aerospace

MDPI

Article

A Two-Stage Framework for Microsatellite Thermal Mode Identification and Fault Detection via Clustering and Sequence Prediction

Weijian Pang [1,2,*], Jun Zhou [2], Jingwen Xu [2,*] and Xinian Zhi [1]

[1] Ningbo Institute of Northwestern Polytechnical University, Ningbo 315103, China
[2] Institute of Precision Guidance and Control, School of Astronautics, Northwestern Polytechnical University, Xi'an 710072, China
* Correspondence: pangwj@mail.nwpu.edu.cn (W.P.); xjw0324@nwpu.edu.cn (J.X.)

Abstract

Microsatellites operate in highly dynamic thermal environments due to severe physical constraints, making temperature telemetry a critical onboard health indicator. Conventional threshold-based monitoring fails to distinguish normal operational mode transitions from genuine faults, causing excessive false alarms. To address this, we propose a two-stage framework integrating unsupervised thermal mode discovery with mode-specific deep learning prediction. Raw temperature telemetry is downsampled and segmented into orbital cycles. Unsupervised clustering identifies two nominal thermal regimes and four canonical fault-type libraries (step, spike, drift, and noise), each corresponding to distinct in-orbit failure mechanisms. For each nominal mode, a Convolutional Neural Network–Long Short-Term Memory (CNN-LSTM) is trained on 7-day historical windows to forecast 3-day temperature evolution. Post-downlink, incoming cycle mode is inferred via nearest-neighbor DTW classification; anomalies are flagged when prediction residuals exceed mode-adaptive thresholds. Validation on Macau Science Satellite-1B (MSS-1B, COSPAR 2023-069-B, NORAD 56732) in-orbit telemetry from a 41° inclination low-Earth orbit—where solar illumination dominates external thermal loading and internal heat from the data-communication module and scientific payload constitutes the primary internal thermal source—shows the method reduces anomaly flags by 96.6% and improves prediction mean absolute error by 51.3% compared to a non-classified global baseline under nominal operating conditions, correctly detecting a known operational transient while suppressing spurious alarms. A synthetic fault injection experiment with four anomaly types and five baseline methods further confirms the framework's detection capability, achieving an overall F1 score of 0.725 vs. 0.258 for the global baseline—a 2.8× improvement driven primarily by a 4× precision gain. Sensitivity analysis reveals that the two-stage advantage is most pronounced for low-magnitude and short-duration faults, where mode-specific context is essential. This work advances microsatellite autonomous health management by providing reliable anomaly detection with quantified fault detection performance.

Keywords: fault detection; microsatellites; thermal control system; clustering; Dynamic Time Warping (DTW); anomaly pattern matching

Check for updates

Academic Editor: Hyun-Ung Oh

Received: 8 May 2026
Revised: 3 June 2026
Accepted: 8 June 2026
Published: 11 June 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

The evolution of microsatellites and CubeSats over the past two decades has transformed space access. Since the seminal CubeSat concept [1] and the IAA's advocacy for

cost-effective Earth observation missions [2], small satellites have progressed from educational platforms to operationally capable systems [3,4]. Today, microsatellites play an expanding role in telecommunications, Earth observation, and scientific exploration [5,6], which has correspondingly increased demands for their reliability and autonomy. Among onboard subsystems, the thermal control system is critical [7,8], as temperature anomalies often indicate or manifest in-orbit failures. Due to size, mass, and power constraints, microsatellites have low thermal inertia, making their temperature highly sensitive to internal power dissipation and the external orbital heat flux. Embedded attitude control actuators, such as printed magnetorquers integrated within satellite structures, further influence local thermal profiles during operation [9]. Consequently, microsatellite telemetry exhibits rich non-stationary dynamics, challenging conventional monitoring techniques for accurate interpretation.

In current engineering practice, temperature monitoring relies on fixed-threshold alarms, which fail to account for significant normal temperature fluctuations during operational mode changes (e.g., payload activation, attitude maneuvers, or high-speed data downlink). A temperature rise normal during high-power transmission may be misflagged as anomalous if calibrated for quiescent Earth-observation mode. Misclassifying nominal transitions as faults produces excessive false alarms, obscuring genuine anomalies, desensitizing ground operators, and undermining automated health monitoring credibility. Thus, a detection methodology capable of contextualizing satellite operational modes and distinguishing anomalies within each mode is urgently required.

Numerous machine learning-based time-series anomaly detection techniques exist [10,11], including prediction-based methods, reconstruction-based methods, and distance-based methods. However, their application to satellite systems has predominantly been limited to global binary anomaly classification. Current research exhibits two critical limitations. First, explicit modeling of embedded operational modes in temperature telemetry is rare. Most approaches treat the entire telemetry stream as a stationary process, failing to account for the cyclic, context-dependent nature of normal thermal behavior. Second, even when temporal dynamics are captured (e.g., via LSTM), single non-classified global methods are typically trained on all data. Such models must approximate complex, multi-modal functions, leading to averaging effects and poor generalization across distinct operational regimes. This constrains both detection granularity and accuracy, while providing no diagnostic insight into anomaly type. These two limitations motivate the corresponding contributions presented below.

This study addresses these gaps via an 'identify first, detect second' paradigm. The core idea is to unsupervisedly discover the satellite's intrinsic thermal operational modes directly from telemetry, followed by anomaly detection conditioned on inferred mode context. The main contributions are as follows. First, addressing the lack of mode-aware modeling, a two-stage clustering strategy for thermal mode discovery: temperature time series are segmented into orbital cycles and resampled; Dynamic Time Warping (DTW) distances are computed between all normalized cycles; DBSCAN is applied to the precomputed DTW matrix to isolate anomalous cycles, while K-means partitions normal cycles into two dominant thermal modes (low-power baseline and high-power operational modes) and the anomalous cycles into four canonical fault-type libraries—step, spike, drift, and noise—each corresponding to a distinct in-orbit failure mechanism (e.g., heater relay stuck-on, transient electrical overload, gradual sensor degradation, and electromagnetic interference), providing an interpretable, physically meaningful context for subsequent monitoring. Second, addressing the averaging effect of global models, mode-specific CNN-LSTM predictors for precise thermal modeling: for each normal mode, historical data are concatenated chronologically into a long sequence, and a sliding window scheme is used to

train a dedicated CNN-LSTM network whose convolutional layer extracts local temporal features while the LSTM captures long-range dependencies, yielding specialized predictors that faithfully reproduce mode-specific thermal dynamics. Third, comprehensive fault detection validation through synthetic anomaly injection: four types of artificial faults with varying magnitudes and durations are injected into the validation time series, and the two-stage method is benchmarked against five baseline approaches, demonstrating a 2.8× F1 improvement over the global CNN-LSTM baseline; a parameter sensitivity analysis further reveals that mode-specific context is most critical for detecting low-magnitude and short-duration anomalies.

It should be noted that the present study validates the framework on telemetry from a single temperature sensor of one satellite, and the anomaly pattern interpretations are derived from unsupervised cluster morphology rather than verified mission event logs. While synthetic fault injection provides quantitative detection performance metrics, real fault ground truth remains unavailable. These limitations temper the generalizability of the conclusions and are discussed in detail in Section 6.

The remainder of this paper is organized as follows. Section 2 reviews related work and positions the contribution. Section 3 details the two-stage methodology, including data preprocessing, clustering, mode-specific CNN-LSTM training, and anomaly pattern matching. Section 4 describes the dataset and preprocessing steps. Sections 5 and 6 present and discuss experimental results, respectively. Section 7 concludes the paper and outlines future research directions.

2. Related Work

This section situates the present work within the broader landscape of satellite health monitoring and draws methodological inspiration from adjacent fields. We first review data-driven spacecraft anomaly detection approaches, highlighting their achievements and the specific gap motivating the proposed two-stage framework. We then discuss key techniques from industrial process monitoring and computer vision that inform the methodology.

2.1. Data-Driven Satellite Health Monitoring: From Global Models to Context-Aware Detection

Data-driven approaches have emerged as a promising alternative to fixed-threshold monitoring for spacecraft health management [12,13]. Over the past decade, machine learning has driven a paradigm shift in anomaly detection for satellite systems. Long Short-Term Memory (LSTM) networks have emerged as a powerful tool for modeling satellite system dynamics, leveraging their ability to capture long-term temporal dependencies. Islam and Rahimi [14] demonstrated LSTM outperforms traditional statistical models (e.g., ARIMA) in predicting reaction wheel telemetry, while Gong et al. [15] extended this to space–ground collaborative fault diagnosis. Similarly, Cipollone et al. [16] deployed an LSTM-based algorithm for satellite maneuver detection from pattern-of-life data. These studies confirm the efficacy of recurrent architectures for learning nominal behavior from telemetry [17,18]. More recently, CNN-based approaches have demonstrated remarkable performance in satellite fault classification [19,20].

Beyond prediction-based methods, distance- and reconstruction-based techniques have also been extensively explored. Cui et al. [21] successfully applied fast Dynamic Time Warping (fast-DTW) with improved k-nearest neighbors for satellite time-series anomaly detection, demonstrating DTW's robustness to phase shifts and non-linear temporal distortions [22]—a critical property for comparing orbital cycles, which may exhibit slight alignment variations. Autoencoders and their variants [23,24] reconstruct nominal telemetry; anomalies are flagged when reconstruction error exceeds a threshold.

