ENGINEERING Artificial Intelligence
https://doi.org/10.2738/ENGAI.2026.0007

Original Research

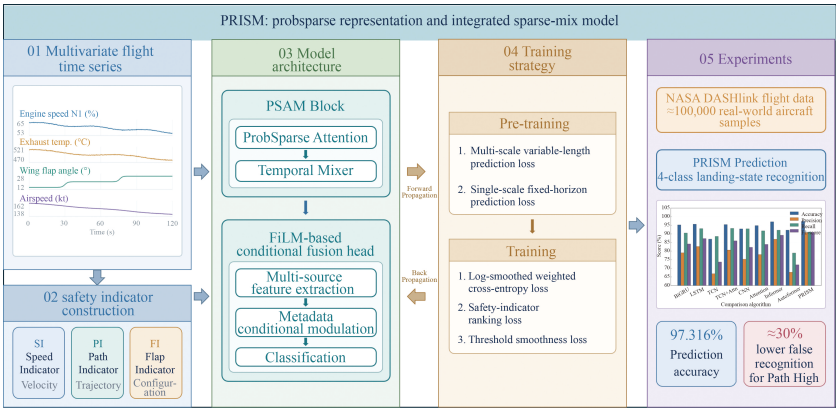
PRISM: Mechanism-Aware Learning with Sparse-Mix Attention for Aircraft Abnormal Landing Recognition

Fanbiao Li (✉)[1], Xiong Peng[1], Yiyun Zhao[1], Jose I. Leon[2], Weihua Gui[1], Haonan Xu[3]

1 School of Automation, Central South University, Changsha 410083, China
2 Department of Electronic Engineering, University of Seville, Seville 41092, Spain
3 COMAC Shanghai Aircraft Design & Research Institute, Shanghai 200232, China

© The Author(s) 2026. This article is published by Higher Education Press.

GRAPHICAL ABSTRACT



HIGHLIGHTS

- An imbalanced classification network is proposed for aircraft landing risk recognition.
- Structured safety indicators are constructed from normal flight statistics for interpretable risk characterization.
- A safety-aware fusion mechanism integrates structured indicators with deep learning for enhanced recognition.
- Extensive experiments on Quick Access Recorder data demonstrate improved overall and minority-class recognition performance.

ABSTRACT During the process of aircraft approach and landing, there are challenges such as inadequate adaptability to complex operational scenarios and class imbalance in abnormal-state recognition. This paper proposes ProbSparse Representation and Integrated Sparse-Mix Model (PRISM), which is an abnormal landing recognition deep learning model that combines structured safety indicators with conditional discrimination mechanisms. Firstly, structured safety indicators are constructed from velocity, trajectory, and configuration dimensions to explicitly characterize critical risk states. Subsequently, a ProbSparse Attention and Multi-Scale Mixing (PSAM) block combining ProbSparse Self-Attention and a Multi-Scale Temporal Mixer is designed to jointly capture global dependencies and local dynamic patterns. In addition, a Feature-wise Linear Modulation (FiLM) based conditional classification head is introduced to dynamically modulate decision boundaries using contextual information such as airport and runway conditions, thereby improving model generalization under complex operational environments. Experiments conducted on nearly 100,000 real-world aircraft samples from the NASA DASHlink dataset demonstrate that PRISM achieves superior performance over baseline models, reaching 97.316% Accuracy. In particular, the false recognition rate for the high-risk Path High category is reduced by approximately 30% compared with Informer. Ablation studies and SHAP (SHapley Additive exPlanations)-based interpretability analysis further verify the effectiveness of the proposed modules and structured indicators. This verifies that PRISM enables accurate, robust, and physically interpretable abnormal landing recognition in complex airport environments, providing a promising solution for intelligent aircraft safety monitoring and early risk warning.

KEYWORDS Aircraft abnormal landing recognition; Multivariate time series classification; Deep learning; Conditional feature fusion

Received 31 Jul 2026; accepted 13 Sep 2026

E-mail: fanbiaoli@csu.edu.cn (Li F).

1 Introduction

In air transportation engineering, aircraft safety has always been one of the most critical research concerns. Compared to relatively stable aircraft phases such as cruising, the landing phase is widely considered a critical stage that affects flight safety due to its operational complexity, frequent state transitions, and high sensitivity to external environmental disturbances. According to the IATA Annual Safety Report [1], runway excursions and tail strikes remain the two most frequently reported accident categories worldwide, both of which predominantly occur during the approach and landing phases. Moreover, although commercial aviation has achieved historically low accident rates, each landing-related safety event still poses substantial operational and economic consequences. These observations highlight the importance of developing intelligent abnormal landing recognition methods for the approach–landing process to enhance operational safety and support flight decision-making [1,2].

