Sensors & Transducers, Vol. 272, Issue 1, April 2026, pp. 59-68

www.sensorsportal.com

Sensors & Transducers

Published by IFSA Publishing, S. L., 2026
http://www.sensorsportal.com

A Security-Aware Edge Artificial Intelligence Framework for Robust Autonomous Decision-Making in Unmanned Aerial Systems

Sérgio SILVA
Department of Communication Sciences and Information Technologies,
University of Maia, 4475-690 Maia, Portugal
E-mail: D012196@umaia.pt

Received:26 Dec. 2025 /Revised: 31 March 2026 /Accepted: 20 April 2026 /Published: 28 April 2026

Abstract: This article presents a security-aware Edge Artificial Intelligence framework designed to enhance autonomous decision-making in Unmanned Aerial Systems. As aerial vehicles increasingly rely on onboard artificial intelligence to interpret sensor data, plan flight trajectories, and respond to environmental conditions, they are exposed to cyber-physical threats such as data manipulation, Global Navigation Satellite System (GNSS) spoofing, and adversarial visual perturbations. The proposed framework integrates lightweight deep learning models optimized for embedded processors with a multilayer cybersecurity architecture that includes real-time integrity verification of sensor streams, adversarial robustness modules, and secure decision-validation routines. Experimental evaluation on a quadrotor platform demonstrates that the system preserves autonomy performance while significantly reducing the risk of unsafe commands under adversarial conditions, achieving over an 85 % reduction in unsafe actions with minimal false alarms and negligible impact on mission continuity. These findings highlight the importance of combining Edge Artificial Intelligence with embedded security mechanisms to ensure the resilience, safety, and reliability of autonomous unmanned systems operating in contested or communication-limited environments.

Keywords: Edge artificial intelligence, Unmanned aerial systems, UAV cybersecurity, Autonomous decision-making, Embedded AI, Adversarial robustness, Sensor integrity monitoring.

1. Introduction

Autonomous Unmanned Aerial Systems (UAS) have advanced rapidly due to the integration of onboard Artificial Intelligence (AI), enabling complex missions with limited human intervention. These systems rely on digital sensors, communication links, and machine learning models to interpret environmental data, plan trajectories, and execute decisions in real-time. However, this increasing reliance on digital intelligence exposes unmanned systems to a range of cyber-physical threats, including data manipulation, GNSS spoofing, and adversarial perturbations in visual sensors. Such attacks can compromise decision-making, reduce operational safety, and threaten mission success.

Recent research has increasingly focused on improving the security and reliability of autonomous Unmanned Aerial Systems (UAS) through the integration of AI and advanced anomaly detection techniques. Several studies have explored the use of Machine Learning models to detect cyber-physical threats affecting aerial platforms. For example, AI-based frameworks have been proposed to detect cyber-physical attacks in Internet-of-Things enabled drones by monitoring system behavior and identifying abnormal patterns in sensor data [1].

Other works have specifically addressed GNSS spoofing attacks, which represent one of the most critical threats to autonomous navigation. Machine learning and adversarial learning techniques have been investigated to identify spoofing patterns and mitigate

navigation errors in Unmanned Aerial Vehicles (UAV) [2, 3].

Recent approaches also explore ensemble learning and deep learning methods for detecting navigation signal manipulation and improving the robustness of positioning systems in aerial platforms [4].

In addition, research has been conducted on explainable reinforcement learning methods capable of detecting adversarial attacks affecting guidance and path planning modules [5].

Experimental studies have also demonstrated the feasibility of detecting spoofing attacks through sensor-fusion analysis and recovery mechanisms based on consistency verification between inertial and navigation data [6-8].

Recent advances in autonomous aerial systems have highlighted the importance of integrating AI with cybersecurity mechanisms to ensure safe operation in dynamic environments [9].

Several studies have investigated the use of deep learning models for real-time perception and navigation in unmanned aerial vehicles, demonstrating significant improvements in obstacle detection and environment understanding [10, 11].

However, the integration of such perception systems with security-aware mechanisms remains a relatively unexplored research area. In particular, the challenge of ensuring trustworthy autonomous decision-making under adversarial conditions has motivated the development of embedded security frameworks capable of detecting sensor anomalies, verifying data integrity, and validating navigation commands before execution [12, 13].

Despite these advances, many existing solutions focus primarily on individual attack detection mechanisms and often rely on centralized processing or external monitoring systems. This can limit their applicability in real-time autonomous platforms operating with constrained computational resources or in communication-limited environments [14, 15]. Therefore, there remains a need for integrated frameworks that combine embedded AI perception with real-time cybersecurity mechanisms capable of validating both sensor data and autonomous decisions directly on the onboard computing platform [9, 16].