Despite these advances [25–27], most existing methods remain global models that treat the entire telemetry stream as a single, stationary process. They rarely model the underlying operational modes governing temperature dynamics. Consequently, legitimate mode transitions (e.g., Earth observation to high-speed downlink) can be misclassified as anomalies, causing high false-alarm rates in practice. Additionally, global models must approximate a complex, multi-modal function, producing averaged predictions that fail to accurately represent individual modes. This limits both detection sensitivity and the ability to distinguish fault types. Table 1 summarizes these limitations and how the proposed framework addresses each.

Table 1. Limitations of existing approaches.

Existing Method Category	Key Limitation
Fixed-threshold monitoring	Cannot adapt to multi-regime thermal dynamics
Global prediction models	Averaging effect across modes; poor generalization for individual regimes
Global reconstruction models	Single reconstruction target for multi-modal data; limited diagnostic granularity
Distance-based methods	Binary classification without contextual mode awareness
Supervised multi-class approaches	Require expensive manual mode labeling

2.2. Methodological Inspirations from Adjacent Fields

The proposed two-stage strategy and its algorithmic components draw conceptual and technical inspiration from two well-established domains.

1. Industrial Process Monitoring: The 'Divide and Conquer' Paradigm
   In complex industrial systems, distinct operational regimes (working conditions) are first identified, followed by local monitoring models for each regime. This 'divide and conquer' philosophy [28] effectively handles non-stationary, multi-modal processes, and its strength lies in converting an intractable multi-modal modeling problem into several simpler sub-problems—each with homogeneous data distribution. The proposed framework adopts this principle: DTW-based clustering identifies thermal regimes, and mode-specific CNN-LSTM networks serve as local models for residual-based anomaly detection. Experimental results confirm this approach enhances predictive fidelity and fault discriminability compared to a single non-classified global method.
2. Computer Vision and Signal Processing: Feature Transformation and Data Augmentation
   Transforming time series into alternative representations to leverage powerful neural network is well-established in fault diagnosis. Zhao et al. [19] converted control moment gyroscope signals into spectrograms via short-time Fourier transform for automated CNN-based fault diagnosis design, while Huang et al. [29] developed a CNN-LSTM approach for complex system fault diagnosis. The strength of representation transformation lies in exposing structural features that are latent in raw sequential data, enabling more effective pattern recognition by deep networks. While we retain the temporal domain and use CNN-LSTM for direct sequence prediction, this cross-domain concept guides the feature engineering: the convolutional layers extract local temporal features analogous to visual edges, providing a rich hierarchical representation of thermal dynamics.
   Furthermore, data augmentation techniques such as DTW oversampling [21] and generative adversarial networks [30] have been adapted to satellite fault diagnosis to address labeled anomaly data scarcity. The key strength of these augmentation strategies is their ability to synthesize realistic fault samples from limited real data, which the proposed framework leverages conceptually by constructing anomaly pattern libraries unsupervisedly.

3. Time-Series Similarity Measures
The anomaly pattern matching module employs a multiscale shape similarity metric that integrates DTW distance, trend correlation, curvature similarity, and event-based features. This composite measure is inspired by recent advances in time-series classification [22,31], which show that combining complementary shape aspects enhances discrimination performance. The strength of multi-aspect similarity metrics over single measures is their robustness: no single distance metric captures all relevant temporal characteristics, and their weighted combination compensates for individual weaknesses. Adapting these techniques to satellite telemetry enables not only anomaly detection but also fault type assignment—a significant improvement over binary detection.

3. Materials and Methods

This section presents a two-stage framework for thermal mode identification and fault detection in microsatellites. The methodology comprises five logical components: data preprocessing, unsupervised thermal operational mode discovery via two-stage clustering, construction of mode-specific CNN-LSTM predictors, residual-based fault detection, and anomaly pattern matching. Detailed parameter settings are deferred to Section 4 to maintain focus on methodological principles.

3.1. Two-Stage Overview

The framework operates in two sequential phases: unsupervised mode discovery followed by mode-conditioned anomaly detection, as illustrated in Figure 1. During the first phase, the satellite's orbital temperature cycles are partitioned into distinct operational regimes—two normal modes and four canonical fault-type libraries. During online monitoring, the current mode is inferred via nearest-neighbor DTW classification; the corresponding mode-specific model generates a prediction, and a fault is flagged when the prediction residual exceeds a mode-adaptive threshold. Flagged segments are then matched against pre-defined anomaly libraries for preliminary diagnosis.

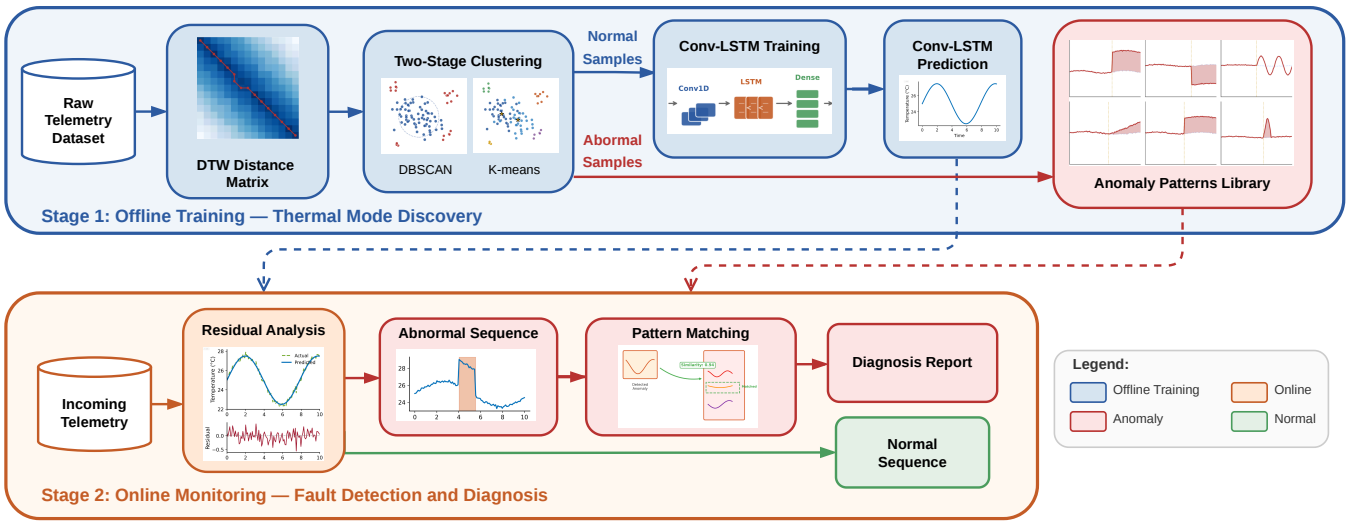


Figure 1. Two-stage framework for microsatellite thermal mode identification and fault detection.

3.2. Thermal Mode Discovery via Two-Stage Clustering

The unsupervised mode discovery relies on a robust similarity measure for orbital temperature cycles, specifically Dynamic Time Warping (DTW), which non-linearly aligns sequences and is insensitive to minor temporal shifts or phase variations between orbits. Pairwise DTW distances for all orbital cycles are computed using the fastdtw algorithm [21], generating a distance matrix.

A two-stage clustering procedure is applied. In the first stage, Density-Based Spatial Clustering of Applications with Noise (DBSCAN) is performed directly on the precomputed distance matrix. DBSCAN groups densely connected cycles (i.e., with small DTW distances) and labels isolated cycles as noise. Density parameters are set based on the DTW distances distribution. Resulting labels are binary: cycles in the main dense cluster are nominal, while noise-labeled cycles are potentially anomalous.

In the second stage, K-means clustering is applied separately to the nominal and anomalous subset, using their corresponding sub-matrices as features. Although K-means inherently assumes Euclidean geometry and centroid averaging, its application to the DTW sub-matrix is justified in this context for three reasons. First, DBSCAN has already performed the primary separation (nominal vs. anomalous) using a density-based criterion fully compatible with the precomputed distance matrix; K-means operates only on the resulting homogeneous subsets, where intra-cluster distances are relatively compact and the Euclidean approximation is more reasonable. Second, within each subset, the DTW distances exhibit near-Euclidean behavior: pairwise distances among nominal cycles are small and smoothly varying, and the same holds for anomalous cycles within each pattern group. Third, a comparative study using the precomputed DTW matrix confirmed that hierarchical clustering with average linkage yields the same two nominal clusters and highly similar anomalous partitions, validating the robustness of the K-means solution. Alternatives such as K-medoids or spectral clustering were also evaluated but offered no significant improvement in cluster separation while incurring substantially higher computational cost for the anomalous subset. For nominal cycles, the optimal cluster number (determined via the Elbow method and confirmed by physical interpretability) is two. These two clusters correspond to the satellite's two fundamental thermal operational states: low-power baseline mode (quiescent standby), and high-power operational mode (payload activation). Distinguishing these regimes is essential, as their nominal temperature ranges and temporal dynamics differ substantially.

For the anomalous subset, K-means with four clusters yields well-separated groups that correspond to four canonical fault types commonly observed in microsatellite thermal systems. Each fault type exhibits a distinct temporal morphology and is associated with a plausible in-orbit failure mechanism. It is important to note that the operational contexts assigned to these patterns are hypothesized interpretations inferred from cluster morphology rather than confirmed by ground-truth mission event logs; they should be understood as plausible scenarios pending independent validation. Table 2 summarizes the thermal characteristics and hypothesized operational contexts for each fault type.

