CSIS Ecosystem: Impacts of Pragmatic Detection on a Campus Surveillance Case Study

Babacar Mane[1], Fernando H. de A. Moraes Neto[1], Roberto de Cerqueira Figueiredo [1], Caio Nery[1], Diana Romero Clavijo[1], Samuel Rios da Silva[1], Daniela Barreiro Claro[1], Celia Ghedini Ralha[1], Ana Patricia Magalhães[2], Rita Suzana Pitangueira Maciel, Marlo Vieira dos Santos e Souza[1], George Marconi de Araújo Lima[1], Bruno Pereira dos Santos[1], Robespierre Pita[1], Edlane Proença, Luis Emanuel Neves de Jesus[1], Iala Patrícia de Jesus Monteiro de Jesus[1]

[1]FORMAS Research Center on Data and Natural Language
Institute of Computing - Federal University of Bahia (UFBA)
Av. Milton Santos, s/n - Ondina, Salvador/BA, 40170-110,

[2]Department of Exact Sciences and Earth - State University of Bahia (UNEB)
Rua Silveira Martins, 2555, Cabula, Salvador/BA, 41150-000,

{babacarm, fernando.humberto, roberedo, caionms, dianaclavijo, samuelrs, dclaro, celiaralha, rita.suzana, msouza1, gmlima, brunops, ropespierre.pita, edlane, luis.emanuel}@ufba.br,ialapjmonteiro@gmail.com, anapatriciamagalhaes@uneb.br

Abstract. ***Research Context:*** *In modern universities, where there is a constant flow of people, it is essential to implement computer vision surveillance systems to detect incidents and alert security personnel, in order to ensure the safety of all members of the academic community.* ***Practical Problem:*** *Universities are faced with numerous security issues, including limited resources, a diverse range of technical infrastructure, and a mix of old and new systems. Besides, traditional security methods are reactive, slow to notice incidents, which in turn makes quick response in emergencies very hard.* ***Proposed Solution:*** *We present the solution in the form of the Campus Surveillance Interoperability System (CSIS), which is an open-source, modular, and interoperable architecture that we put forth for the specific surveillance needs of universities. CSIS utilizes a variety of computer vision models from the YOLOv11x set, specialized in recognizing weapons, fires, floods, graffiti, suspicious behavior, and license plates.* ***Related IS Theory:*** *Our paper adopts a Socio-technical theory, which recognizes that the effectiveness of information systems depends on the interaction between social and technical components rather than on technology alone.* ***Research Method:*** *In a real-world university setting, we put our proof of concept to the test. We had real-time video stream input, ensemble-based inference, and a centralized alert management system. We assessed the accuracy of the models.* ***Summary of Results:*** *We achieved an average accuracy of 83% for license plate recognition, weapon detection, graffiti, and smoke detections. Although we achieved moderate accuracy in fire detection and lower performance in flood detection. Considering sociotechnical concerns, we evaluate CSIS's ability to automate surveillance tasks.* ***Contributions and Impact to IS***

Area: *CSIS architecture introduces an interoperable architecture for intelligent surveillance, demonstrating a model of interoperability in real time;*

1. Introduction

The organizational and functional complexity of modern Federal Universities transforms them into dynamic sociotechnical ecosystems. These institutions are characterized as open academic environments where the free circulation of people and knowledge is fundamental to the proper functioning of academic life. They manage student and staff data and sensitive infrastructure (Information Technology, laboratories). This leaves universities susceptible to both physical and digital security threats.

To ensure adequate security, traditional surveillance approaches employed by universities are often insufficient to capture the diversity of incidents and the contextual nuances that characterize academic environments [Laroca et al. 2021, Delnevo et al. 2024, Keerthana et al. 2023, Zhou et al. 2022]. Moreover, many universities operate with limited financial resources and rely on heterogeneous infrastructures that must interoperate legacy systems with modern digital platforms. Within this context, universities increasingly confront incidents such as unauthorized access, armed robbery, fire outbreaks, flooding, suspicious behavior, and acts of vandalism.

In this context, the proposed solution is to invest in AI-based surveillance technologies to address the limitations of traditional surveillance approaches. The Campus Surveillance Interoperability System (CSIS) architecture, an open-source system designed to enhance university surveillance and real-time security response, integrates a set of YOLOv11x [Jocher and Qiu 2024] models to detect incidents such as weapon in hand, suspicious behavior, fire outbreaks, flooding, vehicle monitoring through license plate recognition, and acts of vandalism in a semantic level of interoperability [Figueiredo et al. 2025]. The selection of YOLOv11x is guided by practical considerations rather than comparative experimentation. YOLOv11x provides efficient real-time inference on available university hardware, integrates natively with the Ultralytics training and inference ecosystem to facilitate the development of multiple detectors, and benefits from stable documentation, tooling, and community support. At the time of implementation, this ecosystem was more mature than those of YOLOv9/10 or RT-DETR, and its performance-to-complexity trade-off is deemed adequate under the institutional constraints. More than this, the CSIS solution integrates a pragmatic level to enable the detection of suspicious behavior, considering vehicles and people, and to account for intention and context, thereby interoperating with traditional systems.

A proof of concept was implemented at a Federal University to evaluate its effectiveness. Technical experimental results demonstrated high detection accuracy, particularly in license plate recognition (96.79%). The CSIS models perform in detecting various security threats, with high accuracy for weapon detection (90%) and graffiti recognition (76%), moderate accuracy for fire and smoke (71%), and lower performance for flood detection.

Beyond such proofs of concept, we aim to describe the impacts of our approach within the Security Coordination sector of a Federal University. We assess these impacts by considering the staff, people, processes, and technologies involved in mediating or investigating an incident.

The findings suggest that CSIS solutions within the university environment are likely to yield substantial impacts, not only at the individual level by enhancing privacy protection, personal safety, and well-being but also at the organizational level, through improved institutional security management, operational efficiency, resilience, and technological integration, providing, thus, a socio-technical solution.

The remainder of this paper is organized as follows: Section 2 reviews related work; Section 3 describes the sociotechnical ecosystem currently operating within the Security Coordination sector, followed by an overview of the CSIS ecosystem at Section 4 that depicts CSIS social, organizational and technical impacts as surveillance solution for academic environments; Section 5 evaluates the performance of the CSIS architecture in a real-world context, and finally, Section 7 presents our Conclusions and Future Work.

2. Related Works

The integration of Artificial Intelligence (AI) into surveillance systems has marked a shift from traditional passive video monitoring to intelligent, automated detection and response. This shift has been driven by advances in machine learning algorithms and increased computational power, enabling real-time analysis and decision-making capabilities.