During the approach and landing phases, the aircraft is required to perform precise maneuvers under the multiple constraints of critical parameters such as airspeed, path deviation, and navigation deviation. In particular, Glideslope deviation, excessive approach speed, localizer deviation, and delayed flap deployment have been identified as significant predictive precursors of landing risk [3]. Moreover, typical safety events such as runway excursions and hard landings are rarely triggered by a single instantaneous factor; instead, they usually result from the progressive accumulation and coupled interaction of multiple unsafe factors over time. For example, excessive float distance, delayed braking, and delayed reverse-thrust deployment exhibit strong path-dependent relationships during the landing process [4]. Consequently, landing risk is not caused by isolated anomalies at specific moments but rather emerges from the continuous deviation and interaction of multivariate aircraft parameters throughout the temporal evolution process [5]. This characteristic implies that abnormal landing recognition is fundamentally a dynamic temporal modeling problem, requiring continuous analysis of multivariate aircraft states over the entire approach and landing window.

It is worth noting that the approach and landing tasks can be described as a typical time series classification (TSC) problem [6], in which multivariable aircraft parameter sequences are used to determine the landing status. In recent years, the TSC field has achieved substantial progress through various representative approaches, including distance-based methods, feature-transformation methods, and end-to-end deep learning models such as Long Short-Term Memory (LSTM), Transformer, and InceptionTime [7]. However, most existing TSC approaches are designed for generic industrial or sensor-based applications and do not adequately consider the unique characteristics of aviation risk evolution. As a result, they are difficult to directly apply to landing risk recognition tasks. Their limitations are primarily reflected in two aspects: temporal modeling complexity and class imbalance.

On the one hand, from the perspective of temporal feature modeling, landing risk is not determined by instantaneous aircraft states at isolated time points, but rather by the continuous evolution of multivariate states over a period of time. Risk factors such as excessive speed, glide trajectory deviation, and abnormal configuration deployment generally exhibit progressive accumulation patterns with strong temporal dependency. In addition, complex coupling relationships exist among multivariate aircraft parameters, such as the dynamic interaction between pitch angle and vertical speed, resulting in highly correlated risk patterns across both temporal and feature dimensions. Nevertheless, existing generic TSC methods mainly focus on coarse-grained normal-versus-abnormal discrimination and lack the capability to characterize fine-grained abnormal patterns together with their underlying physical semantics [8].

On the other hand, from the data distribution perspective, the vast majority of real-world operational samples correspond to normal landings [9], whereas abnormal events are relatively rare, leading to severe class imbalance. Most generic TSC methods implicitly assume relatively balanced class distributions [10], which causes trained models to be dominated by majority classes and consequently weakens their ability to recognize minority-class anomalies. This issue has been identified as a critical bottleneck in tasks such as hard-landing prediction and runway excursion recognition [11]. Furthermore, most existing approaches directly output classification results without providing structured explanations associated with aircraft operation mechanisms, making it difficult to satisfy the practical requirements for interpretability and reliability in aviation safety applications.

To address the above challenges, this study models the landing process as a multivariate time-series classification problem and develops a neural network architecture tailored for imbalanced learning to improve the recognition capability for typical abnormal landing events. Meanwhile, based on the statistical characteristics of normal landing data and aircraft operation principles, a set of structured safety indicators with explicit physical meanings is constructed to quantify deviations in critical risk factors, including speed, trajectory, and aircraft

configuration. These indicators are incorporated into the proposed model as auxiliary features to enhance both classification performance and model interpretability. Moreover, the proposed structured safety indicators provide stable performance improvements across multiple classification models and strong generalization ability. The main contributions of this paper are summarized as follows.

(1) An imbalanced classification neural network model for aircraft landing risk recognition is proposed, which enhances the detection capability for minority-class abnormal events.

(2) A set of structured safety indicators based on the statistical characteristics of normal aircraft operations is constructed, enabling interpretable modeling of critical risk factors.

(3) A fusion mechanism integrating structured safety indicators with deep learning models is designed, introducing explicit safety-oriented representations into the learning model to improve both classification performance and interpretability.