3.3. Mode-Specific CNN-LSTM Prediction Models

For each normal thermal mode, a dedicated CNN-LSTM network is trained to predict the future temperature evolution from a historical window. The model operates on raw temperature values as two-dimensional tensor of shape (batch_size, sequence_length, 1). The architecture includes a 1D convolutional layer (Conv1D) to extract local temporal features, followed by an LSTM layer to capture long-range dependencies, and a dense output layer to produce a forecast sequence of length F. Conv1D filters count is tuned per

mode via cross-validation, while the kernel size and LSTM units are fixed. The complete set of hyperparameters is summarized in Table 3.

Table 2. Canonical fault types and hypothesized in-orbit mechanisms.

Fault Type	Temporal Morphology	Hypothesized In-Orbit Mechanism	Example Scenario
Step	Sustained constant offset from nominal value	Sudden change in thermal load or heater relay malfunction	Heater stuck on/off; payload power state latched
Spike	Brief, intense temperature excursion with rapid decay	Transient energy dissipation event	Electrical short-circuit; momentary overload during switching
Drift	Gradual, monotonic deviation from nominal trajectory	Slow degradation of thermal interface or sensor drift	Thermal pad aging; thermistor calibration shift; contact resistance increase
Noise	Persistent elevated random fluctuation around nominal mean	Intermittent electromagnetic or electrical interference	Switching power supply noise; ground loop; antenna coupling

Table 3. CNN-LSTM hyperparameter configuration.

Hyperparameter	Low-Power Mode	High-Power Mode
Conv1D filters	32	32
Conv1D kernel size	3	3
LSTM units	64	64
Dense units	864	864
Learning rate	0.001	0.001
Batch size	16	16
Patience (early stopping)	4	4

Each mode-specific model is trained to minimize the mean squared error (MSE) between predicted and true temperature values using the Adam optimizer with an initial learning rate of 0.001. Early stopping with a patience of 4 epochs is employed to prevent overfitting, and the best weights are restored. A single global CNN-LSTM model with identical architecture (32 Conv1D filters, kernel size 3, 64 LSTM units) is also trained on the chronologically interleaved concatenated data of both normal modes. This non-classified global method serves as a baseline to quantify the benefit of mode-specific specialization.

3.4. Residual-Based Fault Detection and Anomaly Pattern Matching

Upon downlinking telemetry data to the ground, the proposed framework rapidly analyzes satellite thermal status. For each incoming orbital cycle, the operational mode is first inferred by computing its DTW distance to the mean prediction curve of each nominal thermal regime, with the mode yielding the smallest distance selected. The corresponding mode-specific CNN-LSTM model then predicts the most recent historical window of prescribed length. The absolute prediction error at each time step is computed and smoothed via a moving average filter to suppress short-term fluctuations. A fault is flagged if the smoothed residual exceeds a mode-specific threshold, defined as a high percentile of the absolute error distribution observed on the validation set for that mode.

For each flagged anomalous segment, the framework identifies its fault type by matching against the pre-defined anomaly pattern libraries derived from the clustering stage. Let $Q = \{q_1, q_2, \ldots, q_n\}$ denote the anomalous segment and $P = \{p_1, p_2, \ldots, p_m\}$ denote a reference pattern from the library. Both sequences are first normalized to $[0, 1]$ via

min-max scaling. A multiscale shape similarity metric integrates four complementary temporal aspects:

Shape alignment (S_{shape}): The DTW distance $d_{DTW}(P, Q)$ quantifies overall morphological similarity while accommodating non-linear temporal warping. It is normalized to $[0, 1]$ via min-max scaling across all patterns in the library:

$$S_{shape}(P, Q) = 1 - \frac{d_{DTW}(P, Q) - d_{min}}{d_{max} - d_{min}} \tag{1}$$

where d_{min} and d_{max} denote the minimum and maximum DTW distances from Q to all reference patterns.

Trend correlation (S_{trend}): Pearson correlation coefficient between the first differences of P and Q, capturing directional agreement:

$$S_{trend}(P, Q) = \frac{\sum_{i=1}^{n-1}(P'_i - \bar{P}')(Q'_i - \bar{Q}')}{\sqrt{\sum_{i=1}^{n-1}(P'_i - \bar{P}')^2} \cdot \sqrt{\sum_{i=1}^{n-1}(Q'_i - \bar{Q}')^2}} \tag{2}$$

where $P' = \{p_2 - p_1, p_3 - p_2, \ldots\}$ and $Q' = \{q_2 - q_1, q_3 - q_2, \ldots\}$ denote first differences, and $\bar{P}'$, $\bar{Q}'$ their respective means.

Curvature similarity ($S_{curvature}$): Correlation of second differences, characterizing convexity/concavity patterns that indicate inflection points:

$$S_{curvature}(P, Q) = \max(0, \text{corr}(P'', Q'')) \tag{3}$$

where P'' and Q'' denote second differences, and negative correlations are clipped to zero to penalize opposing curvature patterns.

Event-based similarity (S_{event}): Captures alignment of abrupt change events via normalized overlap of significant derivative magnitudes:

$$S_{event}(P, Q) = \frac{|E_P \cap E_Q| + \epsilon}{|E_P \cup E_Q| + \epsilon} \tag{4}$$

where $E_P = \{i : |P'_i| > \tau_P\}$ and $E_Q = \{i : |Q'_i| > \tau_Q\}$ are sets of event indices with derivative magnitude exceeding threshold τ (set to the 80th percentile of $|P'|$ and $|Q'|$, respectively), and ϵ is a small constant to prevent division by zero.

The four components are aggregated into a total similarity score via weighted convex combination:

$$S(P, Q) = w_1 S_{shape} + w_2 S_{trend} + w_3 S_{curvature} + w_4 S_{event} \tag{5}$$

subject to $w_1 + w_2 + w_3 + w_4 = 1$. Weights are determined empirically via grid search on labeled anomalous segments. The reference pattern yielding the highest similarity score is selected. If $S(P^*, Q) \geq \theta$ where θ is a preset acceptance threshold, the segment is diagnosed with the fault type associated with pattern P^*; otherwise, it remains uncategorized, indicating a potentially novel anomaly type.

4. Data

This section describes the satellite telemetry dataset used to develop and validate the proposed framework. It covers the data source, sensor selection rationale, preprocessing configuration, two-stage clustering parameterization, and fault detection/pattern matching thresholds and weights. All numerical settings are critical for results reproducibility.

4.1. Data Source and Sensor Selection

Experimental data were acquired from Satellite B of the Macau Science Satellite-1 mission (MSS-1B, COSPAR 2023-069-B, NORAD 56732) as illustrated in Figure 2, a low-Earth-orbit scientific satellite with an initial orbit of approximately 500 km × 450 km altitude and 41° inclination, co-developed by the Chinese mainland and Macau for solar X-ray and medium-energy particle detection. Satellite B is equipped with space environment payloads; its precise pointing and frequent mission mode switches impose stringent thermal control requirements.



Figure 2. Satellite B of the Macau Science Satellite-1 mission.

This study focuses on the critical temperature sensor (telemetry channel TK46) mounted near a high-power communication module on the satellite platform. This location was selected due to three reasons: the module undergoes frequent activation/deactivation, causing rapid power dissipation variations; the sensor is directly exposed to the orbital sunlight-eclipse cycle-induced external heat flux changes; and historical telemetry data shows this channel exhibits the widest dynamic range and the clearest separation between operational modes.

The raw dataset comprises approximately six months of continuous in-orbit operation covering diverse mission scenarios including routine Earth observation, high-speed data downlink, and attitude maneuvers. This period was selected because it contains the richest operational mode diversity and the most frequent mode transitions, providing a compact yet representative dataset for framework development and evaluation. Telemetry is sampled at 1 min intervals. Descriptive statistics of the raw temperature record over this period are summarized in Table 4. A contiguous seven-day segment drawn from a later period is reserved as an independent validation set, excluding it from training and clustering procedures. This segment, representative of routine operations with no known hardware failures, serves as an ideal benchmark for false-alarm rate assessment. The chronological partitioning ensures that clustering and CNN-LSTM training are performed exclusively on the earlier period; the held-out test period is used only for final evaluation and is never accessed during model development, threshold calibration, or weight tuning.

Table 4. Raw temperature telemetry summary statistics (training period).

Statistic	Value (°C)
Mean	10.14
Std. dev.	1.07
Minimum	7.98
Maximum	20.28
5th percentile	8.80
95th percentile	11.66

4.2. Data Preprocessing Parameters

Raw temperature telemetry, originally sampled at 1 min intervals, is first downsampled to a uniform 5 min interval via linear interpolation. This resampling reduces short-term noise and computational burden while preserving the dominant orbital cycle thermal dynamics (orbital period $\approx$ 90 min, yielding $\approx$ 18 points/orbit). However, it inevitably smooths transient thermal events shorter than five minutes, which may attenuate or miss sub-five-minute spikes; this trade-off is discussed further in Section 6. The continuous temperature stream is then segmented into individual orbital cycles using precise satellite ephemeris and eclipse flags, with each cycle treated as an independent sample. Cycles containing telemetry gaps longer than 10 min ($\leq$1% total) are discarded. No amplitude normalization is applied, as empirical experiments showed raw temperature values yield better model convergence and prediction accuracy.

For supervised model training, all orbital cycles belonging to the same nominal thermal mode (identified via clustering) are concatenated chronologically to form a single long time series, preserving temporal order. Sliding windows are then extracted from this sequence to generate input–output pairs. Windowing parameters were selected to balance three criteria: (1) sufficient length to capture complete diurnal and orbital thermal cycles; (2) adequate forecast lead time for ground operators to respond; and (3) minimal redundancy while generating sufficient training samples.The resulting parameters are given in Table 5.

Table 5. Sliding window parameters.