To address the security challenges inherent in complex environments such as universities, several solutions have emerged. For example, [Santos et al. 2023, Moura et al. 2021, Santos et al. 2024] proposed object detection frameworks for public safety based on YOLO models, focusing primarily on weapon and fire detection. Although effective in specific contexts, these solutions are often limited in terms of scalability and multi-threat integration.

The CSIS architecture proposes a comprehensive framework for real-time interoperability between surveillance devices and institutional systems. At its core, CSIS integrates YOLO-based computer vision models for object detection and image classification. Originally introduced by [Redmon et al. 2016], YOLO formulates object detection as a single regression problem, enabling fast and accurate inference in a single forward pass. Owing to its real-time performance and architectural efficiency, YOLO has become a widely adopted solution in surveillance applications.

Over the years, the YOLO framework has evolved. Notable developments include the C3k2 module introduced in YOLOv11x [Jocher and Qiu 2024], which improves feature fusion, and the attention-centric enhancements of YOLO12 [Tian et al. 2025], which aim to improve detection in complex and cluttered environments. These architectural refinements address persistent challenges related to inference speed, multiscale accuracy, and resource efficiency [Wang and Liao 2024].

Recent applications have demonstrated YOLO's versatility in university surveillance contexts. Authors in [Xiao et al. 2024] developed an autonomous inspection robot equipped with YOLO to detect threats during patrols across academic premises, exemplifying the integration of object detection with robotic systems. Similarly, [Tantra and Widjaja 2024] implemented a YOLO model to enforce campus dress code compliance, achieving a mean precision of 51.8% and an F1-score of 45%. Although modest in accuracy, their work highlights the potential of computer vision for non-traditional monitoring tasks aligned with institutional policies.

These advancements demonstrate the feasibility of processing video streams in real time to identify diverse events, including fires, smoke, cold weapons, firearms, floods, graffiti, and vehicle license plates. By integrating such models, CSIS provides simultaneous detection of multiple incident types, surpassing earlier solutions that often focused on isolated threats.

Beyond the technical perspective, multiple studies have underscored the relevance of AI-based surveillance in educational environments. Authors in [de Andrade et al. 2024] highlight that intelligent monitoring systems enhance public perception of safety and aid in crime prevention. Authors in [Shiri 2024] highlight their effectiveness in deterring recidivist offenders, while [Kerich et al. 2024] outlines common institutional threats, such as theft and vandalism, that disrupt campus operations.

Despite growing adoption of these technologies, security incidents persist. Authors in [Ekpoh et al. 2020] identified insufficient technology and a lack of trained personnel as key factors behind the ongoing crime in Nigerian universities. This reveals a persistent gap between existing surveillance infrastructure and the evolving demands of campus safety. In this context, interoperability emerges as a critical requirement.

Authors in [Anagnostopoulos et al. 2021] conducted a comprehensive survey on smart campus surveillance systems and emphasized the need for seamless integration across heterogeneous technologies. However, the diversity of heterogeneous systems in a campus university requires a direction for a full interoperability solution [Maciel et al. 2019]. In this perspective, authors in [Mane et al. 2021, Ribeiro et al. 2019] propose solutions for interoperability at the syntactic, semantic, and pragmatic levels, the latter referring to shared expectations regarding the effects of exchanged messages. In such a complex environment, taking a three-fold perspective of full interoperability [Maciel et al. 2019], sociotechnical [Ralha et al. 2025a], and AI-based Information Systems [Ralha et al. 2025b], this work introduces CSIS, an interoperable approach through syntactic, semantic, and pragmatic levels to provide smart campus university surveillance as its case study.

By incorporating advanced object detection models and supporting interoperability at syntactic, semantic, pragmatic levels, the CSIS architecture directly addresses the gaps identified in the literature, offering a scalable and comprehensive solution for modern university surveillance.

3. Sociotechnical Ecosystem

From a sociotechnical ecosystem perspective, the current processes of the Security Coordination Unit at a Federal University integrate human, technological, and organizational components that operate interdependently to ensure institutional security.

At the social level, the ecosystem comprises surveillance operations carried out by a rotating team of 12 operators. Three operators are assigned to day shifts and three to night shifts, thereby ensuring uninterrupted 24/7 monitoring of a network of over 650 surveillance cameras. The six operators are scheduled to work on alternate days. Their primary responsibility is to observe, identify, and document incidents captured by the monitoring system, thereby contributing to the university's overall safety and security. The broader security team consists of 95 agents engaged in daily field activities, and

the Security Coordination Unit is under two supervisors. Human intervention remains central. As evidenced in the monitoring process, manual actions continue to outweigh automated solutions in overseeing the university environment. For example, it is practically unfeasible for three operators to continuously monitor all events captured by 650 surveillance cameras. The fatigue associated with such intensive human surveillance, coupled with the limited resources typically available in universities, underscores the need for new technical and scientific advancements to improve monitoring practices.

At the technical level, the process involves the use of monitoring tools, access control systems, cameras, networks, radio communication, a dedicated social media application (Short and Instant Message Service), and information platforms that support the collection, analysis, and dissemination of security data. The university operates a network of 650 surveillance cameras distributed in an area of approximately 50,000 km². Within this traditional surveillance framework, threats are typically detected through three primary channels: direct observation by surveillance operators via monitoring screens, identification by local security agents, or reporting by members of the institution. Once a threat is identified, a team of security agents is promptly alerted through radio communication and a dedicated social media–based notification system (including Short and Instant Message Service) to intervene at the incident site. Simultaneously, the two supervisors of the Security Coordination Unit are generally informed in real time. The incident response process is adapted to the specific nature of each threat. For instance, in the event of a fire, if the intervention by security personnel proves insufficient to contain the situation, the fire department is automatically notified and dispatched. Likewise, in cases of suspicious behavior, once the threat is confirmed, a security officer is deployed to assess the individual. If the person is found to be armed, the Federal Police are immediately contacted to intervene on-site.

At the organizational level, policies, decision flows, procedures, and responsibilities define how human and technological resources are coordinated to detect, report, and mitigate incidents.

Accordingly, our ecosystem is designed to assist and complement the existing surveillance infrastructure, providing more robust support for informed and timely decision-making. In this context, the Security Coordination Unit operates as an adaptive ecosystem, where technological innovations, emerging threats, and changes in work practices continuously reshape the interactions among people, processes, and technologies.

4. CSIS Ecosystem

The CSIS (Campus Surveillance Interoperability System) architecture was developed to provide an intelligent, interoperable, and modular surveillance solution for academic environments.

