IJIE

UTHM Penerbit
Universiti Tun Hussein Onn Malaysia

INTERNATIONAL JOURNAL OF INTEGRATED ENGINEERING
ISSN: 2229-838X e-ISSN: 2600-7916

Vol. 18 No. 4 (2026) 50-62
https://publisher.uthm.edu.my/ojs/index.php/ijie

Anomaly Detection in Autonomous Fixed-Wing UAV Agricultural Mapping Missions Using Deep Learning

Zainal Arifin[1], Lilik Anifah[1]*, IGP Asto Buditjahjanto[1]

[1] *Department of Electrical Engineering, Faculty of Engineering, State University of Surabaya, Surabaya 60231, INDONESIA*

*Corresponding Author: lilikanifah@unesa.ac.id
DOI: https://doi.org/10.30880/ijie.2026.18.04.004

Article Info

Received: 28 February 2026
Accepted: 7 May 2026
Available online: 17 June 2026

Keywords

UAV, anomaly detection, autoencoder, precision agriculture, deep learning

Abstract

Autonomous fixed-wing UAVs are now a cornerstone of precision agriculture, especially for large-scale mapping. But even small aerodynamic disturbances or slight shifts in the center of gravity (CG) can introduce subtle instabilities in flight. These small changes may not seem dramatic, yet they can reduce mapping accuracy and potentially affect operational safety. In this study, we present an unsupervised deep learning framework for detecting flight anomalies using multivariate telemetry data (specifically, roll, pitch, and altitude signals). Three autoencoder architectures were tested: LSTM, GRU, and a 1D-CNN. The evaluation was based on telemetry collected from 13 real autonomous agricultural flights, where CG variations were intentionally controlled to create realistic instability conditions. Anomalies were identified using window-level reconstruction errors, and performance was evaluated through threshold-independent ROC–AUC analysis. All three models perfectly classified severe anomalies, achieving accuracies and F1 Scores of 1.00. When it came to milder CG deviations, differences became more apparent. The GRU autoencoder showed the strongest sensitivity, reaching an ROC–AUC of 1.00 and detecting over 98% of mild anomaly cases. The LSTM achieved near-perfect separability (AUC = 0.99), while the 1D-CNN delivered solid performance (AUC = 0.97) with the added advantage of lower computational complexity. Overall, the results confirm that recurrent autoencoder architectures are highly effective for real-time UAV anomaly monitoring. Among them, the GRU offers the best balance between detection performance and model efficiency, making it particularly well-suited for embedded agricultural UAV applications.

1. Introduction

Fixed-wing UAVs have become an important platform in precision agriculture due to their ability to cover larger operational areas and achieve longer flight endurance than multirotor systems. Tokekar et al. [13] Although autonomous waypoint navigation is commonly employed, flight stability can still be influenced by aerodynamic imbalance, center of gravity (CG) displacement, and environmental disturbances such as wind turbulence. These factors may introduce longitudinal oscillations, degrading mapping quality and reducing the reliability of autonomous missions. Deep learning techniques have shown strong potential for anomaly detection in cyber-physical and aerospace systems. Pang et al. [5] presented a comprehensive review of deep learning approaches to anomaly detection, emphasizing the effectiveness of neural architectures for modeling nonlinear temporal relationships in multivariate datasets. Similarly, Chalapathy and Chawla [8] reviewed deep learning-based anomaly detection methods and highlighted the increasing importance of unsupervised learning frameworks for complex sensor-driven systems.

Autoencoder-based approaches have become widely adopted for unsupervised anomaly detection in time-series applications. Xu et al. [7] proposed a variational autoencoder framework for detecting anomalies in seasonal multivariate data through reconstruction behavior. Hundman et al. [3] introduced an LSTM-based anomaly detection framework for spacecraft telemetry, demonstrating that recurrent reconstruction models can effectively identify temporal irregularities via dynamic thresholding. These findings suggest that reconstruction-based learning is particularly well-suited for telemetry-driven anomaly detection in aerospace systems. In UAV applications, deep learning has increasingly been utilized for fault diagnosis and anomaly monitoring. Ozkat [4] developed a vibration-based anomaly-detection framework using a wavelet-scattering LSTM autoencoder and demonstrated its effectiveness in identifying structural abnormalities in UAV systems. Zhao and Xu [2] proposed a multi-sensor deep metric-learning method for detecting unknown UAV anomalies using heterogeneous telemetry signals. In addition, Feng et al. [1] introduced a causality-enhanced graph neural network that models the structural relationships among UAV sensor nodes to improve anomaly localization performance.

Recent research has also explored spatio-temporal and distributed learning approaches for UAV monitoring systems. Yang et al. [9] proposed a spatio-temporal regression framework for anomaly detection and recovery in UAV flight data by modeling correlations among dynamic flight states. Hafiz et al. [6] evaluated a deep learning framework for UAV network anomaly detection across multiple datasets, demonstrating the robustness of deep architectures in UAV-related monitoring environments. Furthermore, Chen et al. [12] discussed the integration of edge-cloud computing architectures in UAV systems, emphasizing the importance of computational efficiency for real-time onboard intelligent processing.

Despite the promising results reported in previous studies, several limitations remain. Most existing works focus primarily on severe fault conditions, network intrusion scenarios, or vibration-based monitoring, while subtle aerodynamic instability caused by CG shifts during autonomous agricultural cruise missions remains insufficiently explored. Moreover, comparative evaluations between lightweight recurrent and convolutional autoencoder architectures for real-time UAV deployment are still relatively limited. This study evaluates LSTM, GRU, and 1D-CNN autoencoders for detecting CG-induced flight anomalies using real fixed-wing UAV telemetry collected from autonomous agricultural mapping missions. The main contributions of this study are summarized as follows: This study contributes to the field in several important ways. First, a real experimental dataset of fixed-wing UAV telemetry with controlled CG variations was developed to represent realistic aerodynamic instability conditions. Second, an unsupervised reconstruction-based anomaly detection framework was implemented to detect subtle longitudinal instability patterns. Finally, a comparative evaluation of LSTM, GRU, and 1D-CNN autoencoders was conducted to examine the trade-off between anomaly detection capability and computational complexity for real-time onboard UAV deployment. The comparative evaluation framework implemented in this study treats LSTM, GRU, and 1D-CNN autoencoders as baseline deep learning architectures representing recurrent and convolutional temporal learning approaches for UAV telemetry anomaly detection.