1.1. Extension over Previous DAUS 2026 Conference Paper

This article extends previous work presented at the DAUS 2026 conference [17]. While the conference version introduced the initial concept of a secure Edge AI framework, the present manuscript provides significant extensions. These include enhanced real-time security layers, a more detailed and formalized system architecture, and expanded experimental evaluation under multiple adversarial scenarios. In addition, this work introduces improved implementation details and formally defined evaluation metrics, strengthening reproducibility and scientific rigor.

The key differences between the conference and journal versions are summarized in Table 1.

Table 1. Conference versus Journal versions.

Component	DAUS 2026	Journal article
Architecture	Conceptual design	Full multi-layer implementation
EKF Monitoring	Basic description	Formal EKF with statistical detection
Perception Model	Preliminary YOLOv5	Optimized and embedded deployment
Experiments	Limited validation	Multiple adversarial scenarios
Evaluation Metrics	Partial	Comprehensive and formally defined
Reproducibility	Limited details	Full implementation detail

The main contributions of this work can be summarized as follows:

- A security-aware Edge AI architecture integrating perception, integrity monitoring, and decision validation for autonomous UAS;
- A real-time sensor integrity monitoring approach based on sensor fusion and anomaly detection techniques;
- A secure decision validation mechanism that prevents unsafe or compromised commands from being executed;
- An embedded implementation optimized for resource-constrained UAV platforms;
- Experimental validation under multiple cyber-physical attack scenarios, demonstrating improved system resilience and safety.

The main novelty of this work lies in the integration of lightweight perception models, sensor integrity monitoring, and secure decision validation within a unified Edge AI architecture capable of operating fully onboard resource-constrained UAV platforms.

The remainder of this article is organized as follows. Section 2 presents the proposed methodology. Section 3 describes the system architecture and implementation. Section 4 provides experimental results and evaluation. Section 5 discusses the findings and limitations, and Section 6 concludes the article and outlines future work.

2. Methodology

The proposed framework integrates secure Edge AI capabilities into the autonomous decision-making pipeline of UAS. As illustrated in Fig. 1, the methodology consists of three main components: (i) onboard perception and inference, (ii) cybersecurity and data integrity verification, and (iii) secure decision validation. All components are

optimized to operate in real-time on embedded hardware with limited computational resources, ensuring reliable and safe autonomy even under adversarial conditions.

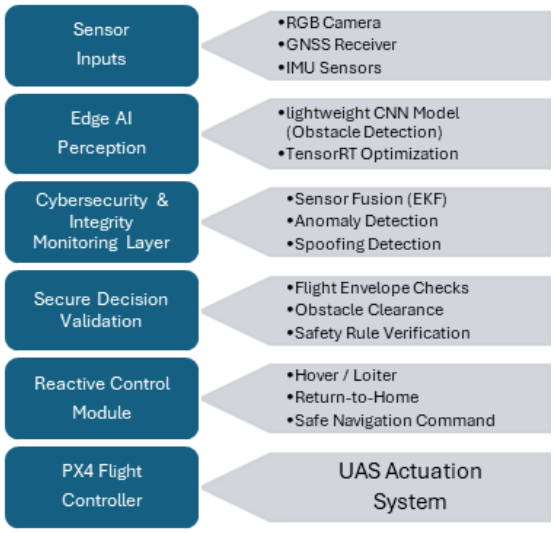

Fig. 1. Architecture of secure Edge AI framework for autonomous UAS.

Fig. 1 illustrates the overall architecture of the proposed secure Edge AI framework. Sensor data from the onboard camera, global navigation satellite receiver, and inertial measurement unit are first processed by the perception and sensor fusion modules.

The perception module performs obstacle detection using a lightweight Convolutional Neural Network (CNN), while the integrity monitoring layer evaluates the consistency of navigation and inertial data. The perception layer processes visual information acquired from the onboard camera using and extracts environmental features and identifies potential obstacles in real-time.

Detected anomalies are forwarded to the decision validation module, which enforces safety constraints before navigation commands are transmitted to the flight controller.

The sensor fusion and integrity monitoring layer combines data from the GNSS receiver and inertial measurement unit through an Extended Kalman Filter (EKF). By evaluating the innovation residual between predicted and observed measurements, the system can identify inconsistencies that may indicate sensor spoofing or data manipulation attacks.

The final layer performs secure decision validation by enforcing operational safety constraints on the navigation commands generated by the autonomous control system. If anomalies are detected or safety conditions are violated, the system triggers fallback behaviors such as hover or return-to-home maneuvers.

This layered architecture provides multiple levels of protection against cyber-physical threats while maintaining real-time operation on resource-constrained embedded platforms.