4.1. CSIS Ecosystem Architecture

The CSIS architecture monitors the university surveillance environment through security cameras and delivers real-time notifications to a centralized dashboard [Figueiredo et al. 2025], as illustrated in Figure 1. It can detect and process incidents such as fire, smoke, weapon possession, flooding, suspicious behavior, vandalism (e.g., graffiti), and vehicle license plate recognition. The monitoring services are structured upon

multiple interoperability layers, such as syntactic, semantic, and pragmatic, supported by a centralized event handling hub module and an integrated events dashboard.

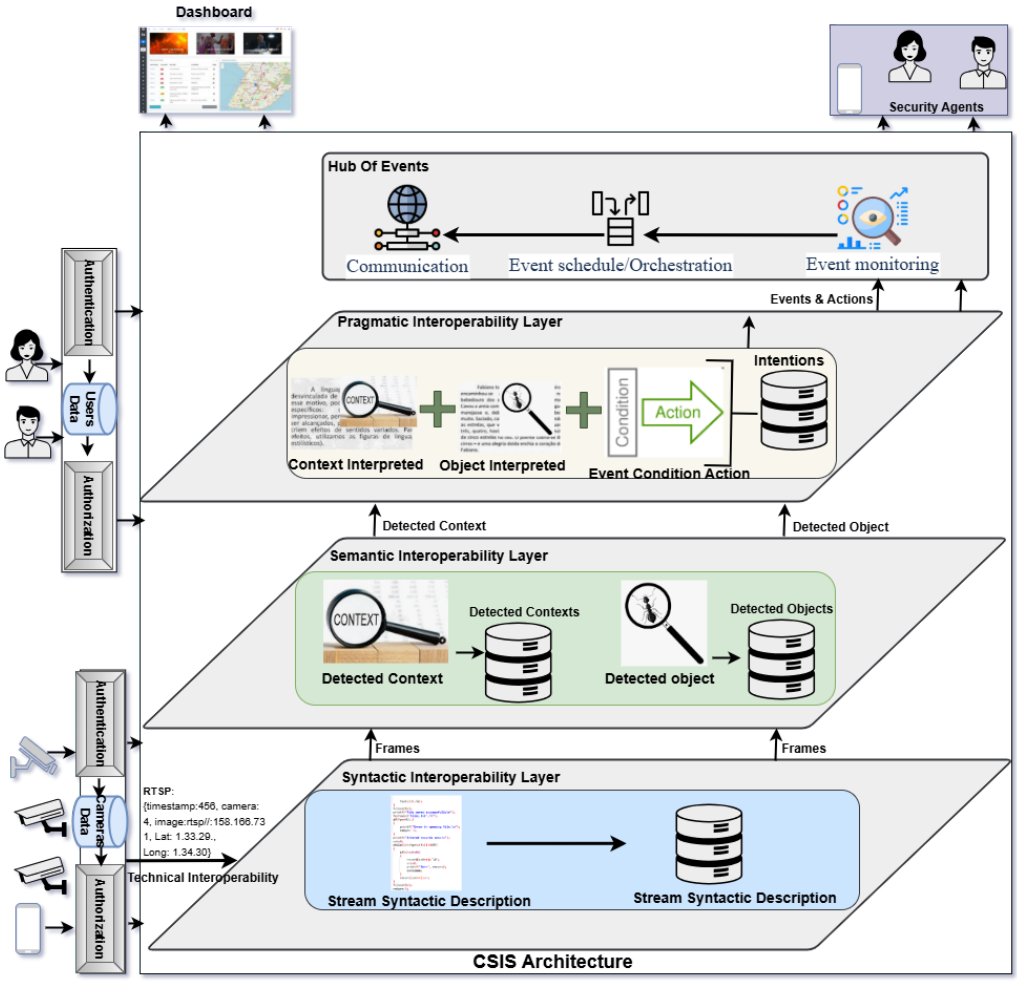

Figure 1. CSIS Architecture. Source: the authors

4.1.1. Syntactic Interoperability Layer

The syntactic layer ensures that data exchanged among system components adheres to a well-defined structure and format. It emphasizes the standardization of communication protocols, message encoding, and data schema, thereby enabling different subsystems to consistently interpret and parse the exchanged information [Ribeiro et al. 2021].

Within CSIS, each surveillance camera is connected and authenticated in the system environment, establishing the technical foundation for seamless integration. Communication between the cameras and the system is managed through the Real-Time Publishing and Subscription (RTPS) protocol, as defined by the Internet Engineering Task Force (IETF) under Request for Comments (RFC) 2326 [Schulzrinne et al. 1998]. Depending on latency and reliability requirements, either the Transmission Control Protocol (TCP) or the User Datagram Protocol (UDP) is employed for multimedia stream transmission.

At the syntactic level, raw metadata from the streams is also collected, including camera identifiers, timestamps, frame structures, latency and frames per second (fps). These syntactic descriptors are critical for synchronization, traceability, and reliable data handling across CSIS services.

4.1.2. Semantic Interoperability Layer

The semantic layer assigns meaning to structured data received from the Syntactic Interoperability Layer. It interprets and enriches raw information by identifying relevant entities, patterns, and contexts, making data machine-understandable and enabling intelligent decision-making across heterogeneous systems [Ribeiro et al. 2021, Mane et al. 2021].

Our architecture implements the semantic layer by using different frameworks and strategies. A key task is to analyze video frames to identify meaningful objects and contextual elements within the monitored environment. To accomplish this task, our approach employs pre-trained object detection models based on the YOLOv11x architecture [Jocher and Qiu 2024] to perform semantic detection of entities such as fire, smoke, graffiti [Moura et al. 2021], suspicious behavior [Santos et al. 2024], car plates [Santos et al. 2023], and persons with knives or weapons.

Each detection includes contextual metadata, such as the event location and the specific camera identification. These elements contribute to a structured understanding of the monitored scene, and they are crucial for generating meaningful alerts [Gondim et al. 2025]. The semantic layer ensures that data transmitted between components carries a shared, machine-readable meaning, which is essential for interoperability across heterogeneous environments.

4.1.3. Pragmatic Interoperability Layer

At the pragmatic layer, the system interprets object detections and contextual information within predefined intention rules, focusing on achieving goals. Rules are designed to reflect the intended functional behavior of the surveillance system by aligning system actions to achieve their goals.

The primary mechanism supporting this interpretation is the Event-Condition-Action (ECA) model [Dube et al. 2002, Paschke 2006], which enables the system to automatically trigger appropriate responses to specific events detected in the environment. However, due to its simplicity and effectiveness, the ECA model is considered suitable for our approach.