1.1 Research Gap and Contributions

Although previous studies have investigated UAV anomaly detection using deep learning approaches, several limitations still remain. Most existing works primarily focus on severe fault conditions, vibration-based analysis, or simulated datasets, while subtle aerodynamic instability caused by center of gravity (CG) shifts during autonomous agricultural mapping missions has received relatively limited attention. In addition, comparative evaluations of lightweight deep autoencoder architectures for real-time onboard deployment remain underexplored. This study contributes to the field in several important aspects:

- Development of a real experimental fixed-wing UAV dataset with controlled CG variations
- Investigation of mild longitudinal instability detection using unsupervised reconstruction-based learning
- Comparative evaluation of LSTM, GRU, and 1D-CNN autoencoders under identical flight conditions
- Analysis of the trade-off between detection capability and computational complexity for onboard UAV implementation

2. Materials And Methods

This section outlines the experimental setup, including the UAV platform, flight data collection process, preprocessing steps, and the deep learning framework used for anomaly detection. The experiments were conducted using a fixed-wing UAV performing autonomous agricultural mapping missions, during which controlled center-of-gravity (CG) variations were intentionally introduced. Multiple real-world flights were conducted to generate three data categories: normal, mild anomaly, and severe anomaly. The collected telemetry data consisted of multivariate time-series signals (specifically roll, pitch, and altitude) extracted from onboard flight logs. A structured filtering process was applied to isolate stable cruise segments during autonomous flight mode, ensuring consistent data analysis. To prepare the data for deep learning, a sliding-window segmentation method was used to convert continuous telemetry streams into fixed-length sequences suitable for model training. For anomaly detection, an unsupervised framework was developed using three autoencoder architectures: LSTM, GRU, and 1D-CNN. Each model was trained solely on baseline (normal) flight data to learn the UAV's nominal dynamic behavior. Anomalies were then detected by measuring reconstruction errors. Model performance was evaluated using both threshold-based binary classification metrics and threshold-independent ROC–AUC analysis to provide a comprehensive assessment. The following subsections provide detailed explanations of the UAV platform and flight configuration, telemetry extraction methodology, dataset preparation process, and the design of the deep learning architectures used in this study.

2.1 UAV Platform and Flight Setup

Fig. 1 illustrates the autonomous fixed-wing UAV flight scenario used in this study, highlighting the comparison between normal (Fig. 1a) and anomalous flight conditions (Fig. 1b). In the baseline setup, the UAV's center of gravity (CG) sits comfortably within its designed aerodynamic stability range. As a result, the aircraft maintains smooth and steady straight-line flight during agricultural mapping missions. It follows a predefined 1 km straight mapping path at a constant altitude of 60 m and a ground speed of 15 m/s. Under these normal conditions, roll and pitch changes are minimal, which helps maintain a consistent ground sampling distance (GSD) and reliable image overlap for accurate photogrammetric data collection.

The anomalous scenario, however, exhibits different dynamic behavior. Here, the CG is deliberately shifted rearward, reducing the aircraft's longitudinal stability margin. This change leads to noticeable altitude oscillations and periodic pitch fluctuations during cruise. Although the UAV follows the exact same waypoint trajectory as in the baseline case, it exhibits a clear vertical undulating motion along the straight mapping route. By intentionally adjusting the CG in a controlled way, the study introduces a physically meaningful and realistic source of instability—rather than relying on simulated or artificially injected faults. This ensures that the anomaly detection framework is tested against genuine aerodynamic disturbances. In this way, the flight scenario creates a clear link between physical aerodynamic imbalance and measurable deviations in multivariate telemetry data. These measurable changes form the backbone of the proposed deep learning–based anomaly detection approach. A total of thirteen autonomous mapping flights were conducted under the following conditions:

- Cruise altitude: 60 m
- Ground speed: 15 m/s
- Straight mapping trajectory

Flight categories:

- Baseline (normal CG): F02, F03, F05, F06, F07, F09
- Severe anomaly (rear CG shift): F01, F04, F08, F10, F11
- Mild anomaly (-3 mm CG shift): F12, F13

Telemetry signals used:

- Roll
- Pitch
- Barometric Altitude

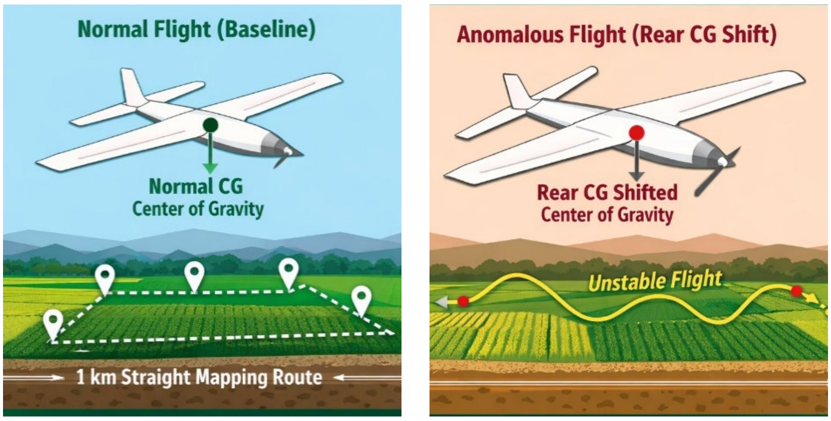
Only AUTO-level cruise segments were used.