2.1. Onboard Perception and Edge AI Inference

The onboard perception module performs obstacle detection and free-space assessment from monocular RGB imagery. A compressed CNN, based on Nano YOLOv5 and optimized for embedded deployment, processes 416×416 images to output obstacle bounding boxes with confidence scores. Training combines the VisDrone dataset with in-house outdoor flight data using augmentation techniques such as brightness variation, Gaussian noise, and motion blur.

Real-time performance on an NVIDIA Jetson Nano is achieved through post-training INT8 quantization, structured channel pruning, and TensorRT optimization. The model achieves a mean average precision (mAP@0.5) of 71.4 %, precision of 0.76, recall of 0.69, and an inference rate of 18–22 frames per second, depending on scene complexity. These results demonstrate that lightweight deep learning models can provide accurate perception while meeting embedded hardware constraints.

The training dataset combines publicly available aerial imagery from the VisDrone dataset with additional flight recordings captured during outdoor test missions. The combined dataset includes urban, semi-urban, and natural environments, providing diverse visual conditions for training the perception model. Data augmentation techniques were applied to simulate variations in illumination, motion blur, and sensor noise. These augmentations improve the robustness of the perception model when operating under challenging environmental conditions commonly encountered during real-world UAV deployments.

To improve reproducibility, the implementation details of the perception model are specified as follows. The Nano YOLOv5 model corresponds to a compressed variant of YOLOv5n adapted for embedded deployment. The training dataset consists of approximately 18,500 annotated images, combining the VisDrone dataset (70 %) and in-house UAV flight data (30 %). The dataset was divided into training (70 %), validation (20 %), and test (10 %) subsets.

Training was performed for 150 epochs using the Adam optimizer with an initial learning rate of 0.001 and cosine decay scheduling. A batch size of 16 was used during training. Data augmentation included brightness variation, Gaussian noise, motion blur, and horizontal flipping.

During inference, a confidence threshold of 0.4 and a non-maximum suppression threshold of 0.5 were applied. Deployment on the NVIDIA Jetson Nano was achieved using TensorRT with INT8 quantization calibrated on a subset of 1,000 representative images. Structured pruning reduced the model size by

approximately 35 % while preserving detection accuracy.

These details enable reproducibility and allow fair comparison with other embedded perception approaches.

2.2. Cybersecurity and Data Integrity Layer

A dedicated integrity-monitoring layer evaluates the consistency of sensor data from GNSS, inertial measurement units, and vision-based estimates using an EKF. Anomaly detection relies on innovation residuals, triggering alerts when thresholds are exceeded. Monitored features include position drift, velocity inconsistencies, inertial sensor bias variation, and disagreement between vision and inertial measurements.

The EKF used for sensor fusion estimates the system state vector, including position, velocity, and orientation of the aerial vehicle. The filter continuously compares predicted and measured sensor values through innovation residuals. When the residual exceeds a predefined statistical threshold, the system identifies a potential anomaly in the sensor measurements. This approach enables the detection of navigation inconsistencies caused by spoofing attacks, sensor faults, or malicious data manipulation.

This layer ensures that the system can detect potential cyber-physical attacks, including sensor spoofing and data injection, within minimal latency. Fig. 1 illustrates the integration of the perception and integrity-monitoring modules in the overall system architecture.

2.3. Secure Decision-Making and Validation Layer

Autonomous navigation commands generated by the onboard AI are validated by a safety filter that enforces flight envelope constraints, obstacle clearance, and integrity alerts. Commands associated with detected anomalies are rejected, triggering a reactive control module that executes safe fallback actions such as hover, loiter, or return-to-home maneuvers.

This approach prevents unsafe actions while maintaining mission continuity. Table 2 summarizes the system performance under various adversarial scenarios, highlighting reductions in unsafe commands, false alarm rates, and reaction latency.

The interaction between the perception, integrity monitoring, and decision validation modules forms a layered defense architecture designed to improve system robustness. The perception module provides environmental awareness through visual analysis, while the integrity monitoring layer evaluates the reliability of sensor measurements and detects inconsistencies that may indicate cyber-physical attacks. The decision validation module enforces a formally defined safety envelope to prevent unsafe navigation commands. The constraints include maximum linear velocity (5 m/s), maximum angular rate (120 °/s), and minimum obstacle clearance distance (1.5 m). In addition, navigation commands are rejected when an integrity alert is active.

A command is classified as unsafe if it results in a predicted trajectory that violates any of these constraints or introduces a collision risk. When such a condition is detected, the command is discarded and a fallback control strategy is activated.

The fallback mechanism follows a hierarchical structure: (i) hover stabilization as default response, (ii) loiter mode when localization remains stable, and (iii) return-to-home if anomalies persist for more than 2 seconds. This ensures safe operation while minimizing mission disruption.

This formalization transforms the decision validation layer into a deterministic safety filter, strengthening the reliability of autonomous decision-making under adversarial conditions.