Parameter	Value	Physical Meaning
Historical window (W)	2016 pts	7 days
Forecast horizon (F)	864 pts	3 days
Sliding step (S)	5 pts	25 min overlap

4.3. Clustering Configuration

For the first clustering stage, DBSCAN is applied directly to the distance matrix. The neighborhood radius is chosen based on the distribution of pairwise DTW distances to effectively isolate cycles with unusual thermal profiles while preserving the dense main cluster. The minimum samples parameter is set to suppress spurious micro-clusters. In the second stage, K-means clustering is performed separately on the nominal and anomalous subsets. The optimal number of clusters for each subset is determined via the Elbow criterion and validated by physical interpretability. The complete set of clustering parameters is summarized in Table 6. A sensitivity analysis indicated stable cluster assignments for the number of anomalous clusters in the range $[5, 7]$.

Table 6. Clustering Parameters.

Stage	Algorithm	Parameter	Value
Stage 1	DBSCAN	eps	5.0
		min_samples	10
Stage 2	K-means	n_clusters (normal)	2
		n_clusters (anomalous)	4

4.4. Fault Detection and Pattern Matching Configuration

For fault detection, the mode-specific threshold is set at the 95th percentile of the absolute prediction error distribution observed on the validation set, with prediction residuals smoothed using a 10-point moving average window. For anomaly pattern matching, the multiscale shape similarity metric defined in Section 3.4 employs a two-level weighting scheme. At the inner level, the four shape components are weighted as $w_1 = 0.3$, $w_2 = 0.3$, $w_3 = 0.2$, and $w_4 = 0.2$; at the outer level, the combined shape similarity, event similarity, and statistical similarity are fused with weights $\alpha_1 = 0.6$, $\alpha_2 = 0.3$, and $\alpha_3 = 0.1$, respectively. These weights were determined via grid search on a set of 12 manually verified anomalous segments spanning different pattern types, with the inner-level weights searched over $\{0.1, 0.2, 0.3, 0.4, 0.5\}$ (subject to the sum-to-one constraint) and the outer-level weights over $\{0.1, 0.2, \ldots, 0.8\}$. Shape similarity was prioritized to emphasize morphological correspondence over numerical proximity. The selected weights were stable across different random subsets of the tuning segments, indicating robustness to tuning set composition. The event detection threshold τ in Equation (4) is set to the 80th percentile of derivative magnitudes. A matching is accepted when the similarity score $S(P, Q) \geq 0.7$, a threshold that maximizes the F1 score on the tuning set.

4.5. Experimental Environment

All experiments were conducted on a workstation with an Intel Xeon W-2245 CPU, 64 GB of RAM, and an NVIDIA GeForce RTX 3060 GPU. Deep learning models were implemented and trained using a widely adopted scientific computing framework. Preprocessing, clustering, and evaluation routines were executed in a unified computational environment with double-precision floating-point arithmetic.

5. Results

This section evaluates the proposed framework on a held-out validation dataset containing no known hardware failures—an ideal benchmark for false-alarm assessment and detection capability verification.

5.1. Overall Fault Detection Performance

Table 7 summarizes the reduction in anomaly flags and prediction error improvement achieved by the two-stage framework compared to the non-classified global baseline on the nominal-condition validation set. Mode-specific predictors capture the distinct dynamics of each thermal regime far more faithfully than a single global model. It is important to note that, since the validation set contains no injected or confirmed hardware failures, the 96.6% reduction in anomaly flags primarily reflects spurious-alarm suppression; the four flags produced by the two-stage method correspond to a known operational transient (data downlink spike), as discussed in Section 5.2, confirming genuine detection capability.

The prediction mean absolute error (MAE) is also substantially lower for the two-stage method (0.225 vs. 0.462, a 51.3% improvement), indicating that mode-specific models capture thermal dynamics more faithfully.

Table 7. Fault detection and prediction accuracy comparison.

Method	Detected Anomaly Points
Two-stage (proposed)	4
Non-classified global	148
Improvement	96.6%

5.2. Prediction Accuracy and Error Analysis

Figure 3 compares the two-stage and global methods in terms of fault detection count and prediction accuracy. The two-stage method flags only 4 anomaly points vs.148 by the global method, a 96.6% reduction in anomaly flags under nominal operating conditions. The four flagged points correspond to a known operational transient (Section 5.2), confirming genuine detection rather than false alarming. The prediction MAE decreases from 0.462 to 0.225, a 51.3% improvement.

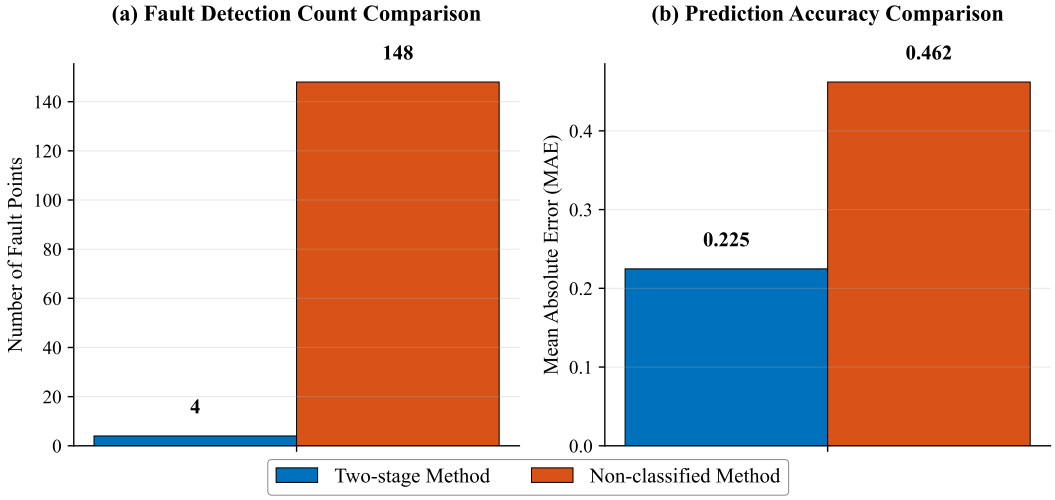


Figure 3. Performance comparison between two-stage and non-classified methods: (**a**) fault detection count, (**b**) prediction MAE.

Figure 4 presents the prediction error distributions. The two-stage method yields a markedly tighter error concentration than the global baseline, confirming that mode-conditioned predictions are both more accurate and more stable.

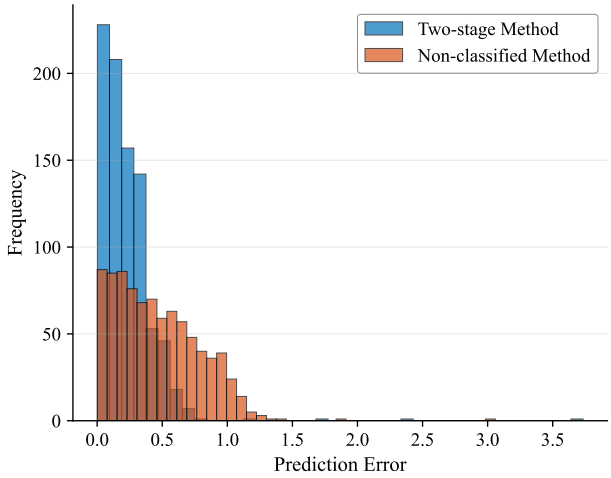


Figure 4. Prediction error distribution. The two-stage method (blue) has errors concentrated near zero, while the global method (orange) shows wider dispersion.

Figure 5 displays a representative segment of the true temperature curve alongside the predictions from the two-stage method and the global baseline. The two-stage predictions closely track the true profile throughout, including mode transition. However, the global predictor exhibits a noticeable lag and systematic under prediction during high power mode, confirming that a single model cannot adequately represent both thermal regimes.

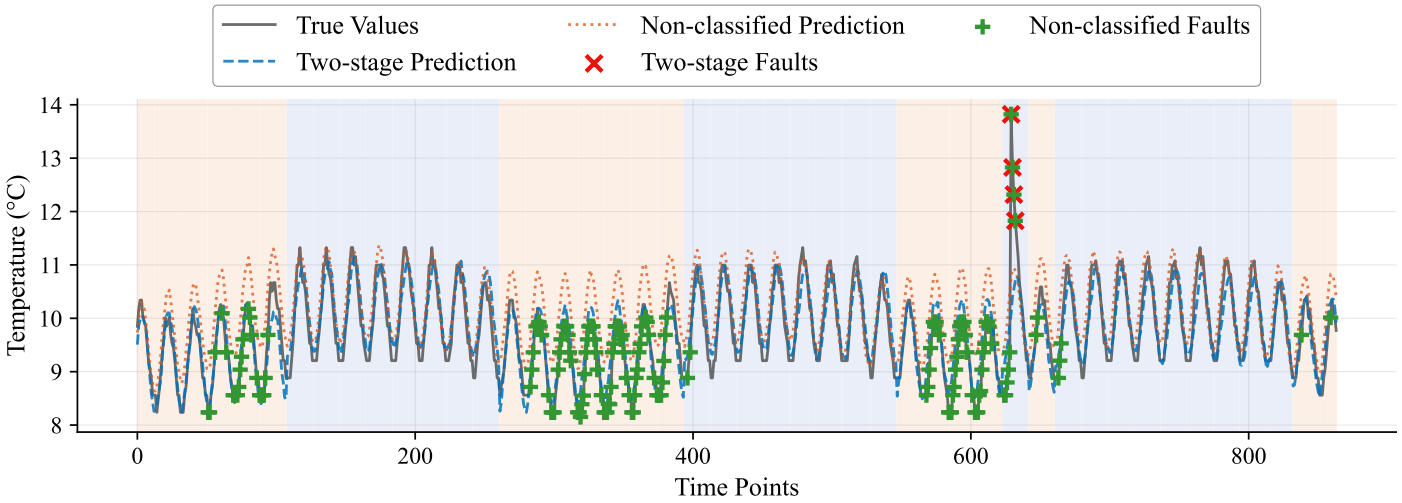


Figure 5. Segmented temperature telemetry with overlaid predictions and detected fault points. The two-stage method (red ×) correctly detects a real operational transient while producing substantially fewer spurious flags than the non-classified global baseline (green +), which misinterprets normal inter-mode transitions as anomalies.

