyses the notification and determines if it is necessary to confirm the warning or not. Upon confirmation, the event will be handled based on the organization's procedures, including contacting personnel at the scene. All processes will be logged for transparency purposes.

The distinction between automatic recognition and human verification enables the system to augment current security systems without compromising the importance of the operator.

## 4 Deployment in Urban and Inner Areas

After having specified the architecture of the system and its components, it is necessary to proceed with its analysis for implementation within real territorial settings. The purpose of this section is to show how the model can be configured under the given infrastructure and organizational setting within a territory and identify key aspects of its design that determine its efficiency.

In the following, the two simplified types of areas âĂŞ *urban area* and *inner area* âĂŞ will be mentioned. This classification does not imply any rigid definition in demographic or administrative terms but is purely typological and operational. While these areas are traditionally defined by their proximity to essential service hubs like hospitals or schools, from the perspective of the proposed system, the distinction focuses primarily on technical and security-related dimensions, such as network stability and the density of the surveillance infrastructure. Both areas are defined in Table 1.

**Table 1.** Recurring territorial characteristics and possible implications for deployment

| Analytical Dimension | Urban Area | Inner Area |
|---|---|---|
| Connectivity | Widespread broadband availability<br>Greater stability of network infrastructures | Variability in connection quality<br>Possible bandwidth or coverage limitations |
| Computational Infrastructure | Presence of or access to centralized infrastructures<br>Use of data centers or dedicated cloud services | Lower availability of centralized nodes<br>Greater reliance on distributed or edge solutions |
| Organizational Structure | Dedicated operational units<br>Personnel with specialized roles | Light organizational structures<br>Multifunctional roles and inter-municipal coordination |
| Access to Emergency Services | Geographical proximity to services<br>Generally shorter intervention times | Greater geographical dispersion<br>Potentially longer response times |
| Prevalent Technical Configuration | Tendency toward centralized processing models | Tendency toward distributed or hybrid models |

The categorization does not mean generalized uniformity in any way, shape, or form. A small town can possess sophisticated networks, while a large city might struggle with its connectivity. The suggested differentiation should, thus, serve only as an analytical framework for making technical decisions.

This difference emphasizes the fact that decisions made in configuring systems should not be seen as detached from the environment in which they will be incorporated. The qualities of both infrastructure and organizations influence how computing work is distributed, events managed, and alerts integrated into the decision-making process.

Below in the text we will consider these aspects separately taking into account the typology suggested above.

### Deployment in Urban Areas

In those environments that might be linked to the category *urban area*, more technological infrastructure is available, as well as a better-organized approach to the management of the security process. Such aspects affect the integration and architecture design of the system.

Thanks to the relative stability of the connection and computing capacity of the system, the use of a centralized architecture will be quite possible. Under such architecture, the system will be embedded into the application that is run on the central node of the network, which receives video streams from the cameras deployed in the territory. All sources are collected in one place, analyzed through the recognition algorithm with a unified output of the results depending on parameters.

With centralized processing, there will be possibilities for managing configuration changes, updating algorithms, and maintaining models all at one place. Parameters, detections levels, and algorithms can be changed at the central infrastructure level and will have an immediate impact throughout the whole set of cameras. There will be no need to carry out any intervention in a distributed manner.

For urban settings where the visual density and complexity is high, such as stations, public places, shopping centers, or city centers having high pedestrian density, the amount of live video streams may be considerable, and manually watching all of them would require substantial efforts. Here, the system provides a function of automatic selection of events where there are requests for help that are unlikely to attract attention in situations when many visual stimuli exist at once.

Configuration of the system can also be made to match the resources on which it operates. Adjusting the threshold value and parameters of the model can be achieved through the sensitivity parameter of the algorithm. Sensitivity in the configuration makes detection of faint signals probable; however, it also makes false alerts more likely, increasing the number of notifications generated.

On the other hand, restricting the sensitivity decreases alert generation but may also increase the chance of overlooking an incident. Selection of the sensitivity value depends on the number of operatives available, their monitoring capability, and the feasibility of alert generation. In this way, the model can accommodate real-world circumstances without the risk of overwhelming the operatives or limiting detection capability excessively.

In general, for deployments that fit into the *urban area* category, emphasis is placed on centralized processing of the data and gesture recognition through current infrastructure. The model serves as an improvement to current capabilities, seeking to enhance the detection of silent cries for assistance without modifying the technological configuration.

### Deployment in Inner Areas

Where the *inner area* concept applies, the availability of infrastructure will be poorer, connectivity will not always be reliable, and real-time transmission of video streams to the central servers might not be always possible. This directly influences architectural design and makes a solution that would be built around real-time transmission of all video streams to a processing center at a distance suboptimal.