2.4. Sensor Fusion and Anomaly Detection Model

The integrity monitoring layer relies on an EKF to estimate the system state using measurements from the inertial measurement unit, visual perception module, and GNSS. The state vector can be represented as:

$$\boldsymbol{x}_{\mathrm{k}} = [\boldsymbol{p}_{\mathrm{k}}, \boldsymbol{v}_{\mathrm{k}}, \boldsymbol{\theta}_{\mathrm{k}}],$$

where p_{k} represents the position, v_{k} the velocity, and θ_{k} the orientation of the aerial vehicle at time step k.

The prediction stage follows the standard EKF formulation:

$$\hat{x}_{\mathrm{k}}|_{\mathrm{k-1}} = f(\hat{x}_{\mathrm{k-1}}, u_{\mathrm{k}}),$$

while the update stage incorporates sensor observations:

$$\hat{x}_{\mathrm{k}} = \hat{x}_{\mathrm{k}}|_{\mathrm{k-1}} + K_{\mathrm{k}}(z_{\mathrm{k}} - h(\hat{x}_{\mathrm{k}}|_{\mathrm{k-1}})),$$

where K_{k} is the Kalman gain and z_{k} represents the measurement vector.

Anomaly detection is performed using the normalized innovation squared (NIS), which provides a statistically grounded measure of consistency between predicted and observed measurements. The innovation residual is defined as:

$$r_{\mathrm{k}} = z_{\mathrm{k}} - h(\hat{x}_{\mathrm{k}}|_{\mathrm{k-1}})$$

The anomaly score is computed as:

$$\epsilon_k = r_k^T S_k^{-1} r_k,$$

where S_k represents the innovation covariance matrix. An anomaly is triggered when:

$$\varepsilon_k > \chi^2(\alpha, d),$$

where $\chi^2(\alpha, d)$ is the chi-square threshold corresponding to confidence level α and d degrees of freedom. In this work, a confidence level of $\alpha = 0.99$ was selected to balance detection sensitivity and false positive rate.

The process and measurement noise covariance matrices were empirically calibrated using nominal flight data. Sensitivity analysis showed that variations of ± 10 % in the threshold resulted in less than 3 % variation in detection performance, indicating robustness of the detection mechanism.

2.5. System Architecture Overview

The overall system architecture follows the layered design introduced in Section 2 and is implemented on an embedded UAV platform. The framework integrates three main components: a perception module, an integrity monitoring module, and a decision validation module.

Rather than repeating the methodological details, this section focuses on the practical integration of these components within the onboard system. The perception module processes visual data in real-time using the optimized Nano YOLOv5 model. The integrity monitoring module evaluates sensor consistency through EKF-based estimation, while the decision validation layer filters navigation commands based on predefined safety constraints.

All modules operate sequentially within the onboard computing pipeline, ensuring that perception outputs and state estimates are continuously verified before control actions are executed. This integrated design enables real-time operation under resource constraints while maintaining robustness against adversarial conditions.

Implementation details, including hardware configuration and optimization techniques, are provided in the following subsection.

2.6. Threat Model

The security analysis of the proposed framework considers a cyber-physical adversary capable of manipulating sensor data or navigation signals during autonomous flight operations. The attacker may attempt to influence the decision-making pipeline by injecting corrupted measurements into the navigation or perception modules.

Three primary attack vectors are considered in this study:

- Navigation spoofing attacks, where the adversary manipulates GNSS signals to introduce false position estimates and alter the flight trajectory;
- Adversarial perception attacks, where carefully crafted perturbations are introduced into camera images to mislead the deep learning perception model;
- Sensor data injection, where malicious values are inserted into inertial or navigation sensor streams at the middleware level.

The attacker is assumed to have the capability to interfere with external sensor inputs but does not have direct access to the internal flight controller firmware or onboard computing hardware. This assumption reflects realistic threat scenarios where adversaries attempt to manipulate external signals or communication interfaces rather than compromising the entire onboard system.

Under this threat model, the proposed security-aware Edge AI framework focuses on detecting inconsistencies between multiple sensor sources and preventing compromised data from influencing the autonomous control pipeline.

2.7. Algorithm: Secure Edge AI Decision Pipeline

Input: Sensor data streams (camera, GNSS, IMU)
Output: Validated navigation command:

i. Acquire sensor data from camera, GNSS receiver, and IMU;
ii. Perform visual perception using compressed CNN model;
iii. Estimate vehicle state using EKF-based sensor fusion;
iv. Compute innovation residual between predicted and observed measurements;
v. If residual exceeds anomaly threshold
   Trigger integrity alert;
vi. Apply decision validation filter to navigation commands;
vii. If command violates safety constraints or integrity alert is active
   Execute fallback maneuver (hover or return-to-home);
viii. Otherwise
   Transmit validated command to flight controller.