2.2 Data Preprocessing and Window Segmentation

Telemetry data extracted from autonomous flight logs were first filtered to isolate stable cruise segments during AUTO flight mode. Only steady-level flight conditions were selected in order to minimize transient maneuvers that could introduce unrelated temporal fluctuations into the anomaly detection process. Three primary telemetry parameters—roll, pitch, and barometric altitude—were selected as multivariate input features because of their direct relationship with longitudinal flight stability and CG-induced oscillatory behavior. Prior to model training, all telemetry signals were normalized using z-score standardization:

$$\mathrm{x}_{norm} = \frac{x-\mu}{\sigma} \tag{1}$$

where μ and σ represent the mean and standard deviation of each telemetry feature computed from baseline flight data. To transform continuous telemetry streams into fixed-length temporal sequences, a sliding-window segmentation approach was used. Each input sample consisted of 80 consecutive telemetry observations with overlapping windows. This multivariate sequence structure enabled the deep autoencoder architectures to learn temporal flight dynamics over short- to medium-duration intervals. Only baseline flight data were used during training to allow unsupervised learning of nominal UAV behavior. Meanwhile, flights containing mild and severe anomalies were reserved exclusively for evaluation and performance analysis.



(a) **(b)**

Fig. 1 *UAV fixed-wing mission illustration (a) Normal flight; (b) Anomalous flight*

Table 1 *Experimental configuration*

Parameter	Value
Total Flights	13
Baseline Flights	6
Severe Anomaly Flights	5
Mild Anomaly Flights	2
Input Features	Roll, Pitch, Altitude
Window Size	80
Normalization	Z-score
Optimizer	Adam
Loss Function	Mean Squared Error
Batch Size	32
Epochs	30
Threshold Method	Mean + 3σ
Evaluation Metrics	Accuracy, Precision, Recall, F1-Score, ROC-AUC

All autoencoder models were trained exclusively using baseline flight telemetry to learn nominal UAV dynamic behavior in an unsupervised manner. During training, the models minimized reconstruction loss using the Adam optimizer with the mean squared error (MSE) objective. Early stopping was not implemented, and all models were trained for 30 epochs using identical preprocessing and window segmentation configurations to ensure a fair comparative evaluation. Following the training stage, anomaly detection was conducted by computing reconstruction errors for unseen telemetry windows. A statistical threshold was subsequently determined using the mean reconstruction error plus three standard deviations calculated from the training data.

2.3 Mathematical Formulation

Multivariate time-series representation:

$$X = \{x\_1, x\_2, \dots, x\_T\}, x\_t \sum \mathbb{R}^d \quad (2)$$

Autoencoder reconstruction loss:

$$L = (1/N) \sum \left\| W\_i - \widehat{W}\_i \right\|^2 \quad (3)$$

Anomaly score per window:

$$S\_i = (1/(L * d)) \sum (x - \hat{x})^2 \quad (4)$$

Threshold determination:

$$\theta = \mu\_train + 3\sigma\_train \quad (5)$$

ROC–AUC formulation:

$$AUC = \int_0^1 TPR(FPR)\, d(FPR) \quad (6)$$

2.4 Deep Learning Architectures

Three deep learning autoencoder architectures were implemented and comparatively evaluated in this study: Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), and a one-dimensional Convolutional Neural Network (1D-CNN) autoencoder. These architectures were selected to represent different temporal learning strategies for multivariate UAV telemetry anomaly detection. Recurrent architectures such as LSTM and GRU are designed to capture sequential temporal dependencies, whereas convolutional architectures focus on localized temporal feature extraction with lower computational complexity.

2.4.1 LSTM Autoencoder

The Long Short-Term Memory (LSTM) autoencoder was designed to capture long-term temporal dependencies within multivariate flight telemetry data. The model employs a stacked encoder–decoder LSTM structure and contains 18,224 trainable parameters. Through gated memory cells and recurrent feedback mechanisms, the LSTM architecture mitigates the vanishing gradient problem while preserving important dynamic state information across extended temporal sequences. This capability is particularly important for identifying longitudinal oscillatory behavior, where instability patterns develop gradually rather than appearing as sudden signal deviations. LSTM-based anomaly detection has demonstrated strong performance in aerospace telemetry applications. Hundman et al. [3] proposed an LSTM-based anomaly detection framework for spacecraft telemetry using dynamic thresholding and sequential reconstruction, demonstrating that recurrent autoencoders are effective at detecting temporal irregularities in complex aerospace systems. In UAV-related applications, Ozkat [4] developed a wavelet-scattering LSTM autoencoder for vibration-driven anomaly detection and demonstrated that temporal memory mechanisms can effectively capture structural abnormalities in UAV flight data. Due to its capability to learn long-term temporal correlations, the LSTM autoencoder is expected to perform effectively in detecting severe CG-induced instability, particularly when oscillatory behavior persists across multiple consecutive telemetry windows.

2.4.2 GRU Autoencoder

The Gated Recurrent Unit (GRU) autoencoder was implemented as a computationally efficient alternative to the LSTM architecture, consisting of 13,869 trainable parameters. Unlike LSTM, the GRU simplifies the recurrent gating mechanism by combining the forget and input operations into a single update gate. This streamlined structure reduces computational complexity while maintaining strong modeling of temporal dependencies. As a result, GRU architectures generally exhibit faster convergence and lower computational overhead, making them suitable for embedded UAV applications with limited onboard resources. GRU-based recurrent models have been widely adopted for multivariate time-series anomaly detection because they balance computational efficiency and temporal modeling capability. Pang et al. [5] highlighted the effectiveness of recurrent neural network architectures, including GRU models, for unsupervised anomaly detection in sequential data environments. Similarly, Chalapathy and Chawla [8] emphasized that lightweight recurrent frameworks are highly suitable for complex sensor-driven anomaly detection systems where temporal continuity is critical. In UAV telemetry applications, flight behavior often exhibits moderate, continuous temporal correlations. Therefore, the GRU architecture is particularly suitable for detecting subtle CG-induced anomalies while maintaining relatively low computational complexity, making it attractive for real-time onboard implementation.