(4) Extensive experiments conducted on real-world Quick Access Recorder (QAR) aircraft data demonstrate that the proposed method consistently outperforms existing mainstream baseline approaches in terms of both overall classification performance and minority-class recognition capability.

The remainder of this paper is organized as follows. Section 2 reviews recent advances in abnormal landing recognition and TSC methods. Section 3 introduces the proposed PRISM, the construction of structured safety indicators, and the overall model training procedure. Section 4 presents the experimental design and comprehensive result analysis, including comparative experiments, ablation studies, SHAP-based interpretability analysis, and early prediction experiments. Finally, Section 5 concludes the paper and discusses potential directions for future research.

2 Literature Review

Landing risk identification is an inherently interdisciplinary problem spanning multiple research domains. Following the evolution of methodological development, this section reviews four closely related research streams: rule-based landing safety recognition, data-driven landing risk modeling, deep learning methods for TSC, and class-imbalance learning. By analyzing the limitations of these approaches in landing risk identification, the motivation and design rationale of this study are presented.

2.1 Rule-based landing safety recognition

Rule- and threshold-based methods represent the earliest and most widely adopted paradigm for aviation safety recognition. Among them, the Stabilized Approach Criteria proposed by the Flight Safety Foundation is the most representative. It requires aircraft to achieve a stabilized approach no later than 1000 feet above airport elevation under instrument meteorological conditions and 500 feet under visual meteorological conditions, while enforcing thresholds on key parameters such as descent rate, thrust configuration, and trajectory deviation [12]. This criterion has been widely adopted by aviation authorities, operators, and manufacturers, serving as a fundamental reference for Flight Operational Quality Assurance (FOQA) systems and accident investigations [13].

However, rule-based approaches have several inherent limitations. First, their checkpoint-based decisions cannot capture the continuous evolution of aircraft states during approach. Second, predefined thresholds are rigid and lack adaptability to different aircraft types, runway conditions, and meteorological environments. Third, a gap exists between rule definitions and operational practice: although unstable approaches account for approximately 3.5%–4.0% of all approaches, 95%–97% of flight crews still continue landing rather than execute a go-around [14], highlighting the limited capability of rule-based systems for risk quantification and early warning. Moreover, conventional FOQA systems rely on predefined exceedance rules, enabling the detection of known anomalies but failing to identify previously unseen risk patterns [15].

These limitations have driven the development of data-driven approaches that learn risk representations and predictive patterns directly from historical flight data.

2.2 Data-driven landing risk modeling

With the widespread deployment of QAR systems in commercial aviation [16], data-driven safety modeling based on aircraft operational data has become an active research direction. Existing studies can be broadly divided into two developmental stages.

(1) Stage Ⅰ: traditional machine learning methods. Early studies primarily relied on statistical analysis and shallow machine learning, where handcrafted features were extracted for landing event classification or regression. [17] employed logistic regression to analyze the relationship between QAR data and long landing events, identifying key aircraft parameters from 1500 feet to touchdown. Subsequently, support vector machines [18] and radial basis function neural networks [19] were

introduced for hard landing prediction, achieving improved predictive performance.

Machine learning methods have also been applied to runway excursion risk analysis. For example, [3] evaluated six algorithms, including decision trees, gradient boosting, and random forests, using aircraft parameters above 500 feet. The decision tree achieved 98% accuracy in predicting pilot decisions during unstable approaches. In addition, explainable AI techniques such as SHAP have been adopted for causal analysis of runway excursions, identifying key risk factors including asymmetric angle of attack, radio altitude, and reverse thrust deployment timing [4].

However, traditional machine learning methods exhibit two major limitations. First, they rely heavily on handcrafted features [20], making it difficult to capture high-order temporal dependencies automatically. Second, most methods formulate landing risk as a static classification problem using single time points or aggregated statistics, failing to characterize the dynamic evolution of aircraft states during approach.

(2) Stage Ⅱ: deep learning methods. To overcome these limitations, deep learning models have been increasingly adopted for end-to-end modeling of QAR time-series data. LSTM networks were among the earliest methods applied to hard landing prediction because of their ability to model temporal dependencies, outperforming traditional approaches on A320 QAR datasets [21].

Subsequently, Ref. [8] proposed a dual-stage attention network to jointly model temporal dependencies and inter-variable relationships, improving both prediction accuracy and interpretability. Ref. [22] introduced SDTAN, which incorporates scalable temporal attention into hard landing prediction while providing parameter-level interpretability.