In the context of license plate recognition [Santos et al. 2023], as illustrated in Figure 2, the producer system performs Automatic Number Plate Recognition (ANPR) combined with Paddle Optical Character Recognition (OCR) to detect vehicle plates, verify their registration status, and compute a confidence score for the recognition. Once the plate is successfully identified, the system triggers the *traffic.plate_recognized* event. On the consumer side, the access-control system interprets this event by applying ECA rules that reflect pragmatic intentions, evaluating the incoming event based on criteria such as the vehicle's authorization status, its frequency of occurrence within a defined pe-

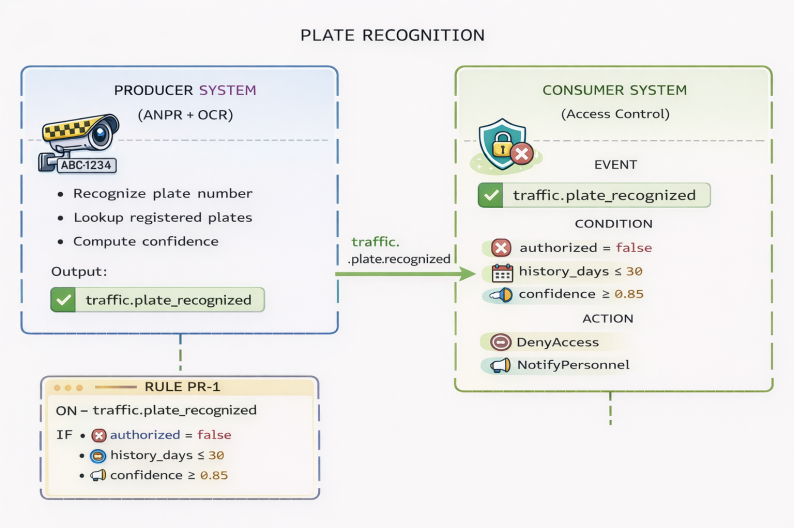

Figure 2. ECA rules for license plate inspection.

riod (e.g., the past 30 days), and whether the recognition confidence exceeds a predefined threshold. When these conditions are met, the system may take actions such as denying access or notifying security personnel. This example highlights how pragmatic interpretation translates low-level recognition outputs into high-level access-control decisions.

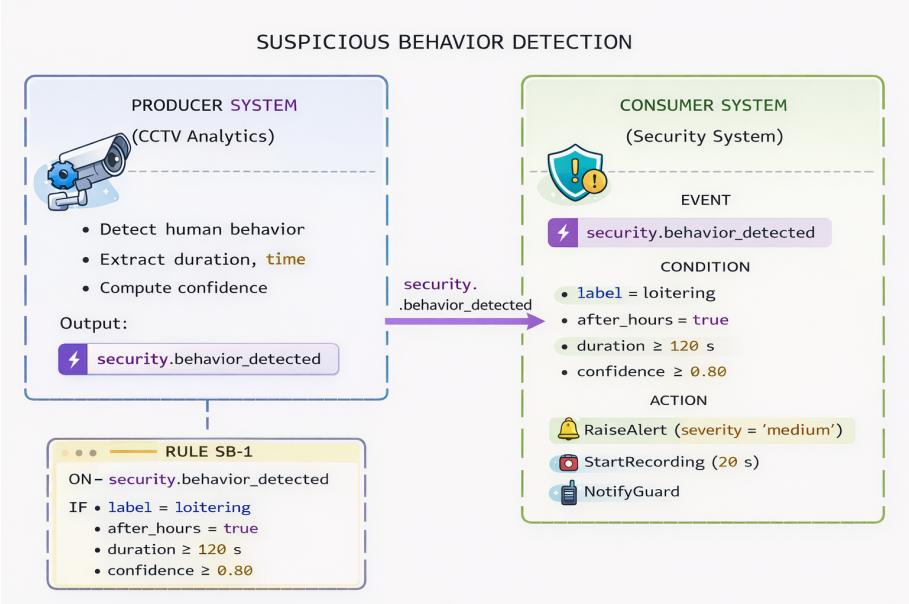

Figure 3. ECA rules for suspicious behavior detection.

In the suspicious behavior detection scenario [Santos et al. 2024], as illustrated in Figure 3, the producer system analyzes CCTV video streams to detect human behaviors, extract temporal attributes such as duration and time of occurrence, and compute a confidence measure. When potentially suspicious activity is detected, the *event security.behavior_detected* is generated. The consumer security system interprets this event by applying ECA rules that reflect pragmatic intentions, such as identifying loitering behavior after hours with sufficient duration and confidence. If these conditions are met,

the system initiates actions including raising an alert with an appropriate severity level, starting video recording, and notifying security personnel. This scenario illustrates how contextual and temporal information is pragmatically interpreted to enable timely and goal-driven security responses.

Collectively, these scenarios demonstrate how ECA rules enable pragmatic interoperability by linking perception-level events to intention-oriented actions, thereby ensuring that heterogeneous producer and consumer systems operate coherently to achieve shared surveillance objectives.

4.1.4. Hub of Events

The hub of events are a structured data consisting of the detection object, its context, and the corresponding action which sent to the Event Bus API (event producer). The **Hub of Events** module handles the entire event processing life cycle.

The **Monitor Service** accesses event data in real time via the WebSocket protocol and forwards this to the **Update Service**. The Event Bus maintains records of both detected events and their associated actions. The Update Service component receives the information over HTTP, stores it in a database, and republishes it to other components such as dashboards and mobile devices used by security personnel. To optimize performance, a memory cache stores redundant information.

4.1.5. User Interface

The user interface is composed of two primary access roles: the **Security Team** and the **Security Manager** or **Dashboard Operator** as illustrated in Figure 4. These users interact with the system through a centralized dashboard designed to display alerts, events, and analytical data in real time.

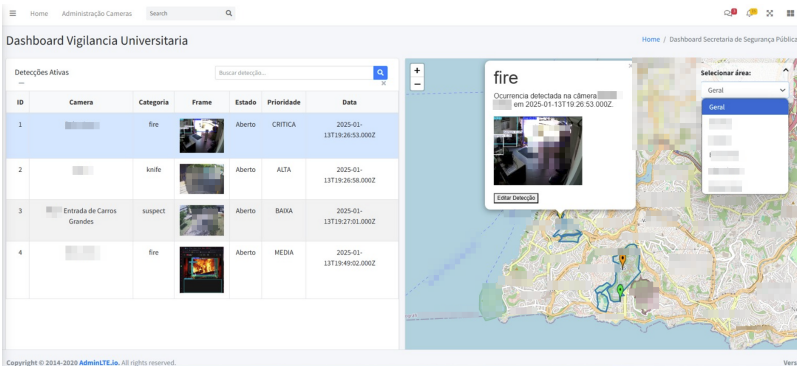

Figure 4. CSIS Dashboard

Detections are published in real time to a centralized dashboard that presents geolocated alerts, associated confidence levels, and annotated video snapshots. This interface enables security personnel to visualize, prioritize, and efficiently respond to critical events.