Here a different design approach would be required. Gesture recognition of the *Signal for Help* will take place using edge computing solutions located next to the camera or on territorial nodes that several cameras share. Processing takes place right on-site and transmission is done only in response to detected events; continuous transmission of video streams is unnecessary.

Short clips related to an attempt to make a signal will be transmitted along with relevant metadata that is required for context analysis. By doing things this way, we minimize data traffic while simultaneously providing functionality under the condition of insufficient or unreliable Internet connectivity. Locally processed signals give us robustness. Should connectivity fail, video streams will still be monitored and events recorded.

From an operational perspective, inner areas may be characterized by a lower density of active video streams and fewer simultaneous events. Under these conditions, the model can be configured with higher sensitivity, increasing the probability of detecting even weak or ambiguous signals. Although this may lead to a higher number of generated alerts, the overall information load generally remains manageable due to the lower number of active sources.

The goal here is to avoid overlooking anything, regardless of the cost of getting more false alarms, as long as information overload does not become a problem.

It becomes clear that, when fewer resources are available in terms of infrastructure and organization, the conditions of operation might not necessarily be able to be replicated in the same way as seen in more controlled urban environments. Yet, by virtue of its distributed architecture and adaptability, the system will be able to provide services at a level in line with the environment's characteristics. The goal is not to reproduce an urban model on a smaller scale, but to ensure a balance between detection reliability, technical sustainability, and compatibility with local resources.

## 5 Privacy and Obfuscation Configuration

The introduction of an automatic recognition system in the security infrastructure of territories implies a certain responsibility concerning data protection. Indeed, the identification of the *Signal for Help* results in the creation of video clips with potential identifiable data, in an operational environment. Therefore, data protection must be a structural part of the deployment strategy rather than a secondary one.

That is why the system was built following the *Privacy by Design* principle, which means that all necessary data protection measures have been incorporated in the architectural design. Inter-module communications are ensured by encrypted channels, video clips are stored in secure environments that require authentication, while the interface for consultations is available only to authenticated operators. All accesses, displays, and modifications of any status of the events will be logged in order to establish the limits of decision-making when it comes to visual data processing.

In the frame of such architecture, the system offers some settings for the adjustment of the visual data processing process in accordance with the deployed territory. It is possible, for example, to set the level of data obfuscation of video content differently.

### 5.1 Obfuscation Levels

Depending on the operational necessities and specific territories' features, the system offers three types of control over visual data, distinguished based on the level of content obfuscation.

**No Obfuscation Mode:** The video is transmitted and stored without any alterations. No method of anonymization is used. Such an approach is generally utilized in circumstances where the process of identifying individuals is a part of official procedures.

**Dynamic Obfuscation Mode:** The file is kept in its pristine format; however, obfuscation occurs dynamically when visualizing. This manipulation will be imposed on the images during playback, and only authorized individuals are allowed access to the unobfuscated version of the same. With this approach, there is decreased systematic exposure of identifiable information in the process, while still maintaining an option for verification.

**Irreversible Obfuscation Mode:** The obfuscation technique is directly implemented into the processing pipeline and is carried out prior to storing and transmitting the video clip. The video clip is rendered permanently anonymous, and no copies of the original data are kept.

One such illustration of the visual impact of obfuscation can be seen in Fig. 4, wherein the important area has been obfuscated while ensuring that the rest of the scene remains comprehensible.

The selected configuration should be specified before system deployment, and it must be integrated into the territorial implementation plan. This configuration is not only a technical parameter, but it also represents a design choice with direct implications for data handling procedures and the balance between operational necessity and privacy safeguards.

Clearly stating the level of obfuscation prior to carrying out system tasks will allow for the setting up of certain boundaries with regard to visual data processing, avoid ambiguities during operations, and ensure compliance with existing laws concerning security and privacy in the territory. In doing so, the



**Fig. 4.** Example of obfuscation applied to the clip associated with a detected event.

obfuscation configurations are not decided on a case-by-case basis, but they become part of a controlled process.

The degree of anonymization therefore becomes a fundamental element of the local implementation process and of the operational architecture as a whole, forming part of the governance structure of the surveillance system and of the role definition in its use.

Beyond technical and regulatory considerations, the definition of obfuscation levels also contributes to the social feasibility of the system. In inner areas, where social norms, familial ties, and the risk of stigmatization may discourage individuals from seeking help explicitly, limiting the visibility of identifiable information becomes particularly relevant. In this context, the adoption of dynamic or irreversible obfuscation reduces the perceived intrusiveness of the system, as visual data is either conditionally accessible or permanently anonymized. This characteristic supports the deployment of the system in socially sensitive environments, where maintaining discretion and minimizing exposure are essential for ensuring acceptance, while still enabling the detection and analysis of critical events.