2.4.3 1D-CNN Autoencoder

The one-dimensional Convolutional Neural Network (1D-CNN) autoencoder was implemented as a lightweight anomaly detection architecture with only 4,515 trainable parameters. Unlike recurrent models, the 1D-CNN does not rely on sequential memory mechanisms. Instead, temporal patterns are extracted using convolutional filters operating on localized sliding windows. This enables efficient identification of short-term signal variations while maintaining a significantly lower computational cost than recurrent architectures. Convolution-based anomaly detection methods have shown promising performance in multivariate monitoring systems. Hafiz et al. [6]

evaluated deep learning frameworks for UAV-related anomaly detection across multiple datasets and demonstrated the effectiveness of convolutional feature extraction in learning discriminative monitoring representations. Furthermore, Pang et al. [5] discussed that convolutional architectures are computationally efficient and well-suited for large-scale anomaly detection tasks involving multivariate time-series data. Although 1D-CNN architectures are highly attractive for resource-constrained onboard deployment, their focus on localized temporal features may reduce sensitivity to long-duration oscillatory behavior caused by CG imbalance. Nevertheless, their compact structure and efficient inference capability make them practical candidates for real-time UAV anomaly detection systems.

2.5 Training and Evaluation Procedure

All autoencoder models were trained exclusively using baseline flight telemetry to learn nominal UAV dynamic behavior in an unsupervised manner. During training, the models attempted to reconstruct multivariate telemetry sequences representing stable autonomous cruise conditions. The training process minimized reconstruction error using the Mean Squared Error (MSE) loss function together with the Adam optimizer. To ensure a fair comparative evaluation, identical preprocessing, normalization, and sliding window segmentation settings were applied across all architectures. Each model was trained for 30 epochs with a batch size of 32, while the input sequence length was fixed at 80 telemetry samples per window. No anomalous flight data were included during training in order to maintain the integrity of the unsupervised anomaly detection framework. After training, anomaly detection was performed by calculating reconstruction errors between the original telemetry sequence and the autoencoder's reconstructed output. The reconstruction error for each telemetry window was computed using:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(x_i - \hat{x}_i)^2 \quad (7)$$

where x_i represents the original telemetry signal and $\hat{x}_i$ denotes the model's reconstructed output. A statistical anomaly threshold was then determined using the mean reconstruction error plus three standard deviations derived from the baseline training data:

$$Threshold = \mu + 3\sigma \quad (8)$$

where μ denotes the mean reconstruction error, and σ represents the standard deviation of reconstruction errors computed from normal flight telemetry. Performance evaluation was conducted using Accuracy, Precision, Recall, F1-score, and Receiver Operating Characteristic Area Under Curve (ROC–AUC) metrics. In addition to severe anomaly evaluation, a sensitivity analysis was performed using mild CG-shift flight scenarios to assess each model's capability to detect subtle longitudinal instability patterns.

3. Results and Discussion

This section provides a thorough evaluation of the proposed anomaly detection framework using real experimental flight data. The analysis is presented step by step. It begins with binary classification results for severe anomaly detection, then proceeds to assess model sensitivity to mild CG deviations. Next, threshold-independent performance is examined using ROC–AUC analysis to verify how well the models separate normal and anomalous patterns in their learned feature space. To further interpret the results, reconstruction behavior is visually analyzed to understand how each model responds to instability. Finally, model complexity is evaluated to explore the trade-off between computational cost and detection performance. The evaluation strategy is carefully designed to address both theoretical robustness and practical deployment concerns in autonomous UAV systems. Severe anomalies represent clear cases of longitudinal instability caused by rearward CG shifts, while mild anomalies mimic subtle aerodynamic imbalances that may not be easily noticeable during flight. By integrating quantitative performance metrics, reconstruction-based insights, and computational complexity analysis, this section delivers a well-rounded assessment of model accuracy, sensitivity, and suitability for real-time agricultural UAV operations. Performance evaluation was carried out using multiple quantitative metrics, including Accuracy, Precision, Recall, F1-score, Anomaly Sensitivity Ratio, and Receiver Operating Characteristic Area Under the Curve (ROC–AUC). A comparative analysis of the LSTM, GRU, and 1D-CNN autoencoders was conducted under identical preprocessing and thresholding conditions to ensure a fair and consistent performance assessment. In addition to severe anomaly detection, mild CG-shift scenarios were also evaluated to examine each model's sensitivity to subtle longitudinal instability patterns that are difficult to detect using conventional threshold-based monitoring methods.

Int. Journal of Integrated Engineering Vol. 18 No. 4 (2026) p. 50-62

3.1 Binary Classification Performance

The histogram in Fig. 2 clearly shows a strong separation between the reconstruction mean-squared error (MSE) distributions for normal and anomalous flights. For baseline flights, most windows cluster tightly within a low-error range, suggesting that the autoencoder models effectively captured and learned the aircraft's nominal flight behavior during training. In contrast, anomalous flights display a wider distribution that shifts noticeably toward higher MSE values. This reflects the models' difficulty in accurately reconstructing the oscillatory altitude patterns caused by CG imbalance. Consequently, once instability appears, reconstruction quality drops and the error increases accordingly. The very small overlap between the normal and anomalous distributions highlights the strong discriminative power of the learned latent representations. Because of this clear separation, a statistical threshold such as mean plus three standard deviations (mean + 3σ) can be applied reliably without being overly sensitive to noise. Overall, the findings confirm that CG-induced longitudinal instability consistently amplifies reconstruction error across time windows, providing a solid and dependable foundation for binary anomaly detection.