5.3. Anomaly Pattern Matching and Diagnostic Capability

Beyond binary detection, the proposed framework categorizes flagged anomalous segments by matching against the fault-type libraries constructed during clustering. During validation, a segment from the low-power baseline mode exhibited four consecutive anomalous points. Using the multiscale shape similarity metric, this segment is matched to a spike-type anomaly with a combined similarity score of 0.867. The matched pattern is characterized by a brief, intense temperature excursion, consistent with the transient thermal spike observed during data downlink activation—a known operational event in the validation period. This correlation between a detected anomaly and a confirmed operational transient provides additional confidence in the framework's detection capability, though precise causal attribution without dedicated ground-truth instrumentation warrants further investigation.

Figure 6 compares the true temperature sequence, the two-stage prediction, and the matched anomaly pattern curve, showing high morphological correspondence. This result demonstrates that the framework not only detects anomalies with high precision but also provides actionable diagnostic information beyond a simple binary alert.

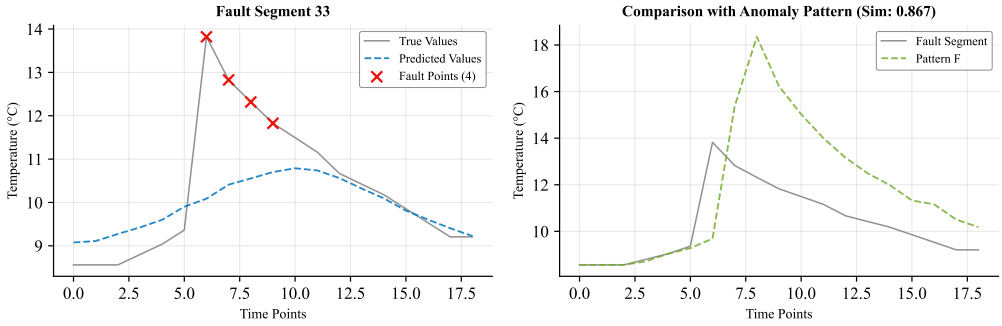


Figure 6. Anomalous segment matched to spike-type fault with similarity 0.867.

5.4. Visualization of Fault Segments and Mode Distribution

Figure 7 visualizes the mode distribution and fault attribution: the detected fault segment lies within the low-power mode cluster but is clearly separated from normal samples, confirming the pattern matching mechanism's ability to identify intra-mode deviations.

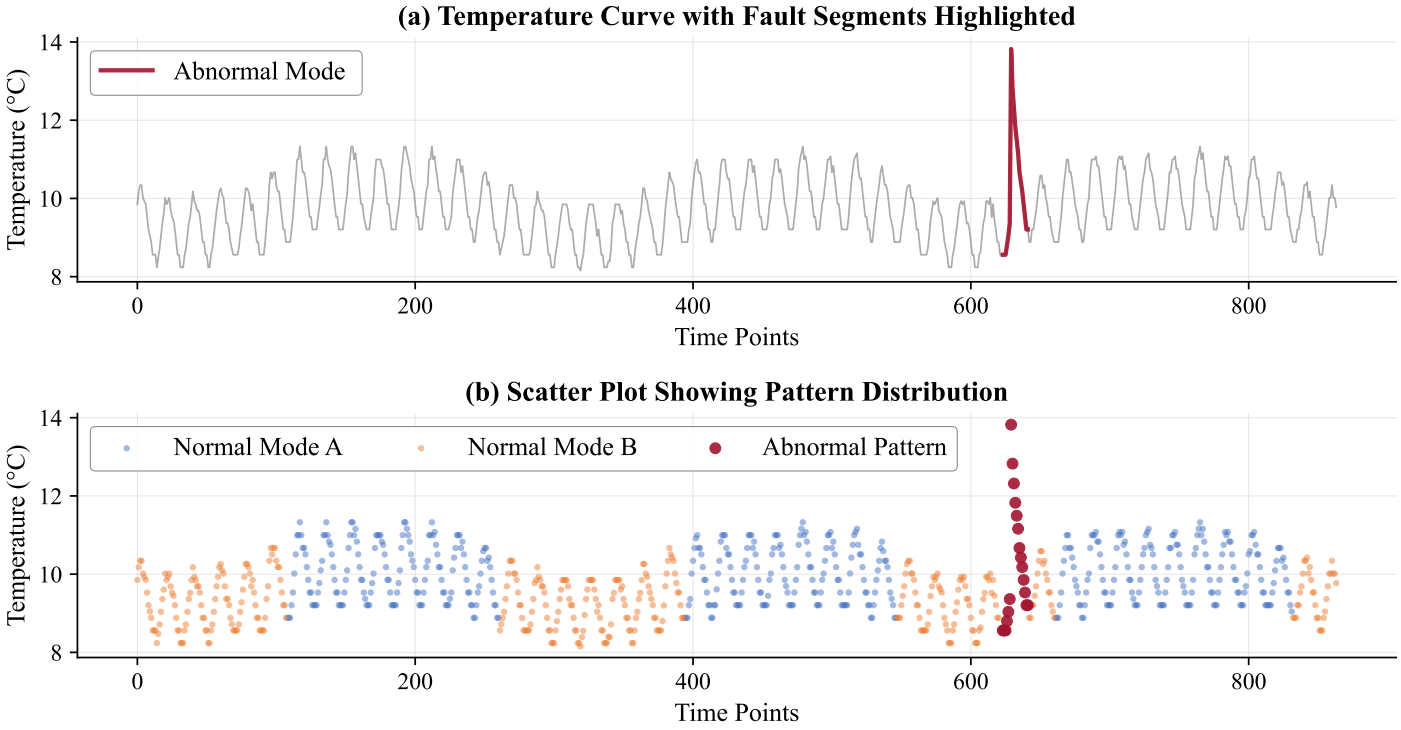


Figure 7. Top (**a**): Temperature curve with fault segments highlighted by matched anomaly pattern color. Bottom (**b**): Scatter plot of pattern mode distribution; normal samples (small dots) are clearly separated by mode, and fault samples (large dots) are correctly attributed to their respective anomaly patterns.

In summary, the two-stage method substantially reduces false alarms and improves prediction accuracy relative to a global baseline. Mode-specific predictors faithfully capture the distinct dynamics of each thermal regime, and detected anomalies are successfully matched to meaningful fault patterns. These findings robustly support the 'identify first, detect second' paradigm for microsatellite thermal monitoring.

6. Discussion

The experimental results demonstrate that the proposed two-stage framework effectively mitigates the core limitation of conventional microsatellite temperature monitoring: conflation of normal operational mode transitions with genuine anomalies. This section elucidates the underlying mechanisms of the observed performance gains, discusses the diagnostic utility of anomaly pattern matching, contextualizes the contribution within the existing literature, acknowledges current limitations, and outlines future research directions.

6.1. Performance Gains Interpretation

The results in Section 5 demonstrate that the two-stage framework effectively disentangles normal operational transients from genuine anomalies. This section interprets the underlying mechanisms behind these performance gains.

The 'identify first, detect second' strategy works because traditional threshold-based and global data-driven models lack contextual awareness: they interpret any deviation from a fixed baseline or a global average as anomalous. In a microsatellite, however, temperature

fluctuations of several degrees Celsius are normal during transitions from low-power baseline state to a high-power operational phase. By first applying unsupervised clustering to identify two dominant thermal regimes, the proposed framework establishes the context against which residuals are evaluated. This divide-and-conquer strategy resolves the conflation that a single non-classified global model cannot.

The prediction MAE improvement similarly stems from eliminating the averaging effect inherent in global models. A global CNN-LSTM trained on interleaved multi-regime data must approximate a multi-modal function, inevitably producing intermediate predictions that are accurate for neither regime. This is visually evident in Figure 5, where the global predictor systematically under-predicts during high-power mode peaks and over-predicts during low-power mode troughs. Each mode-specific predictor, in contrast, operates on a homogeneous data distribution, enabling capture of finer temporal dynamics and state-specific dependencies.

6.2. Synthetic Fault Injection Evaluation

While the nominal-condition validation confirms spurious-alarm suppression (the two-stage method's four flags correspond to a real operational transient, as noted in Section 5.2), quantifying detection sensitivity requires ground-truth fault labels unavailable from real telemetry. To address this, a synthetic fault injection experiment was conducted by embedding four types of artificial anomalies into the validation time series at random positions. Each fault type corresponds to one of the four canonical fault categories identified during clustering (Table 2):

- **Step**: A constant temperature offset added over a contiguous interval, simulating heater relay malfunction or payload power state latching. Magnitudes: 2–5 °C; durations: 10–50 points (50–250 min).
- **Spike**: An intense temperature excursion over the first 3 points, simulating transient electrical overload. Magnitudes: 3–8 °C.
- **Drift**: A linearly increasing offset simulating gradual sensor degradation or thermal contact deterioration. Magnitudes: 2–6 °C (terminal value); durations: 50–200 points (4–17 h).
- **Noise**: Gaussian noise (σ = magnitude/2) superimposed over an interval, simulating electromagnetic interference. Magnitudes: 1.5–5 °C; durations: 20–100 points.