Security managers can review past events, generate reports, adjust rule parameters, and supervise the surveillance process. Meanwhile, field agents receive mobile alerts with contextual information, enabling rapid response to detected incidents. The system also includes a module for user registration and authentication, ensuring that only authorized personnel can interact with sensitive monitoring data. The interface is also flexible to accommodate future extensions to support mobile devices.

5. Experiments and Results at Security Coordination Unit of a federal university

To evaluate the CSIS performance in a real-world context, experiments are conducted at the Security Coordination Unit of a federal university. The aim is to evaluate the models' detection efficacy in unconstrained surveillance environments, specifically their capacity for low-latency and real-time event recognition. The experiments are conducted using video streams captured by surveillance cameras integrated into the system. Each scenario was selected based on real incidents recorded by a Federal University's Security Coordination Unit, with the aim of reproducing the system's practical operation under typical risk situations.

5.1. Datasets and Hyperparameter Setup

The quality, diversity, and volume of training data are crucial for developing robust, accurate, and generalizable object detection models. Selecting appropriate datasets directly influences a model's performance in real-world scenarios. Table 1 provides a detailed list of the training datasets used for each object detection model within our architecture.

Table 1. Datasets for specific detection tasks.

Task	Dataset	N
Car plate	[Laroca et al. 2022]	20,000
Fire and smoke	[Cazzolato et al. 2017]	9,448
weapon in hand	[Assalim 2024]	4,098
	[Lim et al. 2021]	5,500
Flood	[Aquarium 2023]	700
	[Bhutad and Patil 2021]	2,000
Graffiti and spray	[Moura et al. 2021]	516
	[Marcin 2023]	42

The variation in dataset sizes (for example, 20,000 for vehicle plates versus 42 for graffiti) implicitly reveals the challenges associated with acquiring large, high-quality, and well-annotated datasets for all possible campus security scenarios. This disparity in data volume often correlates with the model's performance across different tasks. The diverse selection of datasets is crucial for the robustness and versatility of the CSIS models, allowing them to operate effectively in a wide range of real-world scenarios. Although Table 1 lists datasets of different sizes, it is important to emphasize that these datasets are not used jointly in a single multi-class training process. Each detection task in the CSIS architecture is handled by an independent model, trained separately with its own

data and pretrained weights. For all models, the dataset was split approximately into 80% for training, 10% for validation, and 10% for testing.

The evaluation metrics employed to assess the performance of the YOLOv11x model across its various detection tasks are standard in object detection research and provide a comprehensive assessment of the model's effectiveness: Precision (P) reflects the model's accuracy in identifying relevant objects, Recall (R) reflects the model's sensitivity in detecting all relevant objects, and Mean Average Precision (mAP) provides an overall evaluation of detection quality across all classes and Intersection over Union (IoU) thresholds. Specifically, mAP50 indicates that the average precision is calculated at an IoU threshold of 0.5, a common benchmark in object detection [Silva et al. 2025].

Hyperparameter Setup. The models for detecting fire and smoke, firearms or bladed weapons, graffiti, and flooding were trained using the YOLOv1.1 architecture from the Ultralytics library, with input images resized to 640 × 640 pixels. This resolution offers a balance between spatial accuracy and computational cost for near real-time object detection. Training was conducted with a batch size of 16 and the Stochastic Gradient Descent (SGD) optimizer, using an initial learning rate of 0.01, momentum of 0.937, and weight decay of 0.0005, ensuring stable convergence and robust performance, as depicted in Table 2. An early stopping strategy was applied after 10 epochs without performance improvement to mitigate overfitting. All experiments were trained on *CMCAD Fabiola-Greve* on an A100 GPU with 80 GB of RAM.

Table 2. Training setup and essential hyperparameters (YOLOv11x)

Category	Setting / Value
Model	YOLOv11xx
Input image size	640×640
Batch size	16
Early stopping	10 epochs
Optimizer	Stochastic Gradient Descent (SGD)
Initial learning rate	0.01
Momentum	0.937
Weight decay	0.0005
Loss function	Default YOLOv11x detection loss
Data augmentation	Mosaic, flip, color augmentation, Auto Augment

5.2. Campus scenario

Experiments were carried out on training real-time camera streams, enabling rapid and accurate identification and supporting the immediate generation of alerts or automated responses to ensure timely intervention. This scenario comprises seven surveillance cameras distributed throughout the university campus. Among these cameras, one positioned at the university's main entrance is linked to a model dedicated to performing license plate recognition for automobiles, motorcycles, and buses. Another camera, located in one of the university's main parking areas, is connected to a model specialized in detecting suspicious behavior. The remaining five cameras are configured with task-specific

deep learning models for graffiti detection, fire and smoke detection, flood detection, and the identification of individuals carrying firearms or bladed weapons.

5.2.1. Semantic interoperability

In the context of image analysis, semantic interoperability is achieved when different systems can exchange and interpret image-related information based on a shared understanding of meaning. CSIS incorporates specialized models for vehicle license plate recognition, enabling efficient, real-time identification and tracking. Figure 5 illustrates an example of automatic license plate recognition[1].

Figure 5. Real-time vehicle license plate recognition (anonymized data).

After plate detection, we implemented a text recognition model (PaddleOCR) and achieved 96.79% accuracy, demonstrating strong performance in extracting text across different lighting and visibility conditions. This result validates the effectiveness of license plate recognition. Whenever the system detects vehicle license plates not registered in the university database, an alert is triggered. Figure 6 illustrates the detection of the flooding model on the campus, specifically designed to assist in managing natural disasters within the university environment.

Figure 6. Flood detection

Figure 7. Graffiti and spray detection.

Figure 7 depicts graffiti detection. The detection model identified the act of graffiti in progress, correctly classifying the event and triggering the corresponding alert. This

[1]For anonymization purposes, the license plate has been removed from the original image.

type of task is essential for asset preservation and enabling rapid intervention by security teams, thereby helping preserve patrimony and ensuring timely responses to acts of vandalism.

In addition, we have implemented fire and smoke detection systems, which analyze camera feeds and environmental data to detect signs of fire or smoke at the earliest stages. These models help to ensure rapid responses to potential fire incidents, minimizing the risks to life and property within monitored areas.