Table 2 *Binary classification performance*

Model	Accuracy	Precision	Recall	F1-Score
LSTM Autoencoder	1	1	1	1
GRU Autoencoder	1	1	1	1
1D-CNN Autoencoder	1	1	1	1

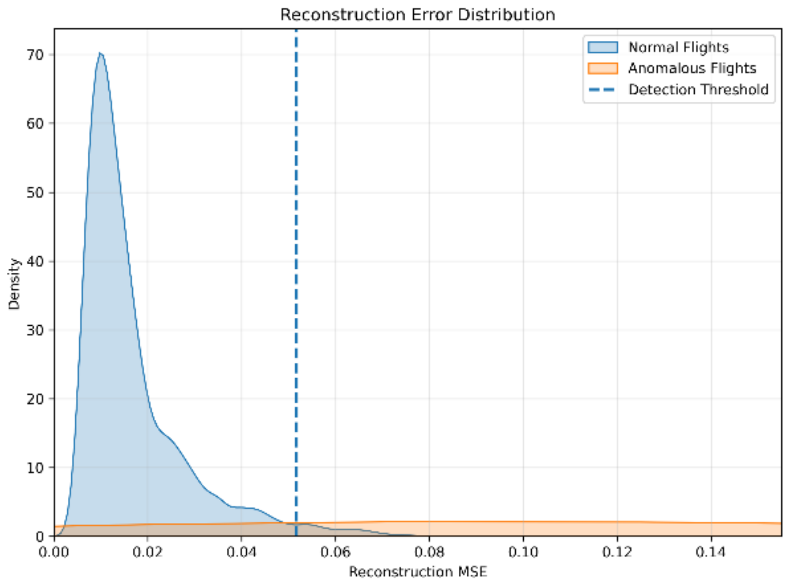


Fig. 2 *Distribution of reconstruction mean squared error (MSE) for normal and anomalous flights*

Table 3 *All three deep learning models achieved perfect flight-level binary classification*

Model	TP	TN	FP	FN	Accuracy
LSTM Autoencoder	7	6	0	0	1.00
GRU Autoencoder	7	6	0	0	1.00
1D-CNN Autoencoder	7	6	0	0	1.00

The results clearly show that severe anomalies caused by CG shifts can be easily distinguished from normal cruising behavior when using reconstruction-based anomaly scoring. The absence of false positives (FP = 0) indicates the models never flagged stable baseline flights as problematic. At the same time, zero false negatives indicate that every flight with a severe anomaly was successfully identified. These findings indicate that aerodynamic instability induced by a rearward CG shift leads to noticeable, statistically significant changes in the multivariate time series, particularly in roll, pitch, and altitude. The reconstruction errors for anomalous flight segments are distinctly separated from those of normal flights, making threshold-based classification both clear and reliable. From a practical standpoint, this kind of performance is exactly what's needed in real-time UAV monitoring. False alarms can unnecessarily interrupt autonomous missions, while missed detections could compromise flight safety and reduce mapping accuracy. Achieving both zero false positives and zero false negatives highlights the framework's strong potential for real-world deployment.

3.2 Mild Anomaly Sensitivity Analysis

Table 2 summarizes how well each model detected mild CG deviations in flights F12 and F13. Unlike the severe cases, which all models identified perfectly, these mild anomalies represent subtle longitudinal instability, making them much more challenging to detect. The GRU autoencoder delivered the strongest performance, correctly identifying 100% of anomalous windows in F12 and 98.8% in F13. With an average sensitivity of 0.99, the GRU clearly maintains high responsiveness even when the anomaly magnitude is small. This shows that its architecture remains sensitive to slight but meaningful changes in flight dynamics. The LSTM autoencoder also performed well, though with slightly lower sensitivity. It achieved detection ratios of 0.826 for F12 and 0.855 for F13, resulting in an average of 0.84. While LSTM networks are designed to capture long-term temporal dependencies, mild CG shifts may not always yield strong enough reconstruction errors across all windows. As a result, some subtle anomalies go partially undetected. Even so, the LSTM model still successfully recognized most instability patterns. In contrast, the 1D-CNN autoencoder showed considerably lower sensitivity, with detection ratios of 0.616 for F12 and 0.653 for F13. Its average sensitivity of 0.63 means that nearly one-third of mild anomaly windows were missed. This suggests that mild CG-induced oscillations primarily affect broader temporal dynamics rather than short, localized signal variations, an area where convolutional models are less effective compared to recurrent architectures. Overall, the results in Table 4 highlight the importance of model architecture when targeting subtle aerodynamic imbalances. Although all models can reliably detect severe instability, the GRU architecture excels at early-stage anomaly detection, making it especially suitable for real-time UAV monitoring, where early warning capability is critical.

Table 4 *Although all models perfectly detected severe anomalies, further evaluation on mild CG deviations (F12 and F13) reveals significant performance differences ROC–AUC analysis*

Model	F12	F13	Average
LSTM Autoencoder	0.826	0.855	0.84
GRU Autoencoder	1	0.988	0.99
1D-CNN Autoencoder	0.616	0.653	0.63

3.3 ROC-AUC Analysis

Although all three models showed perfect performance in detecting severe anomalies, a deeper analysis of the milder CG deviations, particularly in flights F12 and F13, reveals noticeable differences. These subtle instability cases provide a more demanding test, highlighting how sensitive each architecture is to anomalies when they are less pronounced. As shown in Fig. 3, the GRU autoencoder achieves the highest area under the receiver operating characteristic (ROC) curve (AUC = 1.00), indicating perfect discrimination across all threshold values. In practical terms, this means the GRU can consistently and clearly separate normal flight behavior from anomalous patterns, no matter where the classification threshold is set. The LSTM model performs nearly as well, with an AUC of 0.99. This reflects very strong separability, with only slight differences compared to GRU. On the other hand, the 1D-CNN model achieves an AUC of 0.97, still a strong result, but marginally lower. While it demonstrates strong discrimination capability, it is somewhat less precise than recurrent models at identifying subtle instability patterns. Overall, these findings reinforce the idea that recurrent architectures, especially GRUs, are better suited for detecting mild CG-induced deviations, where sensitivity to small temporal changes is crucial.