For each fault type, nine combinations (3 magnitudes × 3 durations) were tested at a random injection position, yielding 36 fault instances. Detection performance was evaluated using precision (p), recall (r), and F1 score, comparing the proposed two-stage method against five baselines: (1) global CNN-LSTM (same architecture, trained on interleaved data); (2) fixed-threshold (3σ sliding window); (3) IQR (1.5 × IQR rule); (4) Isolation Forest (sklearn, 48-point feature window, contamination=0.05); and (5) Moving Average (288-point window, same 0.8 °C threshold as the two-stage method).

6.2.1. Overall Detection Performance

Table 8 presents the overall detection metrics averaged across all fault types and configurations. The two-stage method achieves an F1 of 0.725, a 2.8× improvement over the global CNN-LSTM (0.258). This advantage is driven primarily by precision: the two-stage method yields p = 0.749 vs. p = 0.185 for the global baseline—a 4× improvement. Recall values are comparable (0.804 vs. 0.816), indicating that both methods detect a similar fraction of true faults, but the global baseline produces overwhelming false alarms that severely degrade precision. The three statistical baselines (fixed-threshold, IQR, Isolation Forest) achieve F1 scores of 0.21–0.16, confirming that rule-based approaches are inadequate for the subtle, context-dependent anomalies in satellite telemetry. Notably, the Moving Average

method uses the identical 0.8 °C threshold as the two-stage method yet achieves only F1 = 0.085, demonstrating that the performance gain stems from mode-specific prediction rather than threshold selection.

Table 8. Overall fault detection performance (synthetic injection).

Method	Precision	Recall	F1
Two-stage (Ours)	0.749	0.804	0.725
Global CNN-LSTM	0.185	0.816	0.258
Fixed-threshold (3σ)	0.267	0.307	0.210
IQR-based ($1.5 \times$ IQR)	0.227	0.362	0.216
Isolation Forest	0.119	0.293	0.158
Moving Average (0.8 °C)	0.049	0.624	0.085

6.2.2. Per-Fault-Type Analysis

Table 9 breaks down detection performance by fault type. Step anomalies are best detected (F1 = 0.884) owing to their sustained, unambiguous deviation from the predicted trajectory. Spike anomalies achieve perfect recall ($r = 1.000$) but lower precision ($p = 0.429$), as the extremely short fault duration (3 points) means that even a small number of co-occurring false positives significantly degrades precision. Drift anomalies yield the highest precision ($p = 0.945$) among all types, as gradual deviations, once they exceed the threshold, are rarely misclassified; however, recall is lower ($r = 0.693$) because the initial, low-magnitude phase of the drift falls within the 0.8 °C tolerance. Noise anomalies are the most challenging (F1 = 0.625), as random fluctuations occasionally produce values close to the prediction by chance, reducing both precision and recall.

Table 9. Per-fault-type detection performance (two-stage method).

Fault Type	Precision	Recall	F1
Step	0.801	1.000	0.884
Spike	0.429	1.000	0.600
Drift	0.945	0.693	0.791
Noise	0.822	0.524	0.625

The global CNN-LSTM exhibits a qualitatively different failure mode: high recall but extremely low precision across all fault types. This confirms that the global model is sensitive to deviations but cannot distinguish genuine faults from normal inter-mode transitions, resulting in chronic false-alarm contamination. The two-stage method resolves this by conditioning detection on the inferred operational mode, which provides the contextual baseline against which residuals are evaluated.

6.2.3. Fault Parameter Sensitivity

To characterize the conditions under which mode-specific context is most beneficial, a finer 5×5 parameter grid (5 magnitudes $\times$ 5 durations) was evaluated for each fault type, comparing the two-stage method against the global CNN-LSTM. Figure 8 presents the F1 difference ($\Delta F1 = F1_{\text{two-stage}} - F1_{\text{global}}$) as a function of fault magnitude and duration.

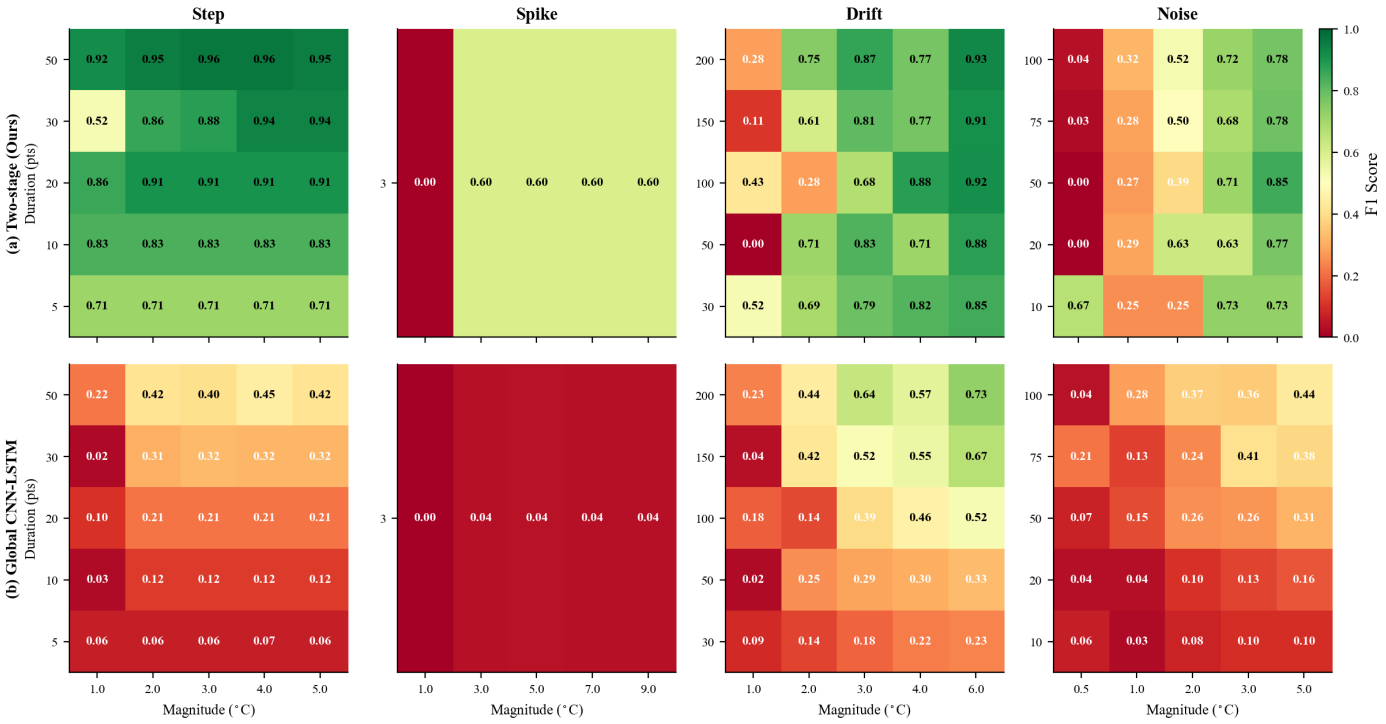


Figure 8. F1 advantage of the two-stage method over the global CNN-LSTM (ΔF1) as a function of fault magnitude and duration for step, spike, drift, and noise anomalies. Green indicates two-stage superiority; larger values indicate greater advantage.

The sensitivity analysis reveals two key findings. First, the two-stage advantage is most pronounced for **short-duration** faults. For step anomalies with duration ≤20 points, ΔF1 exceeds 0.65, whereas for longer durations (≥100 points), it decreases to 0.5–0.55 as the global baseline also begins to detect the sustained deviation. Second, for **drift** anomalies, the two-stage method maintains a consistent advantage (ΔF1 ≈ 0.5–0.6) for magnitudes ≥2 °C, but both methods struggle at 1 °C where the average deviation per point is below the 0.8 °C threshold. This boundary is not a methodological artifact but reflects a fundamental detectability limit: faults whose mean per-point deviation is smaller than the detection threshold are inherently difficult to identify with any residual-based approach. The noise anomaly sensitivity shows a similar pattern, with the two-stage advantage growing steadily from ΔF1 ≈ 0.06 at 0.5 °C to ΔF1 ≈ 0.50 at 5 °C.

6.3. Diagnostic Value of Anomaly Pattern Matching

Beyond accurate detection, the proposed framework introduces preliminary diagnostic capability by matching flagged anomalous segments to a library of recurrent thermal patterns derived from historical data. Successful matching of the detected segment to a distinct anomaly pattern (similarity 0.867) demonstrates that the multiscale shape similarity metric effectively quantifies morphological correspondence. This transforms generic anomaly alert into actionable information guiding operators toward appropriate response actions.

Pattern libraries are constructed unsupervisedly during clustering, requiring no manual labeling. They emerge naturally as clusters of cycles distinct from nominal regimes. This self-supervised property is a key framework strength, circumventing the scarcity of labeled fault data in operational satellite telemetry.

6.4. Comparison with Existing Methods

The results align with and extend prior work in key aspects [32–34]. Dynamic Time Warping (DTW) effectiveness for time-series clustering in satellite telemetry, previously

validated by Cui et al. [21], is corroborated by its successful application in identifying physically meaningful thermal modes. Leveraging and refining Long Short-Term Memory (LSTM) networks for spacecraft system dynamics capture [14,15], the proposed approach replaces a single global LSTM with multiple mode-conditioned CNN-LSTM predictors, each specialized for a homogeneous regime. Compared to single LSTM satellite maneuver detection studies [16], this federated framework achieves substantially higher accuracy and robustness for cyclical, multi-modal systems.