Furthermore, our models for detecting firearms and bladed weapons enhance surveillance capabilities, enabling the identification of dangerous objects in real time. This technology helps prevent violent incidents by enabling security personnel to respond proactively before threats escalate.

5.2.2. Pragmatic interoperability

In the context of image analysis, pragmatic interoperability is achieved when different systems can exchange and interpret messages through image-related information with a shared understanding of both context and intention.

The interpretation of suspicious behavior in monitored environments, such as parking lots, depends on the situational context in which the action occurs. Although computer vision systems can identify similar patterns of movement or posture, the pragmatic meaning of these actions varies according to contextual factors such as location, object of interest, and the presumed purpose of the observed agent.

Our model for detecting suspicious behavior in parking areas considers an individual who is crouching or seated on a bench in close proximity to a vehicle for approximately one minute as potentially exhibiting suspicious behavior. Such conduct may be associated with attempts to damage the vehicle, remove personal property, or gain unauthorized access. In this context, the combination of spatial proximity, body posture, and the duration of the action provides contextual indicators suggestive of a possible intention to engage in unlawful interaction. Although the duration of the stay is similar, the relation between the action and the environment differs. A single semantic analysis of the image based on objects, positions, and movements is insufficient to determine the nature of the action.

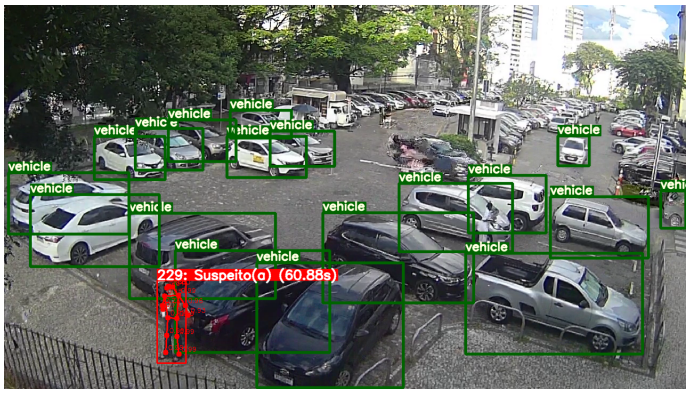

Figure 8. Detection of suspicious behavior for a person standing near a vehicle.

Figure 8 illustrates the detection of suspicious behavior in a parking area. This type of detection is particularly relevant for preventive alerts and for monitoring potential risk situations. For the detection of suspicious behavior, we employed the YOLOv11x-Pose model, which provides native human keypoints and integrates seamlessly with the model track function, enabling the system to analyze posture and temporal persistence using tracked identities across frames.

5.2.3. Quantitative Results

Table 3 shows the performance models across several datasets, evaluating ***P**recision*, ***R**ecall*, and ***m**ean **A**verage **P**recision*. Precision indicates the model's accuracy in identifying relevant objects; recall reflects its sensitivity in detecting all relevant objects; and mAP provides an overall evaluation of detection quality.

Table 3. Precision (P), recall (R), and mean Average Precision (mAP) results for YOLOv11xx model in each dataset.

Models	P	R	mAP50
Fire and smoke	0.71	0.52	0.57
Weapon in hand	0.90	0.58	0.75
Flood	0.49	0.30	0.32
Graffiti and spray	0.76	0.71	0.76
Plate	0.99	0.99	0.99

Our results demonstrate that the models can perform detection tasks with satisfactory precision and robustness across diverse scenarios. Each specialized detector (fire, weapons, floods, graffiti, spray, and license plates) contributes to an integrated and efficient surveillance framework. These findings validate the practical applicability of CSIS and support its deployment in real-world environments. We also presented experiments conducted at the Security Coordination Unit of a Federal University, illustrating the models' performance under real monitoring conditions.

Although some tasks, such as the flood detection scenario, still require dataset expansion, the overall performance remained consistent. The results reinforce the system's potential to enhance the Security unit's operational capacity and optimize incident response through automated alerts and integrated analyses.

6. Impacts and Limitations

6.1. Social Impacts

The implementation on the university campus has generated impacts for students, faculty, staff, and visitors. Our approach improves the safety of individuals operating, helping prevent threats of theft and assault. A real-time automated monitoring enhances confidence and well-being by automatically detecting incidents, reinforcing a sense of security and control over the campus environment. The sense of security provided by CSIS allows students and staff to use the university's public spaces with less concern than previously.

CSIS's ability to detect fires, floods, or suspicious behavior at an early stage reduces students' and staff's exposure to risks and helps protect their physical integrity. The agility in responding to incidents, with automatic alerts directed to security agents (property security or the fire brigade), reduces response time and minimizes the impact of critical situations.

Automated threat detection and classification reduce operators' cognitive load and enhance overall operational efficiency.

CSIS adheres to ethical standards, ensuring the privacy of individuals involved in the monitoring process and complying with the regulations of the General Data Protection Law (LGPD).

6.2. Organizational Impacts

Implementing our approach in a university environment entails substantial organizational transformation. The alerts and events generated by CSIS are transmitted via a centralized dashboard housed within the Security Unit. Such a dashboard enables supervisors and operators to visualize incidents in real time and coordinate appropriate interventions. By consolidating event data into a single platform, the dashboard mitigates information overload and enhances situational awareness, thereby facilitating timely and evidence-based decision-making in critical security scenarios.

At the organizational level, the dashboard impacts by fostering integration across university sectors, reducing data fragmentation, and improving communication among operators, supervisors, and security managers. Consequently, institutional workflows are reinforced, response procedures are standardized, and decision-making processes become more agile and evidence-driven, ultimately minimizing response times during critical incidents.

Moreover, the dashboard's systematic event recording enables the Security Unit to identify recurring patterns, assess structural vulnerabilities, and implement preventive measures. In this way, CSIS generates organizational intelligence that strengthens institutional governance and enhances the overall resilience of the university's security infrastructure.

6.3. Technical Impacts

CSIS has provided technical advances in a university surveillance environment. It is structured around three interoperability layers (syntactic, semantic, and pragmatic) that together enable seamless integration of heterogeneous surveillance infrastructures. Devices are registered and authenticated within the system, ensuring the synchronized collection of essential metadata, such as frame rate and latency, for accurate frame description at the *syntactic level*. At the *semantic layer*, incident detection in the syntactic layer is performed using computer vision models optimized for real-time operation, thereby ensuring both accuracy and timeliness in event recognition. Suspicious incidents and their contextual information are processed at the *pragmatic layer*, where they are evaluated against predefined conditions. Only after this analysis is the event conditioned on the *Hub Of Events* module for subsequent handling. Once in the Hub Of Events module, the event is interpreted as a representation of the underlying intention associated with the detected

incident. The *Events Hub* module is responsible for managing duplicate events, storing them in a database, and forwarding validated records to the Dashboard.