3. Experimental Setup and Results

This section presents the experimental evaluation of the proposed secure Edge AI framework. The objective of the experiments is to assess the capability of the system to detect and mitigate cyber-physical threats while maintaining real-time autonomous operation on embedded hardware.

The evaluation focuses on three representative attack scenarios affecting autonomous UAS, including navigation signal spoofing, visual adversarial perturbations, and sensor data manipulation.

Detection accuracy, false positive rate, latency, and inference speed were measured to assess system performance.

All flight experiments were conducted in controlled outdoor environments following standard UAV safety procedures.

3.1. Experimental Platform

Experiments were conducted on a quadrotor UAS equipped with an embedded computing platform and multiple onboard sensors. The embedded computing platform used in the experiments provides a balance between computational performance and energy efficiency, which is critical for small aerial vehicles. The NVIDIA Jetson Nano offers a GPU-accelerated architecture capable of executing deep learning inference while maintaining relatively low power consumption. This makes it suitable for real-time Edge AI applications in autonomous unmanned systems.

The PX4 autopilot stack was used to control the flight dynamics of the quadrotor platform and to interface with the onboard sensors. The quadrotor platform used in the experiments has a diagonal wheelbase of 450 mm and a total takeoff weight of approximately 1.8 kg. The onboard computing module operates under Ubuntu Linux and runs the Robot Operating System (ROS) middleware for sensor integration and communication between perception, navigation, and security modules. Sensor data streams are synchronized through ROS topics, allowing real-time access to navigation and perception information by the integrity monitoring processes.

The Edge AI inference module was implemented using TensorRT optimization to achieve real-time performance on the embedded hardware. The integrity monitoring and security verification processes were executed as companion processes on the same platform, allowing synchronized access to sensor data and navigation commands. The complete system architecture is illustrated in Fig. 1.

3.2. Test Scenarios

Three representative cyber-physical threat scenarios were implemented to assess robustness during controlled flight experiments.

The first scenario simulates a global navigation satellite spoofing attack by gradually injecting position drift into the navigation data stream. This attack aims to cause navigation errors and potentially unsafe trajectory deviations.

The second scenario introduces adversarial perturbations into the visual perception pipeline. These perturbations are generated using a fast gradient sign method applied to camera frames to evaluate the robustness of the perception model against manipulated visual inputs.

The third scenario simulates sensor data injection by modifying inertial measurement unit readings at the middleware level. This test evaluates the ability of the integrity monitoring layer to detect inconsistencies between inertial, visual, and navigation measurements.

Ground truth reference data were obtained from unaltered sensor logs and reference flights recorded under normal operating conditions.

3.3. Experimental Results

Each scenario was evaluated through twenty flight trials with an average duration of approximately five minutes. Detection accuracy, false positive rate, detection latency, and real-time inference performance were measured for each threat type. Table 2 summarizes the results obtained.

The results demonstrate that the proposed security-aware architecture maintains high detection accuracy across different threat scenarios while preserving real-time system performance. The relatively low detection latency enables the activation of safety mechanisms before compromised sensor data can significantly affect flight control decisions.

Table 2. Performance of the secure Edge AI framework under adversarial scenarios.

Threat Scenario	Detection Rate (%)	False Positive Rate (%)	Detection Latency (ms)	Inference Speed (FPS)
GNSS Spoofing	92 ± 3	4.1	180	20
Visual Perturbation	88 ± 4	5.3	210	18
Sensor Data Injection	90 ± 2	4.7	165	22
Average	90	4.7	185	20

Detection latency remains below 210 ms across all evaluated scenarios, allowing the security mechanisms to respond quickly and prevent unsafe navigation commands. In addition, the perception module maintains real-time inference speeds between 18 and 22 frames per second, demonstrating that the integration of cybersecurity mechanisms does not significantly degrade system performance.

3.4. Ablation Study

An ablation study was also conducted to evaluate the contribution of individual system components. The comparison included four configurations: a baseline system without security mechanisms, a system using only the integrity monitoring layer, a system using

only the safety validation filter, and the full proposed architecture. The results are presented in Table 3.

The ablation study highlights the importance of combining integrity monitoring and decision validation within a unified architecture. While individual modules provide moderate improvements in detection performance, the complete system significantly reduces unsafe commands and improves operational resilience. This demonstrates that layered security mechanisms are essential for protecting autonomous aerial systems operating in adversarial environments.

The complete system reduced unsafe commands by more than 85 % compared with the baseline configuration while maintaining a mission performance degradation of less than 3 % during normal operation.

The results demonstrate that integrating integrity monitoring and secure decision validation with Edge AI significantly improves the resilience of autonomous UAS operating in adversarial environments.