Int. Journal of Integrated Engineering Vol. 18 No. 4 (2026) p. 50-62

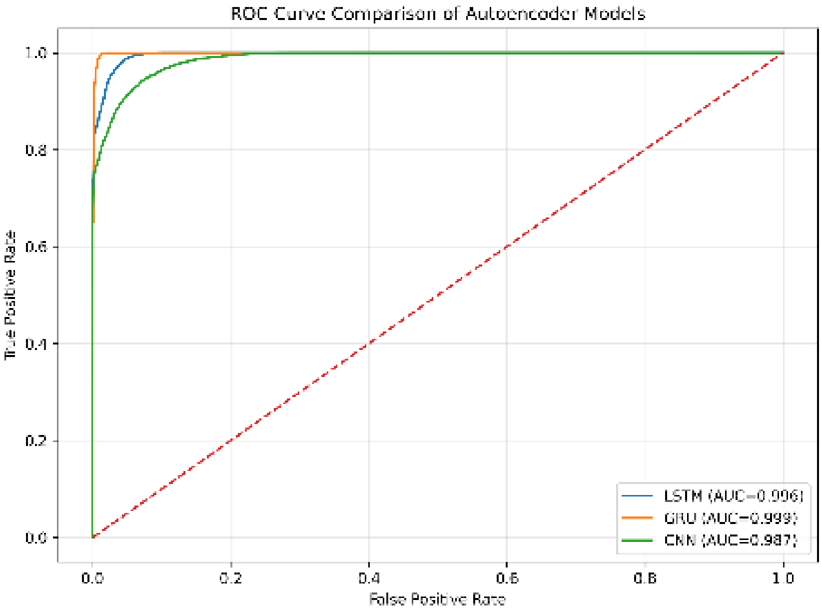


Fig. 3 *The ROC curves for the three autoencoder architectures*

Table 5 *The ROC–AUC results alongside model complexity metrics*

Model	AUC	Mild Sensitivity	Parameter	Complexity
LSTM Autoencoder	0.99	0.84	18,224	High
GRU Autoencoder	1	0.99	13,869	Medium
1D-CNN Autoencoder	0.97	0.63	4,515	Low

All three models achieved AUC values greater than 0.95, indicating a very strong ability to separate normal flight windows from anomalous ones. Simply put, each model can reliably distinguish stable cruise behavior from instability with high confidence. More importantly, the ROC–AUC analysis confirms that this performance isn't just due to selecting a favorable threshold. Instead, it demonstrates genuine separability within the learned latent feature space. In other words, the models are not just classified well by chance; they are truly learning and capturing the fundamental differences between nominal flight dynamics and instability patterns. Although all evaluated models achieved perfect binary classification at the selected threshold configuration, an ROC–AUC analysis was performed to assess separability performance independently of threshold selection. This provides a more robust evaluation of each model's discrimination capability across varying operating conditions.

3.4 Reconstruction Error Distribution

Reconstruction-based anomaly detection relies on a straightforward yet powerful principle: when a model is trained exclusively on normal data, it becomes very good at reproducing familiar patterns, but it tends to struggle when it encounters unexpected data. In practice, normal behavior is reconstructed accurately, whereas abnormal patterns lead to noticeably higher reconstruction errors. For this reason, examining how the model reconstructs signals under different flight conditions provides valuable insight into its sensitivity to instability. If the reconstruction quality clearly deteriorates during unstable or anomalous flights, it serves as direct evidence that the model is effectively detecting and responding to irregular flight dynamics.

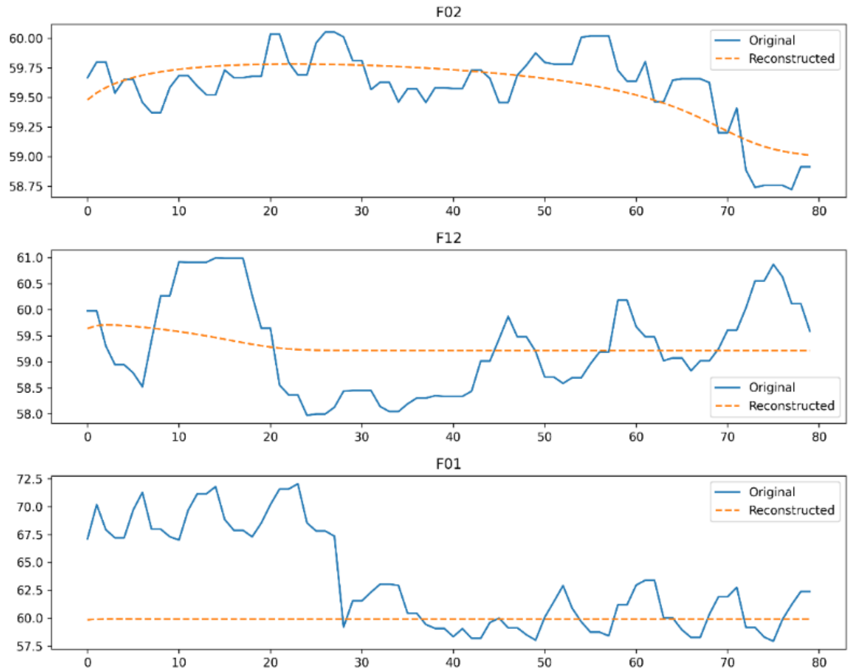


Fig. 4 *Comparison between original and autoencoder-predicted altitude signals for normal (F02), mild anomaly (F12), and severe anomaly (F01) flights. Increased deviation between original and reconstructed signals indicates instability severity*

The reconstruction comparison between the original altitude signals and the autoencoder-predicted outputs for normal (F02), mild anomaly (F12), and severe anomaly (F01) flights clearly illustrates how instability affects model performance. As the gap between the original and reconstructed signals widens, it reflects an increase in instability severity. These visual patterns are consistent with the earlier statistical analysis of reconstruction errors, which showed that anomalous flights had significantly higher mean squared errors (MSEs) than baseline missions. The gradual increase in reconstruction deviation, from normal to mild and then severe anomalies, further confirms that temporal autoencoder models are sensitive to oscillatory instability behavior. Overall, the findings highlight the strength of temporal sequence modeling. Recurrent architectures can learn and track dynamic dependencies in flight telemetry data. However, when instability introduces oscillatory disturbances, it disrupts these learned temporal patterns, leading to significantly larger reconstruction errors.