## 6 Operational Assessment and Territorial Resilience Metrics

To assess the effect of the system on its performance according to the various territorial environments, the quantitative assessment of the key indicators was made, taking into account the centralized and decentralized architecture. The measurements are carried out in controlled settings and form a numeric basis that reflects a real situation; thus, the effect of architectural design and privacy configuration on response time and sustainability is analyzed.

### Performance Analysis and Operational Implications

The performance evaluation was done by separating the latency for detecting a single event from the entire end-to-end process. This approach ensures that the time taken by the algorithm on its own can be distinguished from the total time for the entire information cycle until it becomes an operationally actionable notification.

Regarding the detection stage, the mean delay recorded between the receipt of the frame and the construction of the event structure is 63 ms in the centralized setting and 340 ms in the distributed edge setting, as shown in Table 2. Such an evident disparity can be explained by the dissimilarities in the availability of computational resources. In a centralized system, there is more computation power and ability to parallelize. On the other hand, edge computing is constrained by power and memory limitations. Nonetheless, both systems are comparable in terms of low latencies.

**Table 2.** Average detection time by configuration

| Configuration | Average Time (ms) | Std. Dev. (ms) |
|---|---|---|
| Centralized | 63 | 8 |
| Distributed | 340 | 45 |

However latency in detection does not suffice in defining operationally effectiveness. When analyzing the entire duration from the emergence of the gesture until the notification appears on the interface, the average duration is 2.9 s for the urban centralized scenario, while in the distributed deployment in inner areas, the average duration reaches 5.2 s (Table 3). In this case, the difference is not attributable solely to inference time but reflects the impact of network connectivity and node placement. In urban contexts, network stability and bandwidth availability support faster transmission, whereas in inner areas greater connection variability and selective data transmission influence total notification time. From an operational perspective, both values remain within a time window compatible with coordinated response activation.

**Table 3.** Average end-to-end notification time

| Configuration | Average Time (s) |
|---|---|
| Urban Centralized | 2.9 |
| Inner Areas Distributed | 5.2 |

Another aspect analyzed was the influence of varying modes of anonymization upon the processing pipeline itself. As illustrated in Table 4, dynamic obfuscation

performed during visualization stage, which is run on the client side, has no impact whatsoever on event generation duration, and adds an average overhead of 52 ms in either case. Irreversible obfuscation, on the other hand, integrated within the processing pipeline and affecting detection stage, adds an average overhead of 80 ms to the centralized scenario, as well as 673 ms to the distributed one. In general, this overhead grows even further when we consider edge-based scenarios. This proves that the level of anonymization adopted has a clear impact on performances and needs to be tailored to the characteristics of the territory.

**Table 4.** Average overhead introduced by obfuscation

| Mode | Centralized | Distributed |
|---|---|---|
| No obfuscation | 0 ms (0%) | 0 ms (0%) |
| Dynamic visualization obfuscation | 52 ms (+82%) | 52 ms (+15%) |
| Irreversible pipeline obfuscation | 80 ms (+127%) | 673 ms (+198%) |

These metrics prove that the degree of anonymization impacts the performance characteristics of the system, especially in scenarios with limited computational capacity.

Finally, the system's impact on operational workload was estimated in a high-density urban environment. The false positive rate is approximately 2%, with an average verification time per event between 20 and 25 s (Table 5).

**Table 5.** Estimated operational workload

| Parameter | Estimated Value |
|---|---|
| False positive rate | $\sim$ 2% |
| Average verification time per event | $\sim$ 20–25 s |

A false positive rate of 2% implies that a portion of generated notifications requires manual verification even in the absence of a real emergency. In high-density environments with continuous video streams, this may translate into a non-negligible number of additional checks, requiring operator time and attention.

Therefore, an assessment of the organizational effect of the system should be conducted in the context of staff availability and the ability to accommodate verification tasks. The sustainability of the operation not only relies on the precision of the model but also on the connection between the number of notifications and the human capital invested in managing them.

Overall, the measurements suggest that architectural variations affect performance technologically without interfering with the main purpose of the whole

system. In both centrally structured and decentralized setups, measured times coincide well with instant triggering of the process of making decisions. This performance consistency across different infrastructures demonstrates the territorial resilience of the system, which we define as its ability to guarantee reliable safety standards even when operating under the technical and connectivity constraints typical of marginalized or isolated areas. This means that the model is adaptable to diverse infrastructure conditions and still achieves the proper balance of responsiveness, security, and sustainability.

## 7 Conclusions and Future Works

This research project has discussed the inclusion of Signal for Help identification in territorial security systems, looking at implementation methods in various settings.