To the best of the authors' knowledge, this work represents one of the first frameworks explicitly integrating unsupervised mode discovery with supervised, mode-specific deep learning for satellite telemetry anomaly detection. Previous data-driven methods either focused on binary classification without contextual modeling, or relied on supervised segmentation requiring expensive manual operational mode labeling. The two-stage clustering procedure eliminates this labeling burden while providing the context necessary for precise, regime-aware monitoring.

6.5. Sensitivity Analysis and Baseline Comparison

6.5.1. DBSCAN Parameter Sensitivity

A systematic sensitivity analysis was conducted by varying the DBSCAN neighborhood radius ε from 3.0 to 8.0 (in steps of 0.5) and `min_samples` from 5 to 15, yielding 60 parameter combinations. Across the default parameter region ($\varepsilon \in [4.5, 6.0]$, `min_samples` $\in [8, 12]$), DBSCAN consistently produces a single dense nominal cluster with 9–11% of cycles labeled as noise (anomalous). This stable separation is critical: it ensures that the downstream K-means always receives the same qualitative partition—a compact nominal set and a scattered anomalous set—regardless of small perturbations in ε or `min_samples`. The two nominal thermal modes (low-power and high-power) then emerge reliably from K-means applied to the dense cluster, confirming that the overall two-stage discovery is robust to DBSCAN parameter variation within a reasonable range. Only when ε is reduced below 4.0 does DBSCAN begin to fragment the nominal cluster, yielding two to three sub-clusters and elevated noise ratios (>13%), indicating that overly small neighborhood radii are inappropriate for this dataset.

6.5.2. Baseline Comparison on Validation Set

Table 10 compares six detection strategies on the 864-point validation set (3-day prediction horizon). The fixed-threshold method (3σ sliding window over the preceding day) and the interquartile range (IQR) method represent conventional rule-based approaches. Isolation Forest and Moving Average represent classical machine learning and simple statistical baselines, respectively. The global CNN-LSTM and the proposed two-stage method represent data-driven approaches.

The validation segment contains no injected or confirmed hardware faults but includes a known operational transient—a brief temperature spike caused by data downlink activation, previously identified and analyzed in Section 5.2. The two-stage method correctly flags four points corresponding to this real event, demonstrating genuine detection capability rather than false alarming. In contrast, Isolation Forest flags zero points, entirely missing this real transient; the fixed-threshold and IQR methods flag only three points each, partially capturing the event with limited sensitivity. Notably, while these three methods produce fewer total flags than the two-stage method, they achieve drastically lower F1 scores on injected faults (Table 8), confirming that their low flag counts reflect insensitivity rather than precision. The Moving Average and global CNN-LSTM methods produce 204 and 148 flags respectively—overwhelmingly spurious alarms arising from misinterpretation of normal inter-mode transitions. The two-stage method thus uniquely

achieves both correct detection of real transients and low spurious flagging, a balance that neither the insensitive statistical methods nor the over-sensitive global model can attain.

Table 10. Baseline comparison on validation set (864-point prediction horizon). The segment contains no injected hardware faults but includes one known operational transient (data downlink spike).

Method	Anomaly Flags	Flag Rate (%)	MAE (°C)
Fixed-threshold (3σ)	3	0.35	—
IQR-based ($1.5 \times$ IQR)	3	0.35	—
Isolation Forest	0	0.00	—
Moving Average (0.8 °C)	204	23.61	—
Global CNN-LSTM	148	17.13	0.462
Two-stage (Ours)	4	0.46	0.225

6.6. Limitations and Future Work

Despite promising results, this study has limitations. The anomaly pattern library, constructed unsupervisedly, currently contains four canonical fault types, whose representativeness for novel unseen fault types remains insufficiently evaluated. As on-orbit fault data accumulate, the library should be expanded and validated against a broader set of real anomalies. Exploring techniques like few-shot learning or open-set recognition could enable the framework to identify previously unencountered patterns.

The computational footprint of training and maintaining multiple CNN-LSTM models is non-negligible. While acceptable for ground-based postprocessing or periodic retraining, rapid telemetry analysis via the full framework would benefit from optimization. Future work should explore model compression and lightweight architectures to reduce resource consumption.

This study focuses on a single temperature sensor, providing a clear proof of concept. However, a comprehensive satellite health monitoring system requires integrating multi-sensor data streams, including power, current, and attitude measurements. Extending the framework to a multivariate setting presents challenges in clustering (requiring a suitable similarity measure for multivariate time series) and prediction (needing cross-sensor dependency modeling). Graph neural networks or attention-based multivariate forecasting models may address these challenges.

Several hyperparameters in the framework, including DBSCAN neighborhood radius, K-means clusters count, matching threshold, and similarity weights, were determined empirically from available data. While the DBSCAN sensitivity analysis in Section 6.4 confirms stable cluster separation across a reasonable parameter range on this dataset, systematic evaluation across different satellite missions would further strengthen generalizability claims. Future work should also explore data-driven threshold adaptation, dynamically adjusting fault detection thresholds using recent prediction error statistics to reduce manual calibration requirements.

Five-minute resampling effectively reduces noise and computational load, but inevitably smooths transient thermal events, potentially attenuating or missing spikes shorter than five minutes. A rigorous frequency-domain analysis quantifying the information loss and a systematic evaluation of detection sensitivity as a function of sampling rate are deferred to future work. For critical detection of very short-duration anomalies, a higher sampling rate may be required. A trade-off analysis between sampling frequency and detection sensitivity will guide practical deployment for various mission requirements.

The synthetic fault injection evaluation, while providing quantitative P/R/F1 metrics under controlled conditions, cannot fully substitute for validation on real fault data. Low-magnitude faults (drift < 1 °C, noise < 0.5 °C) remain difficult for all methods, representing

a fundamental detectability boundary rather than a framework limitation. Future work should seek opportunities to validate against confirmed in-orbit anomalies and investigate adaptive thresholding to improve sensitivity for gradual faults.

Finally, integrating semi-supervised and active learning strategies enables continuous refinement of mode discovery and pattern matching modules with new telemetry, advancing the framework toward a fully autonomous and adaptive satellite health management system capable of evolving with the mission.

7. Conclusions

Context-aware monitoring that separates normal operational regimes from fault-like transients can drastically reduce false alarms in microsatellite thermal telemetry. In this work, a two-stage framework was developed to achieve this: unsupervised clustering first partitions orbital temperature cycles into two nominal modes and four canonical fault-type libraries (step, spike, drift, and noise), then mode-specific CNN-LSTM predictors enable high-fidelity forecasting and residual-based anomaly detection. Experimental validation on approximately two months of in-orbit Macau Science Satellite-1B (COSPAR 2023-069-B, NORAD 56732) telemetry yielded a 96.6% reduction in anomaly flags and a 51.3% improvement in prediction mean absolute error compared to a conventional non-classified global model under nominal operating conditions, confirming effective spurious-alarm suppression while correctly detecting a real operational transient.

A synthetic fault injection experiment with four anomaly types and five baseline methods further quantified the framework's detection capability. The two-stage method achieved an overall F1 score of 0.725, a 2.8× improvement over the global CNN-LSTM baseline (F1 = 0.258), driven primarily by a 4× precision gain (0.749 vs. 0.185). This confirms that mode-specific context dramatically reduces false alarms without sacrificing detection sensitivity. Parameter sensitivity analysis across 80 fault configurations revealed that the two-stage advantage is most pronounced for short-duration and low-magnitude faults, where mode-specific prediction is essential for distinguishing genuine anomalies from normal inter-mode transitions. Detected anomalous segments were furthermore matched to distinct fault patterns, demonstrating the framework's preliminary diagnostic capability.

Several limitations temper the generalizability of these findings, including single-sensor validation, the absence of real fault ground truth, hypothesized anomaly interpretations, and the smoothing effect of 5 min downsampling; these are discussed in detail in Section 6. Future efforts will target validation on longer telemetry records spanning full seasonal cycles, multivariate telemetry integration, real in-orbit fault validation, adaptive thresholding for improved drift sensitivity, and validation across multiple satellite platforms to support fully autonomous onboard implementation.

Author Contributions: Conceptualization, W.P. and J.X.; methodology, W.P.; software, W.P.; validation, W.P.; formal analysis, W.P.; investigation, W.P.; resources, J.Z. and X.Z.; data curation, W.P.; writing—original draft preparation, W.P.; writing—review and editing, W.P.; visualization, W.P.; supervision, J.Z.; project administration, J.Z.; funding acquisition, J.Z. and X.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Ningbo Commonweal Programme, Zhejiang Province, under Grant 2025S061.