3.5 Model Complexity vs Detection Capability

This subsection examines the trade-off between model complexity and anomaly detection performance. While achieving high detection accuracy is essential, real-world UAV deployment also demands attention to computational efficiency, memory consumption, and real-time operation. For this reason, the analysis considers not only ROC–AUC performance but also the number of trainable parameters per model. By evaluating detection capability alongside computational footprint, the study provides a clearer picture of which architecture is most suitable for onboard implementation. As illustrated in Fig. 5, the LSTM autoencoder achieves strong detection performance (AUC = 0.99) but incurs the highest computational cost, requiring 18,224 trainable parameters. This makes it the most resource-intensive architecture among the three models evaluated. In contrast, the 1D-CNN autoencoder is much lighter, using only 4,515 parameters, about 75% fewer than LSTM, while still achieving a respectable AUC of 0.97. This sharp reduction in parameter count makes CNN particularly attractive for embedded UAV systems with limited memory and processing power. However, as discussed in Section 4.2, its lower sensitivity to mild CG deviations suggests that it struggles to fully capture longer-term temporal dependencies. The GRU autoencoder offers a well-balanced alternative. With 13,869 trainable parameters, it is significantly simpler than an LSTM yet achieves the highest detection performance (AUC = 1.00) and superior sensitivity to mild anomalies. This combination of strong temporal modeling and moderate computational demand makes GRU especially suitable for real-time onboard anomaly detection in agricultural UAV operations where both reliability and efficiency are critical.

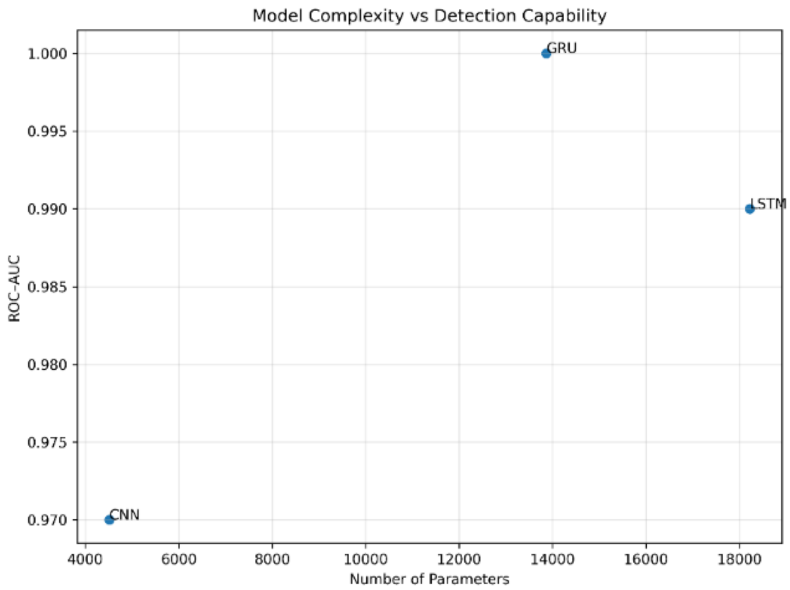


Fig. 5 *Trade-off between model complexity (number of trainable parameters) and anomaly detection capability (ROC–AUC). GRU achieves an optimal balance, delivering near-perfect detection performance with a lower parameter count than LSTM, while CNN offers reduced computational complexity at the expense of sensitivity*

3.6 False Alarm and Robustness Analysis

The false positive rate observed during baseline flights remained consistently low across all recurrent architectures, indicating stable reconstruction behavior under nominal cruise conditions. Both the GRU and LSTM models demonstrated strong robustness against normal telemetry fluctuations and environmental noise, with anomaly ratios approaching zero in most baseline missions. Although the CNN autoencoder achieved competitive binary classification performance, its sensitivity to mild anomalies was noticeably lower. This suggests that convolution-based local feature extraction may be less effective at modeling sustained oscillatory instability than recurrent temporal architectures. The results further indicate that reconstruction-based unsupervised anomaly detection generalizes across autonomous agricultural missions with varying flight durations and telemetry dynamics. This demonstrates the practical applicability of the proposed framework for real-world UAV monitoring systems.

4. Conclusion

This study introduced and experimentally validated an unsupervised deep learning framework for detecting anomalies in autonomous fixed-wing UAVs used for agricultural mapping. By analyzing multivariate telemetry data (specifically roll, pitch, and altitude), the system successfully identified flight instabilities caused by controlled shifts in the center of gravity (CG). The evaluation was based on 13 real autonomous flights, covering baseline, mild anomaly, and severe anomaly conditions. All three tested autoencoder architectures (LSTM, GRU, and 1D-CNN) showed strong performance in detecting severe anomalies, achieving perfect binary classification with accuracies and F1 Scores of 1.00. ROC–AUC analysis further confirmed excellent separability between normal and anomalous flight segments. Among the models, GRU achieved the highest AUC (1.00), followed closely by LSTM (0.99), while the CNN model reached 0.97. Importantly, GRU demonstrated the greatest sensitivity to mild CG shifts, while still maintaining a lower parameter count than LSTM. The 1D-CNN, although the most lightweight option, was less responsive to subtle oscillatory instability patterns. Overall, the results suggest that recurrent autoencoder architectures (especially GRU) strike the most practical balance between detection capability and computational efficiency, making them well-suited for real-time onboard anomaly monitoring in agricultural UAV systems. The study also highlights that CG-induced longitudinal instability can be effectively identified through reconstruction-based unsupervised learning, without the need for labeled anomaly data. Looking ahead, future research will focus on extending the framework to support multiclass anomaly classification and adaptive threshold optimization under varying environmental conditions, such as wind fluctuations and temperature changes. Integrating the anomaly detection system directly with onboard flight control mechanisms for autonomous mitigation also presents a promising next step.

5. Future Research Directions and Research Continuity

While this study confirms that deep autoencoder architectures are highly effective for binary anomaly detection in autonomous fixed-wing agricultural UAV missions, there's still plenty of room to grow, both academically and practically.