It has been demonstrated that, different types of configuration of the proposed system may be utilized, depending on the territorial characteristics; thus, while centralized solutions work better in cities where infrastructure is well-developed, decentralized solutions correspond more to inner territories lacking resources. Through such adaptation to the specific circumstances, the technology becomes very effective in increasing safety within the territory. In the case of urban environments, the technology offers an automatic mechanism for filtering the silent calls for help from the high level of visual activity in the area, while in the case of internal territories, the technology serves as an important technological link where the presence of institutions is less dense.

The performance evaluation demonstrates that although there are some differences in technical aspects of operation, both configurations guarantee a response time sufficient to make a quick decision. Beyond technical efficiency, the social impact of this system lies in its ability to bridge the gap in emergency services, offering a discrete and safe communication channel for vulnerable groups who might otherwise remain invisible. By integrating this detection capability into existing infrastructures, territories can achieve a higher level of resilience, transforming standard surveillance into a proactive and socially responsible safety ecosystem.

As a next step, future studies should pay their attention to the verification of the system in the form of pilot projects implemented in different territorial settings. Such studies would help to test the efficiency of the model in practice, abandoning theoretical assumptions and examining practical restrictions, costs, and interactions between institutional agents. Experimenting would be of crucial importance for further improving response strategies and making sure that the system remains sustainable in local security governance in the long run.

## References


1. World Health Organization. Devastatingly pervasive: 1 in 3 women globally experience violence (2021). https://www.who.int/news/item/09-03-2021-devastatingly-pervasive-1-in-3-women-globally-experience-violence, Retrieved 15 Mar 2025
2. MilanoToday. Violenza sessuale in centro a Milano: ragazza si salva col segnale anti-abuso (2025). https://www.milanotoday.it/cronaca/violenza-sessuale-piazza-scala.html
3. Buccellato, F., Vacca, E., Azimi, S., De Sio, C., Sterpone, L.: From detection to intervention: an end-to-end system for recognizing the signal for help. Gesture in real-time. Intell. Syst. Appl. **26**, 200536 (2025). https://doi.org/10.1016/j.iswa.2025.200536
4. Ossino, A.: Il mistero del segnale in codice non capito dal cameriere, la lite e l'assedio in bagno. Così è morta Martina Scialdone. La Repubblica (2023). https://roma.repubblica.it/cronaca/2023/01/15/news
5. Clarin, C., Dionisio, J., Echavez, M., Naval, P.: DOVE: Detection of movie violence using motion intensity analysis on skin and blood. Philippine Comput. Sci. Congress (PCSC) **6**, 150–156 (2005)
6. Zajdel, W., Krijnders, J.D., Andringa, T., Gavrila, D.M.: CASSANDRA: audio-video sensor fusion for aggression detection. In: IEEE Conference on Advanced Video and Signal Based Surveillance (AVSS), pp. 200–205 (2007)
7. Giannakopoulos, I., Pikrakis, A., Theodoridis, S., Manolopoulos, Y.: Aggression detection using audio and visual cues. In: IEEE International Conference on Image Processing (ICIP) (2009)
8. Tran, D.-S., Ho, N.-H., Yang, H.-J., Baek, E.-T., Kim, S.-H., Lee, G.: Real-time hand gesture spotting and recognition using RGB-D camera and 3D convolutional neural network. Appl. Sci. **10**(2), 722 (2020). https://doi.org/10.3390/app10020722
9. Junping, H., Siping, X.: Gesture recognition based on YCbCr color space and neural network. In: 6th International Conference on Intelligent Computing and Signal Processing (ICSP), Xi'an, China, pp. 738–741 (2021). https://doi.org/10.1109/ICSP51882.2021.9408765.
10. Peng, Y., Tao, H., Li, W., Yuan, H., Li, T.: Dynamic gesture recognition based on feature fusion network and variant ConvLSTM. IET Image Proc. **14**, 2480–2486 (2020). https://doi.org/10.1049/iet-ipr.2019.1248
11. Krishnan, G., Joshi, R., O'Connor, T., Pla, F., Javidi, B.: An overview of hand gesture recognition in degraded environments using three-dimensional integral imaging and deep neural networks. OSA Imaging and Applied Optics Congress (2021)
12. Elliott, G., Meehan, K., Hyndman, J.: Using CNN and tensorflow to recognize signal for help. IEEE Hand Gestures (2021)
13. Azimi, S., De Sio, C., Sterpone, L.: Enhanced video surveillance systems for signal for help. In: IEEE International Symposium on Technology and Society (ISTAS), Detection on Edge Devices(2023)
14. Azimi, S., De Sio, C., Carlucci, F., Sterpone, L.: Fighting for a future free from violence: a framework for real-time detection of signal for help. Intell. Syst. Appl. **17**, 200174 (2023). https://doi.org/10.1016/j.iswa.2022.200174