To further evaluate the effectiveness of the proposed framework, its performance was compared with several recent approaches addressing security and attack detection in autonomous UAS.

3.5. Comparative Evaluation

The comparison focuses on detection accuracy, response latency, and suitability for embedded real-time deployment. The results of this comparison are summarized in Table 4.

Table 3. Ablation Study of Security Components.

System Configuration	Detection Rate (%)	Unsafe Commands	Mission Degradation (%)
Baseline (no security)	42	High	0
Integrity Monitoring only	78	Medium	1.2
Safety Validation only	73	Medium	1.5
Proposed Full Architecture	92	Low	2.8

Table 4. Comparative evaluation of the proposed framework with existing UAV security approaches.

Method	Attack Detection Rate (%)	Detection Latency (ms)	Real-time Embedded Support	Security Scope
ML-based GNSS spoofing detection	85	250	Limited	Navigation attacks
Deep learning UAV protection	87	230	Partial	GNSS spoofing
Ensemble spoofing detection	89	240	No	Navigation attacks
Reinforcement learning attack detection	90	300	No	Guidance attacks
Proposed Secure Edge AI Framework	92	180	Yes	Multi-sensor cyber-physical threats

3.6. Comparative Analysis with Existing Approaches

The comparative results presented in Table 4 highlight several significant differences between the proposed framework and existing approaches for securing autonomous UAS. While previous studies have demonstrated promising results in detecting specific attack vectors such as GNSS spoofing or navigation manipulation, most of these methods focus on a single subsystem of the UAV architecture.

For instance, machine learning–based spoofing detection approaches typically analyze navigation signals in isolation, without considering cross-validation with other onboard sensors. Similarly, deep learning protection mechanisms are often designed specifically for visual perception pipelines and may not address inconsistencies originating from navigation or inertial measurements. As a result, these solutions may fail to detect coordinated cyber-physical attacks that simultaneously affect multiple sensor streams.

By contrast, this approach uses a multi-layer architecture that unifies perception, integrity monitoring, and decision validation within the Edge AI pipeline. By integrating information from multiple onboard sensors through an EKF–based fusion process, the system can identify inconsistencies between navigation, inertial, and visual measurements. This cross-sensor validation significantly improves the reliability of anomaly detection compared with approaches relying on a single data source.

Another important distinction concerns computational feasibility for embedded deployment. Several existing solutions rely on computationally intensive deep learning models or require centralized processing infrastructures for anomaly detection. Such

approaches may introduce significant communication latency or may not be suitable for small UAV platforms with limited processing resources. To overcome this limitation, the system leverages optimized neural networks and integrity monitoring algorithms suitable for embedded platforms like the NVIDIA Jetson Nano.

From a performance perspective, the proposed framework achieves a detection rate of approximately 92 % with an average response latency of 180 ms, outperforming or matching the detection capabilities of several state-of-the-art approaches while maintaining full compatibility with real-time embedded operation. The reduced detection latency is particularly important for autonomous aerial systems, where delayed responses to cyber-physical attacks may lead to unsafe flight behaviors.

Furthermore, the proposed decision validation layer introduces an additional safety barrier that is not commonly present in existing detection frameworks. While many approaches focus exclusively on identifying anomalies, the architecture presented in this study actively prevents compromised data from influencing flight control commands by enforcing safety constraints and triggering fallback control behaviors.

Overall, the comparative analysis indicates that integrating perception, cybersecurity monitoring, and decision validation within a unified Edge AI architecture provides a more comprehensive and resilient protection strategy for autonomous UAS. This integrated design enables reliable operation in adversarial environments while maintaining the computational efficiency required for deployment on resource-constrained aerial platforms.

3.7. Evaluation Metrics

To ensure clarity and reproducibility, the evaluation metrics are formally defined as follows.

Detection rate represents the percentage of attack instances correctly identified by the system. False positive rate indicates the proportion of normal operating conditions incorrectly classified as anomalies.

Detection latency corresponds to the time elapsed between the onset of an attack and its detection, which is a critical parameter for real-time autonomous systems.

An unsafe command is defined as any navigation command that leads to a predicted trajectory violating safety constraints, including obstacle proximity below 1.5 m, excessive velocity, or instability under inconsistent sensor data.

Mission degradation is defined as the relative increase in mission completion time compared to baseline operation:

$$Degradation(\%) = \frac{T_{attack} - T_{baseline}}{T_{baseline}} \times 100$$

All metrics were computed over 20 flight trials per scenario, each lasting approximately five minutes, ensuring statistical consistency.

4. Discussion

The experimental results demonstrate that integrating cybersecurity mechanisms directly within the Edge AI processing pipeline significantly enhances the operational resilience of autonomous UAS. The system effectively detects and mitigates cyber-physical threats while maintaining real-time performance on embedded hardware, highlighting the feasibility of onboard security-aware intelligence.