- Moving from binary to multi-class detection
  Right now, the framework separates flights into normal and anomalous categories. However, real-world UAV operations rarely involve just one type of issue. Future work could expand this into a multi-class anomaly detection system capable of distinguishing between CG imbalance, aerodynamic disturbances, actuator faults, sensor drift, and environmental turbulence. This would allow not just detection. But an accurate diagnosis enables smarter, faster corrective decisions during autonomous missions.
- Real-time onboard optimization
  Although the GRU model already offers a strong balance between detection performance and computational cost, deploying it efficiently on embedded UAV hardware remains an important goal. Techniques such as model pruning, quantization, and knowledge distillation could further reduce model size and inference time. Applying edge AI optimization strategies would help minimize latency while preserving detection sensitivity, an essential requirement for real-time flight monitoring.
- Expanding sensor integration
  This study focused on roll, pitch, and altitude signals, but incorporating additional telemetry sources could significantly enhance robustness. Features such as vibration data, airspeed, current consumption, and control surface commands may provide complementary information about flight health. Advanced multi-sensor fusion strategies, such as attention mechanisms or graph neural networks, could further improve anomaly detection under complex, variable environmental conditions.
- Closing the loop with adaptive flight control
  Perhaps the most impactful extension would be to integrate anomaly detection directly into adaptive flight control systems. Instead of simply flagging irregularities, future frameworks could automatically trigger mitigation strategies such as adaptive pitch adjustments, gain scheduling, or return-to-home procedures. This would shift the system from passive monitoring to active fault-tolerant control, significantly enhancing flight safety.

In summary, this study lays a solid foundation for scalable, intelligent UAV health monitoring. By expanding toward multi-class classification, embedded optimization, multi-sensor fusion, environmental generalization, interpretability, and closed-loop control integration, future research can drive the development of robust, real-time autonomous agricultural UAV systems capable of long-term, safe operation in dynamic field environments.

Acknowledgement

The authors would like to acknowledge the technical support provided during the UAV flight experiments and telemetry data acquisition. The authors also appreciate the assistance of all personnel involved in the preparation and execution of the experimental flight campaigns.

Conflict of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have influenced the work reported in this paper.

Author Contribution

The authors confirm contribution to the paper as follows: ***study conception and design:*** *Zainal Arifin, Lilik Anifah;* ***data collection:*** *Zainal Arifin;* ***analysis and interpretation of results:*** *Zainal Arifin, Lilik Anifah, IGP Asto Buditjahjanto;* ***draft manuscript preparation:*** *Zainal Arifin, Lilik Anifah, IGP Asto Buditjahjanto. All authors reviewed the results and approved the final version of the manuscript.*

References

[1] Feng, C., Fan, J., Liu, Z., Jin, G., & Chen, S. (2025). Unmanned aerial vehicle anomaly detection based on causality-enhanced graph neural networks, Drones, 9(6), 408. https://doi.org/10.3390/drones9060408

[2] Zhao, J., & Xu, X. (2025) A multi-sensor deep metric learning method for unknown anomaly detection in unmanned aerial vehicles, Engineering Applications of Artificial Intelligence, 111584. https://doi.org/10.1016/j.engappai.2025.111584

[3] Hundman, K., Constantinou, V., Laporte, C., Colwell, I., & Soderstrom, T. (2018). Detecting spacecraft anomalies using LSTMs and nonparametric dynamic thresholding, Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 387–395. https://doi.org/10.1145/3219819.3219845

[4] Ozkat, E. C. (2024). Vibration data-driven anomaly detection in UAVs: A deep learning approach, Engineering Science and Technology, an International Journal, 54, 101702. https://doi.org/10.1016/j.jestch.2024.101702

[5] Pang, G., Shen, C., Cao, L., & van den Hengel, A. (2021). Deep learning for anomaly detection: A review, ACM Computing Surveys, 54(2), 1–38. https://doi.org/10.1145/3439950

[6] Hafiz M.A., Inam U.K., M. Mansoor A., Mazliham M.S., Keshav K. (2026). RoboLSTM-IDS: multi-dataset evaluation of deep learning framework for UAV network. PeerJ Computer Science, 11, e3500. https://doi.org/10.7717/peerj-cs.3500

[7] Xu, H., Chen, W., Zhao, N., Li, Z., Bu, J., Li, Z., Liu, Y., Zhao, Y., Pei, D., Feng, Y., Chen, J., & Wang, Z. (2018) Unsupervised anomaly detection via variational auto-encoder for seasonal KPIs in web applications, Proceedings of the World Wide Web Conference, 187–196. https://doi.org/10.1145/3178876.3185996

[8] Chalapathy, R., & Chawla, S. (2019). Deep learning for anomaly detection: A survey, arXiv preprint arXiv:1901.03407. https://doi.org/10.48550/arXiv.1901.03407

[9] Yang, L., Li, S., Zhu, Z., Zhang, A., & Liao, Z. (2024). Spatio-temporal correlation-based multiple regression for anomaly detection and recovery of unmanned aerial vehicle flight data, Advances in Engineering Informatics, 60, 102440. https://doi.org/10.1016/j.aei.2024.102440

[10] Malhotra, P., Vig, L., Shroff, G., & Agarwal, P. (2016). Long short term memory networks for anomaly detection in time series, *Proceedings of the European Symposium on Artificial Neural Networks*, 89–94.

[11] Munir, M., Siddiqui, S.A., Dengel, A., & Ahmed, S. (2019). DeepAnT: A deep learning approach for unsupervised anomaly detection in time series, IEEE Access, 7, 1991–2005. https://doi.org/10.1109/ACCESS.2018.2886457

[12] Chen, W., Liu, B., Huang, H., Guo, S., & Zheng, Z. (2019). When UAV swarm meets edge-cloud computing: The QoS perspective, IEEE Network, 33(2), 36–43. https://doi.org/10.1109/MNET.2019.1800222

[13] Tokekar, P., Vander Hook, J., Mulla, D., & Isler, V. (2016). Sensor Planning for a Symbiotic UAV and UGV System for Precision Agriculture, IEEE Transactions on Robotics, 32(6), 1498–1511. https://doi.org/10.1109/TRO.2016.2603528