Our ecosystem provides an ensemble-based approach to incident detection. Rather than relying on a single model, the system employs multiple YOLOv11x models, each specialized in recognizing a particular class of threat or incident. This modular and scalable architecture not only allows models to be implemented or removed as needed but also enhances detection coverage, improves accuracy, and increases robustness in real-world deployments.

6.4. Limitations

Although the CSIS ecosystem helps address practical issues, several limitations remain.

The performance of detection models is influenced by the availability and balance of training datasets. For example, license plate recognition tasks benefit from extensive datasets, while flood and graffiti detection scenarios are constrained by limited data. This leads to reduced detection performance and limits the models' generalization capacity. Expanding and diversifying datasets for underrepresented threats is essential to enhance model stability and reliability.

The experimental evaluation was limited to a federal university, using a limited number of cameras and context-specific institutional procedures. Although the scenarios represent authentic operational conditions, the findings are not directly generalizable to other campuses or corporate environments with differing infrastructures, security protocols, or spatial configurations. Additional deployments in multi-campus or heterogeneous institutional contexts are necessary to evaluate scalability and external validity.

Pragmatic interoperability in CSIS is implemented using Event–Condition–Action (ECA) rules, which, while transparent and effective, impose limitations. The rule-based approach depends on manually defined conditions and thresholds, making it less adaptable to new suspicious behaviors, ambiguous situations, or conflicting interpretations of context and intention. As a consequence, the current solution could struggle with more complex or emergent situations that require higher-level reasoning or dynamic interpretation of intentions.

Additionally, the system architecture is closely integrated with the YOLOv11x family of models, chosen for their maturity and real-time performance on the available hardware. Although this selection is operationally justified, the absence of a comparative evaluation with alternative detectors limits conclusions about detector-agnostic performance. Future research should investigate the effects of various vision architectures on both semantic and pragmatic interoperability.

An additional aspect concerns the operational analysis of false positives and false negatives. Although standard detection metrics are presented, the study does not quantitatively assess the downstream impact of undetected events resulting from operator workload, alert fatigue, or decision-making processes. A more comprehensive, human-centered assessment would yield further insights into long-term system adoption and user trust.

Finally, while privacy protection and compliance with data protection regulations are considered conceptually, the study does not incorporate a formal privacy impact as-

sessment or user perception analysis. Given that surveillance systems operate in sensitive social environments, further research is needed to assess ethical considerations, transparency, and acceptance across diverse stakeholder groups.

7. Conclusions and Future Work

Our CSIS approach contributes to a university surveillance and security detection environment. Designed for multi-level interoperability (syntactic, semantic, and pragmatic), it embodies an understanding of contemporary security challenges within a complex sociotechnical ecosystem. Its capability to detect and identify a wide range of threats, including weapons, suspicious behavior, fires, floods, and vandalism, in real time demonstrates its foreseen applicability.

From an operational perspective, CSIS has an innovative impact on the Security Unit's daily activities. A centralized security dashboard improves through an automated monitoring, from a predominantly manual workflow to a data-driven, proactive decision-support system. Operators now receive real-time, automated alerts enriched with contextual data, enabling faster and more accurate interventions. Supervisors, in turn, benefit from comprehensive, consolidated event logs, which streamline post-incident analysis and support evidence-based decision-making. These advancements reduce operator fatigue, which is a common challenge in continuous camera monitoring, while enhancing coordination among surveillance operators, field agents, and emergency responders.

For future work, we envisage directions such as expanding and diversifying the dataset and extending to multi-campi environments.

Acknowledgement

We acknowledge the use of an AI-based generative tool for English translation and language correction. This work was partially supported by FAPESB through grants TIC 0002/2015 and CCE 0022/2023. Authors also acknowledge the support from the Coordination of Security Unit and *CMCAD FabiolaGreve* for enabling the training models.

References

Anagnostopoulos, T., Kostakos, P., Zaslavsky, A., Kantzavelou, I., Tsotsolas, N., Salmon, I., Morley, J., and Harle, R. (2021). Challenges and solutions of surveillance systems in iot-enabled smart campus: a survey. *Ieee Access*, 9:131926–131954.

Aquarium (2023). hanyang-puddle-detection-part2 bbox dataset. `https://universe.roboflow.com/aquarium-nfk5n/hanyang-puddle-detection-part2-bbox`. visited on 2024-11-26.

Assalim, J. (2024). Weapon 2 dataset. `https://universe.roboflow.com/joao-assalim-xmovq/weapon-2`. visited on 2024-11-26.

Bhutad, S. and Patil, P. K. (2021). Stagnant water. DOI: `https://dx.doi.org/10.21227/zrhk-tq74`.

Cazzolato, M. T., Avalhais, L. P. S., Chino, D. Y. T., Ramos, J. S., Souza, J. A., Rodrigues-Jr, J. F., and Traina, A. J. M. (2017). Fismo: A compilation of datasets from emergency

situations for fire and smoke analysis. In *SBBD2017 - SBBD Proceedings of Satellite Events of the 32nd Brazilian Symposium on Databases - DSW (Dataset Showcase Workshop)*, pages 213–223. SBC.

de Andrade, A. F., Damasceno, E. Z. C., de Lima Siqueira, J. E., Rocha, A. M. S. A., Carrafa, J. P. P., dos Santos, L. B., and de Jesus, R. F. (2024). A implementação de sistemas de vigilância com tecnologia de inteligência artificial no auxílio da administração pública na prevenção de crimes. *Revista Eletrônica Interdisciplinar*, 16(3).

Delnevo, G., Ghini, V., Fiumana, E., and Mirri, S. (2024). A support tool for emergency management in smart campuses: Reference architecture and enhanced web user interfaces. *Sensors*, 24(18).

Dube, K., Wu, B., and Grimson, J. (2002). Using ECA Rules in Database Systems to Support Clinical Protocols. In *Proceedings of the 13th International Conference on Database and Expert Systems Applications (DEXA 2002)*, volume 2453 of *Lecture Notes in Computer Science*, pages 226–235. Springer.

Ekpoh, U. I., Edet, A. O., and Ukpong, N. N. (2020). Security challenges in universities: Implications for safe school environment. *Journal of Educational and Social Research*, 10(6):112–112.