A key observation is the effectiveness of the integrity monitoring layer in detecting inconsistencies among navigation, inertial, and vision-based estimates with low latency. This capability is particularly important in scenarios where GNSS signals are unreliable or intentionally manipulated. Early detection of anomalies prevents corrupted data from propagating through the decision pipeline, thereby reducing the risk of unsafe system behavior.

The results also underline the importance of decision-level safety enforcement. While anomaly detection identifies suspicious inputs, the validation mechanism ensures that unsafe commands are not executed. This complementary interaction between detection and validation contributes to a significant reduction in unsafe actions, while fallback strategies such as hover stabilization and return-to-home preserve mission continuity.

Another important outcome concerns the balance between security and operational efficiency. The results indicate that improved resilience does not significantly degrade mission performance, suggesting that security-aware processing can be integrated without compromising real-time constraints. This is particularly relevant for embedded platforms, where computational resources are limited.

From a broader perspective, the findings highlight the potential of integrating perception, integrity monitoring, and decision validation within a unified onboard framework. Such integration enables faster response times compared to centralized approaches and improves robustness against multiple attack vectors. This is especially relevant for autonomous systems operating in communication-constrained or adversarial environments.

Despite these advantages, several limitations remain. The current implementation focuses on sensor integrity and adversarial perception attacks, while other threats such as communication link interference or coordinated multi-sensor attacks are not explicitly addressed. Furthermore, the system relies on predefined thresholds and models, which may require adaptation in highly dynamic environments. Future work will explore adaptive and learning-based security mechanisms capable of identifying previously unseen attack patterns.

From an application standpoint, the proposed approach is particularly relevant for mission-critical operations such as infrastructure inspection, environmental monitoring, and emergency response. In these scenarios, reliable onboard detection of sensor anomalies and prevention of unsafe navigation decisions are essential for ensuring safe and continuous operation, especially when communication with ground control is limited.

Overall, the results support the integration of security-aware mechanisms directly within the onboard decision pipeline as an effective strategy for improving the reliability, safety, and autonomy of next-generation aerial systems.

5. Conclusions

The experimental evaluation demonstrates that integrating cybersecurity monitoring directly within the Edge AI pipeline significantly improves the resilience of autonomous UAV systems against cyber-physical threats.

This article presented a security-aware Edge AI framework aimed at improving the resilience of autonomous decision-making in UAS. These findings indicate that security-aware Edge AI architectures represent a promising direction for enabling safe and trustworthy autonomous aerial systems in complex operational environments.

Experimental evaluation on a quadrotor platform demonstrated that the system could detect and mitigate several cyber-physical threats, including navigation spoofing, visual adversarial perturbations, and sensor data manipulation.

The results obtained in this study suggest that integrating security mechanisms directly within Edge AI architectures can significantly improve the reliability of autonomous systems operating in complex environments. Such approaches will play a crucial role in enabling safe large-scale deployment of unmanned aerial platforms in civilian, industrial, and emergency response applications.

The study highlights the importance of incorporating cybersecurity mechanisms directly within the onboard AI pipeline of autonomous aerial systems. Such integration enables safer deployment of unmanned platforms in environments where communication links may be unreliable, or adversarial conditions may be present.

Another important aspect of the proposed framework concerns computational efficiency. The perception model was optimized through model compression techniques including quantization and structured pruning, enabling efficient execution on resource-constrained hardware. Experimental results demonstrate that the integrated architecture maintains real-time inference speeds while simultaneously executing cybersecurity monitoring processes. This confirms the feasibility of deploying security-aware AI systems directly on embedded aerial platforms without requiring high-performance ground infrastructure.

Future research will focus on addressing additional threat vectors, including communication-level attacks and cooperative multi-drone scenarios. Further improvements will also explore adaptive security mechanisms and more advanced perception models capable of improving robustness in complex operational environments.

6. Limitations and Future Work

Although the proposed framework demonstrates promising results for improving the resilience of autonomous UAS, several limitations should be acknowledged.

First, the current implementation focuses primarily on attacks affecting sensor integrity and visual perception. Other types of cyber-physical threats, such as communication link interference, coordinated multi-drone attacks, or malicious firmware modifications, were not addressed in the present study. Developing capabilities to detect and mitigate these threats is a key direction for future research.

Second, the evaluation was conducted on a single quadrotor platform equipped with a specific embedded computing device. While the results demonstrate the feasibility of secure Edge AI on resource-constrained hardware, further experiments on different aerial platforms and processing architectures would provide a more comprehensive assessment of system scalability and robustness.

Another promising direction for future research involves the integration of cooperative security mechanisms in multi-drone systems. In such environments, multiple UAV may share sensor information and collaboratively detect cyber-physical anomalies affecting the swarm. Developing distributed security-aware decision frameworks for cooperative aerial systems could further enhance resilience and mission reliability in complex operational scenarios.