Data Availability Statement: The data presented in this study are available on request from the corresponding author due to privacy restrictions.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

DTW	Dynamic Time Warping
CNN	Convolutional Neural Network
LSTM	Long Short-Term Memory
DBSCAN	Density-Based Spatial Clustering of Applications with Noise
MAE	Mean Absolute Error
MSE	Mean Squared Error
MSS-1	Macau Science Satellite-1

References

1. Heidt, H.; Puig-Suari, J.; Moore, A.; Nakasuka, S.; Twiggs, R. CubeSat: A new generation of picosatellite for education and industry low-cost space experimentation. In Proceedings of the 14th Annual AIAA/USU Conference on Small Satellites, Logan, UT, USA, 21–24 August 2000; pp. 1–8.
2. Sandau, R.; Esper, J.; Paxton, L.; Briess, K.; Bartholomé, E.; Belward, A.; Boshuizen, C.; Carmona-Moreno, C.; Contant, J.-M.; Cutter, M.; et al. *IAA Study Group on Cost Effective Earth Observation Missions, IAA Position Paper on Cost Effective Earth Observation Missions*; IAA: Paris, France, 2004.
3. Weston, S.V.; Burkhard, C.D.; Stupl, J.M.; Ticknor, R.L.; Yost, B.D.; Austin, R.A.; Galchenko, P.; Newman, L.K.; Soto, L.S. State-of-the-Art Small Spacecraft Technology. 2025. Available online: https://ntrs.nasa.gov/citations/20250000142 (accessed on 5 May 2026).
4. Sannino, A.; Mungiguerra, S.; Cassese, S.; Savino, R.; Fedele, A.; Natalucci, S. Fast reconfiguration maneuvers of a micro-satellite constellation based on a hybrid rocket engine. *Aerotec. Missili Spaz.* **2024**, *103*, 401–412. [CrossRef]
5. Kulu, E. Cubesats & nanosatellites—2024 statistics, forecast and reliability. In Proceedings of the International Astronautical Congress, IAC, Milan, Italy, 14–18 October 2024; Volume 2024.
6. Christensen, C.B.; Herrera, C.R.; Mullins, C. Smallsats by the numbers: Growing smallsat activity and expanding commercial capabilities. In Proceedings of the ASCEND 2023, Las Vegas, NV, USA, 23–25 October 2023; p. 4677.
7. Pourtakdoust, S.H.; Fakhari Mehrjardi, M.; Hajkarim, M.H.; Nasihati Gourabi, F. Advanced fault detection and diagnosis in spacecraft attitude control systems: Current state and challenges. *Proc. Inst. Mech. Eng. Part G J. Aerosp. Eng.* **2023**, *237*, 2679–2699. [CrossRef]
8. Gao, Z.; Qiao, K.; Bai, J.; Wang, Z.; Liu, H.; Li, P. Numerical investigation of natural circulation loop with supercritical CO_2 on the thermal control system of micro spacecrafts. *Int. J. Heat Mass Transf.* **2023**, *217*, 124661. [CrossRef]
9. Khan, S.A.; Ali, A.; Yang, S.; Tong, J.; Guerrero, J.M. Optimized design of embedded air coil for small satellites with various dimensions. *J. Aerosp. Inf. Syst.* **2021**, *18*, 269–279. [CrossRef]
10. Crotti, E.; Colagrossi, A. Machine Learning Approaches for Data-Driven Self-Diagnosis and Fault Detection in Spacecraft Systems. *Appl. Sci.* **2025**, *15*, 7761. [CrossRef]
11. Hedayati, M.; Barzegar, A.; Rahimi, A. Fault diagnosis and prognosis of satellites and unmanned aerial vehicles: A review. *Appl. Sci.* **2024**, *14*, 9487. [CrossRef]
12. Xiao, B.; Yin, S. A deep learning based data-driven thruster fault diagnosis approach for satellite attitude control system. *IEEE Trans. Ind. Electron.* **2020**, *68*, 10162–10170. [CrossRef]
13. Mao, Z.; Ma, S.; Liu, W.; Jiang, B. DCDAN-based incipient fault diagnosis for satellite ACS under variable operating conditions. *IEEE Trans. Ind. Inform.* **2023**, *20*, 3115–3123. [CrossRef]
14. Islam, M.S.; Rahimi, A. Fault prognosis of satellite reaction wheels using a two-step LSTM network. In *Proceedings of the 2021 IEEE International Conference on Prognostics and Health Management (ICPHM)*; IEEE: Piscataway, NJ, USA, 2021; pp. 1–7.
15. Gong, J.; Yang, K.; Deng, B.; Bao, L.; Liao, N.; Jiang, W. Research on space-ground collaborative fault diagnosis technology in satellite based on LSTM. In *Proceedings of the 2023 6th International Conference on Intelligent Autonomous Systems (ICoIAS)*; IEEE: Piscataway, NJ, USA, 2023; pp. 118–122.
16. Cipollone, R.; Leonzio, I.; Calabrò, G.; Di Lizia, P. An LSTM-based maneuver detection algorithm from satellites pattern of life. In *Proceedings of the 2023 IEEE 10th International Workshop on Metrology for Aerospace (Metroaerospace)*; IEEE: Piscataway, NJ, USA, 2023; pp. 78–83.
17. Han, K.; Wang, W.; Guo, J. Research on a bearing fault diagnosis method based on a CNN-LSTM-GRU model. *Machines* **2024**, *12*, 927. [CrossRef]
18. Tao, L.; Zhang, T.; Peng, D.; Hao, J.; Jia, Y.; Lu, C.; Ding, Y.; Ma, L. Long-term degradation prediction and assessment with heteroscedasticity telemetry data based on GRU-GARCH and MD hybrid method: An application for satellite. *Aerosp. Sci. Technol.* **2021**, *115*, 106826. [CrossRef]

19. Zhao, H.; Liu, M.; Sun, Y.; Chen, Z.; Duan, G.; Cao, X. Automated design of fault diagnosis CNN network for satellite attitude control systems. *IEEE Trans. Cybern.* **2024**, *54*, 4028–4038. [CrossRef]
20. Lee, K.H.; Lim, S.; Cho, D.H.; Kim, H.D. Development of fault detection and identification algorithm using deep learning for nanosatellite attitude control system. *Int. J. Aeronaut. Space Sci.* **2020**, *21*, 576–585. [CrossRef]
21. Cui, L.; Zhang, Q.; Shi, Y.; Yang, L.; Wang, Y.; Wang, J.; Bai, C. A method for satellite time series anomaly detection based on fast-DTW and improved-KNN. *Chin. J. Aeronaut.* **2023**, *36*, 149–159. [CrossRef]
22. Lacoquelle, C.; Pucel, X.; Massuyès, L.T.; Reymonet, A.; Enaux, B. Warped time series anomaly detection. *IEEE Trans. Signal Process.* **2026**, *74*, 439–452. [CrossRef]
23. Yu, J.; Liu, X.; Ye, L. Convolutional long short-term memory autoencoder-based feature learning for fault detection in industrial processes. *IEEE Trans. Instrum. Meas.* **2020**, *70*, 3505615. [CrossRef]
24. Xi, L.; Miao, D.; Li, M.; Wang, R.; Liu, H.; Huang, X. Adaptive-correlation-aware unsupervised deep learning for anomaly detection in cyber-physical systems. *IEEE Trans. Dependable Secur. Comput.* **2023**, *21*, 2888–2899. [CrossRef]
25. Sun, P.; Jin, G.; Li, W.; Pang, N. Fault Diagnosis of the Thermal Control System based on Temperature Curve Characteristics. In *Proceedings of the 2019 5th International Conference on Big Data and Information Analytics (BigDIA)*; IEEE: Piscataway, NJ, USA, 2019; pp. 92–96.
26. Thoemel, J.; Kanavouras, K.; Sachidanand, M.; Hein, A.; Ortiz del Castillo, M.; Pauly, L.; Rathinam, A.; Aouada, D. Lean demonstration of on-board thermal anomaly detection using machine learning. *Aerospace* **2024**, *11*, 523. [CrossRef]
27. Pang, W.; Zhou, J.; Feng, Z. Thermal Control of Modular Microsatellites: Technologies and Techniques. In *Proceedings of the 2024 10th International Conference on Control, Automation and Robotics (ICCAR)*; IEEE: Piscataway, NJ, USA, 2024; pp. 268–276.
28. Yuan, H.; Cai, Q.; Yin, K. Unsupervised distance metric learning for anomaly detection over multivariate time series. *arXiv* **2024**, arXiv:2403.01895. [CrossRef]
29. Huang, T.; Zhang, Q.; Tang, X.; Zhao, S.; Lu, X. A novel fault diagnosis method based on CNN and LSTM and its application in fault diagnosis for complex systems. *Artif. Intell. Rev.* **2022**, *55*, 1289–1315. [CrossRef]
30. Hedayati, M.; Barzegar, A.; Rahimi, A. Mitigating Data Scarcity for Satellite Reaction Wheel Fault Diagnosis with Wasserstein Generative Adversarial Networks. In *Proceedings of the 2024 IEEE International Conference on Prognostics and Health Management (ICPHM)*; IEEE: Piscataway, NJ, USA, 2024; pp. 367–376.
31. Yan, P.; Abdulkadir, A.; Luley, P.P.; Rosenthal, M.; Schatte, G.A.; Grewe, B.F.; Stadelmann, T. A comprehensive survey of deep transfer learning for anomaly detection in industrial time series: Methods, applications, and directions. *IEEE Access* **2024**, *12*, 3768–3789. [CrossRef]
32. Xie, Y.; Yao, W.; Li, X.; Wang, N.; Zheng, X.; Chen, X. Hybrid digital twin for satellite temperature field perception and attitude control. *Adv. Eng. Inform.* **2024**, *60*, 102405. [CrossRef]
33. Cena, C.; Albertin, U.; Martini, M.; Bucci, S.; Chiaberge, M. Physics-informed real NVP for satellite power system fault detection. In *Proceedings of the 2024 IEEE International Conference on Advanced Intelligent Mechatronics (AIM)*; IEEE: Piscataway, NJ, USA, 2024; pp. 679–684.
34. Ganesan, M.; Lavanya, R.; Nirmala Devi, M. Fault detection in satellite power system using convolutional neural network. *Telecommun. Syst.* **2021**, *76*, 505–511. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.