Figueiredo, R., Mane, B., Clavijo, D. C. R., Silva, S., Moraes, F., Nery, C., Magalhaes, A. P., Maciel, R. S. P., Claro, D., Souza, M., Lima, G., and Leite, J. C. (2025). Towards a SoIS model for University Surveillance. In *Proceedings of the 2025 International Conference on Software Engineering for Systems-of-Systems (SESoS)*, ICSE 2025, page –, Ottawa, Ontario, Canada. ACM.

Gondim, J., Claro, D. B., and Souza, M. (2025). A bilingual analysis of multi-head attention mechanism for image captioning based on morphosyntactic information. *Journal of the Brazilian Computer Society*, 31(1):1063–1076.

Jocher, G. and Qiu, J. (2024). Ultralytics yolo11.

Keerthana, R., Kumar, M., and Prakash, S. (2023). Enhancing campus security through smart surveillance system. In *Proceedings of the International Conference on Smart Technologies*, pages 112–118. IEEE. Accessed: 2025-07-28.

Kerich, N., Omuterema, S. O., and Pepela, M. M. (2024). Evaluating the role of electronic security surveillance in enhancing safety in secondary schools: A case study of trans nzoia county, kenya. *African Journal of Empirical Research*, 5(4):163–173.

Laroca, R., Cardoso, E. V., Lucio, D. R., Estevam, V., and Menotti, D. (2022). On the cross-dataset generalization in license plate recognition. In *International Conference on Computer Vision Theory and Applications (VISAPP)*, pages 166–178.

Laroca, R., Zanlorensi, L. A., Gonçalves, G. R., Todt, E., Schwartz, W. R., and Menotti, D. (2021). An efficient and layout-independent automatic license plate recognition system based on the yolo detector. *IET Intelligent Transport Systems*, 15(4):483–503.

Lim, J., Al Jobayer, M. I., Baskaran, V. M., Lim, J. M., See, J., and Wong, K. (2021). Deep multi-level feature pyramids: Application for non-canonical firearm detection in video surveillance. *Engineering Applications of Artificial Intelligence*, 97:104094.

Maciel, R. S. P., David, J. M. N., Claro, D. B., and Braga, R. (2019). Full interoperability: Challenges and opportunities for future information systems.

Mane, B., Magalhaes, A. P., Quinteiro, G., Maciel, R. S. P., and Claro, D. B. (2021). A domain specific language to provide middleware for interoperability among saas and daas/dbaas through a metamodel approach. In *Proceedings of the 23rd International Conference on Enterprise Information Systems - Volume 1: ICEIS*, pages 83–94. INSTICC, SciTePress.

Marcin, W. (2023). Grafitti finding dataset. `https://universe.roboflow.com/marcin-w-5a1uk/grafitti-finding`. visited on 2024-11-26.

Moura, N., Gondim, J., Claro, D., Souza, M., and Figueiredo, R. (2021). Detection of weapon possession and fire in public safety surveillance cameras. In *Anais do XVIII Encontro Nacional de Inteligência Artificial e Computacional*, pages 290–301, Porto Alegre, RS, Brasil. SBC.

Paschke, A. (2006). ECA-RuleML: An Approach Combining ECA Rules with Temporal Interval-Based KR Event/Action Logics and Transactional Update Logics. CoRR, arXiv:cs/0610167. Accessed: 2025-07-08.

Ralha, C. G., Claro, D. B., and Maciel, R. S. P. (2025a). Conceiving socio-technical information systems from the perspective of digital twins. In *Proceedings of the Workshop on Information Systems Challenges in Brazil (2026-2036) at the Brazilian Symposium of Information Systems*.

Ralha, C. G., Claro, D. B., and Stroele, V. (2025b). Information systems and artificial intelligence integration challenges. In *Proceedings of the Workshop on Information Systems Challenges in Brazil (2026-2036) at the Brazilian Symposium of Information Systems*.

Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 779–788.

Ribeiro, E. L. F., de Jesus, L. E. N., Claro, D. B., and Moura, N. (2021). Towards a pragmatic interoperability on the midas middleware. In *Proceedings of the Brazilian Symposium on Multimedia and the Web*, WebMedia '21, page 161–168, New York, NY, USA. Association for Computing Machinery.

Ribeiro, E. L. F., Monteiro, E. L., Claro, D. B., and Maciel, R. S. P. (2019). A conceptual framework for pragmatic interoperability. In *Proceedings of the XV Brazilian Symposium on Information Systems*, pages 1–8.

Santos, C. N. M., Claro, D. B., Gondim, J. M., and Mane, B. (2024). Suspicious behavior detection near vehicles in university environment: An approach using object detection and body angles. In *Anais do XX Simpósio Brasileiro de Sistemas de Informação (SBSI)*, pages 1–8. Sociedade Brasileira de Computação.

Santos, D., Claro, D. B., and Gondim, J. (2023). Monitoring vehicle plate detection in brazilian universities. In *Proceedings of the 19th Brazilian Symposium on Information Systems (SBSI)*. Sociedade Brasileira de Computação.

Schulzrinne, H., Rao, A., and Lanphier, R. (1998). RFC 2326: Real Time Streaming Protocol (RTSP). `https://www.rfc-editor.org/rfc/rfc2326.html`. Accessed: November 1, 2024.

Shiri, A. (2024). Criminal policy about electronic surveillance. *Iranian Journal of Public Policy*, 10(3):99–111.

Silva, A. S., de Azevedo, A. R., Neto, F. H. d. A. M., and da Silva, P. H. F. (2025). Modelo basado en yolov8 para la detección automática de daños en tejados residenciales. *Revista ALCONPAT*, 15(1):50–63.

Tantra, B. J. and Widjaja, M. (2024). Automatic detection of dress-code surveillance in a university using yolo algorithm. *IAES International Journal of Artificial Intelligence*, 14(2):1568–1575.

Tian, Y., Ye, Q., and Doermann, D. (2025). Yolov12: Attention-centric real-time object detectors. *arXiv preprint arXiv:2502.12524*.

Wang, C.-Y. and Liao, H.-Y. M. (2024). Yolov9: Learning what you want to learn using programmable gradient information.

Xiao, Q., Wu, C., Wang, Y., Lin, B., and Zhang, H. (2024). Campus security inspection robot based on yolo algorithm. In *Proceedings of the 2024 2nd International Conference on Frontiers of Intelligent Manufacturing and Automation*, pages 800–806.

Zhou, Y. et al. (2022). A model for predicting the risk of campus violence in an edge intelligent computing architecture. *Journal of Intelligent Fuzzy Systems*, 43(4):4353–4362.