Future research will focus on extending cybersecurity coverage, enhancing adaptive detection, and validating the approach in larger UAV fleets and real operational environments.

The complete system operates within the computational constraints of a low-power embedded platform, demonstrating that security-aware Edge AI can be effectively deployed in small UAV systems without requiring high-performance computing resources.

References

[1]. A. Hakeem, M. Sabir, R. M. Alhebshi, A. G. Almakky, et al., Integrating artificial intelligence for improved security of IoT-drones through cyber-physical attack detection, *Frontiers in Computer Science*, Vol. 7, 2025, 1545282.

[2]. A. Al-Sabbagh, A. El-Bokhary, S. El-Koussa, A. Jaber, et al., Enhancing UAV security against GPS spoofing attacks through a genetic algorithm-driven deep learning framework, *Information*, Vol. 16, Issue 2, 2025, 115.

[3]. L. Alhoraibi, D. Alghazzawi, R. Alhebshi, Detection of GPS spoofing attacks in UAVs based on adversarial machine learning model, *Sensors*, Vol. 24, Issue 18, 2024, 6156.

[4]. A. Almadhor, J. Baili, S. Alsubai, A. Al Hejaili, et al., CTDNN-Spoof: compact tiny deep learning architecture for detection and multi-label classification of GPS spoofing attacks in small UAVs, *Scientific Reports*, Vol. 15, Issue 1, 2025, 6656.

[5]. T. Hickling, N. Aouf, P. Spencer, Robust adversarial attacks detection based on explainable deep reinforcement learning for UAV guidance and planning, *IEEE Transactions on Intelligent Vehicles*, Vol. 8, Issue 10, 2023, pp. 4381-4394.

[6]. J. Zhou, M. Hu, C. Zhou, Z. Liu, et al., Research on GNSS spoofing detection and autonomous positioning technology for drones, *Electronics*, Vol. 14, Issue 15, 2025, 3316.

[7]. L. Al-Soufi, T. Demirsoy, E. Gelal Soyak, An implementation-based study of the detection of and recovery from GPS spoofing attacks for unmanned aerial vehicles, *International Journal of Advances in Engineering and Pure Sciences*, Vol. 36, Issue 3, 2024, pp. 211-223.

[8]. S. E. Meheretu, E. Nigussie, G. B. Gebremeskel, S. Y. Hailesilassie, A systematic literature review on spoofing and jamming approaches in unmanned aerial vehicles navigation, *Journal of Aerospace Technology and Management*, Vol. 17, 2025, e3425.

[9]. D. Alsadie, J. Tian, Cybersecurity and artificial intelligence in unmanned aerial vehicles: Emerging challenges and advanced countermeasures, *IET Information Security*, Vol. 2025, 2025, 2046868.

[10]. O. Y. Al-Jarrah, A. S. Shatnawi, M. M. Shurman, O. A. Ramadan, et al., Exploring deep learning-based visual localization techniques for UAVs in GPS-denied environments, *IEEE Access*, Vol. 12, 2024, pp. 113049-113071.

[11]. L. A. Fagundes-Junior, K. B. de Carvalho, R. S. Ferreira, A. S. Brandão, Machine learning for unmanned aerial vehicles navigation: An overview, *SN Computer Science*, Vol. 5, Issue 2, 2024, 256.

[12]. M. Adam, M. Hammoudeh, R. Alrawashdeh, B. Alsulaimy, A survey on security, privacy, trust, and architectural challenges in IoT systems, *IEEE Access*, Vol. 12, 2024, pp. 57128-57149.

[13]. S. K. Khan, Cybersecurity modelling for the adoption and deployment of connected and automated vehicles, PhD Thesis, *University of Wollongong*, Wollongong, 2023.

[14]. M. S. Islam, A. S. Mahmoud, T. R. Sheltami, AI-enhanced intrusion detection for UAV systems: A taxonomy and comparative review, *Drones*, Vol. 9, Issue 10, 2025, 682.

[15]. J. P. A. Yaacoub, H. N. Noura, O. Salman, K. Chahine, Toward secure smart grid systems: Risks, threats, challenges, and future directions, *Future Internet*, Vol. 17, Issue 7, 2025, 293.

[16]. A. Oun, K. Wince, X. Cheng, The role of artificial intelligence in boosting cybersecurity and trusted embedded systems performance: A systematic review on current and future trends, *IEEE Access*, Vol. 13, 2025, pp. 55258-55276.

[17]. S. Silva, Secure edge AI for autonomous decision-making in unmanned aerial systems, in *Proceedings of the 2nd International Conference on Drones and Unmanned Systems (DAUS'26)*, 2026, pp. 189-190.