For model interpretability, IMTCN [23] combines multi-scale temporal convolutional networks (TCN) with an explainable framework for multi-scale classification of aircraft exceedance events, providing fine-grained temporal attribution. For QAR anomaly detection, MAD-XFP [24] integrates XGBoost with SHAP to automatically detect and interpret multiple anomaly types during the approach phase, demonstrating the potential of interpretable machine learning for FOQA applications.

Despite notable improvements in predictive performance and interpretability, two important challenges remain. Owing to the scarcity of abnormal landing samples, many studies simplify landing risk identification into binary classification (normal vs. abnormal) or regression-based risk scoring [11]. Although these formulations reduce modeling complexity, they overlook the structural heterogeneity of different anomaly types, limiting the ability to distinguish fine-grained risks such as speed deviations, trajectory divergence, and configuration abnormalities.

Moreover, the features learned by existing data-driven models often lack explicit correspondence to aircraft operational mechanisms, limiting both interpretability and reliability in safety-critical aviation applications.

2.3 Deep learning methods for time series classification

Landing risk identification can be naturally formulated as a multivariate time series classification (MTSC) problem, making it closely related to advances in general TSC research.

In recent years, deep learning has become the dominant paradigm for TSC. The survey by [7] systematically evaluates nine representative architectures, including multilayer perceptron (MLP), FCN, and ResNet, on the UCR/UEA benchmark, demonstrating the superiority of deep learning methods for TSC. Building upon this foundation, InceptionTime [25] employs multi-scale convolutional modules to capture temporal patterns at different resolutions and achieves state-of-the-art performance across multiple datasets. ROCKET [26] introduces random convolutional kernels as feature transformations, achieving competitive accuracy with extremely low computational cost and becoming a strong practical baseline. More recently, Transformer-based architectures have been introduced into TSC. ConvTran [27] demonstrates higher accuracy and training efficiency than ROCKET on large-scale datasets. For multivariate scenarios, methods such as FreConvNet [28] further integrate frequency-domain information and cross-variable interactions to enhance multivariate time-series representations [29].

Despite their success on MTSC benchmarks, these methods are not specifically designed for safety-critical tasks such as landing risk identification and have several limitations. First, they lack robustness to class imbalance, as most TSC benchmarks assume relatively balanced distributions and provide no dedicated mechanisms for extreme imbalance, leading to poor minority-class detection in highly imbalanced QAR datasets. Second, they lack physical semantics, as representations learned solely from raw sequences have limited correspondence with aircraft operational mechanisms and are therefore difficult to interpret from an aviation safety perspective. Third, they insufficiently exploit domain knowledge, as general-purpose MTSC models cannot effectively incorporate operational priors, such as glide path deviation, speed margins, and configuration integrity, into feature representation and model design.

Therefore, combining the temporal modeling capability

of MTSC methods with aviation domain knowledge and imbalance-aware learning remains a key challenge for landing risk identification.

2.4 Class-imbalance learning strategies

Class imbalance is one of the primary challenges in landing risk identification. In real-world QAR datasets, abnormal landing events (e.g., hard and long landings) are far less frequent than normal landings. Consequently, standard classification models tend to be biased toward the majority class, resulting in degraded recall for minority classes.

Existing imbalance learning strategies can be broadly divided into three categories. The first category comprises data-level resampling methods. Oversampling techniques generate synthetic minority samples to balance class distributions, among which the synthetic minority oversampling technique (SMOTE) and its variants are the most widely used. SMOTE creates new samples by interpolating existing minority instances [30]. However, in high-dimensional time-series data, such interpolation often lacks physical interpretability and may introduce unrealistic or noisy samples, degrading decision boundaries [31]. Undersampling alleviates imbalance by removing majority-class samples but inevitably discards potentially informative data.

The second category is algorithm-level cost-sensitive learning, which modifies the loss function by assigning larger penalties to minority-class errors. Among these methods, Focal Loss [32] dynamically reweights training samples to emphasize hard minority instances and has demonstrated strong performance in highly imbalanced tasks such as object detection and medical imaging [33].

The third category consists of ensemble learning methods, which improve minority-class robustness by combining multiple base learners trained on different data subsets. Representative examples include Balanced Bagging and EasyEnsemble.

In aviation safety, class imbalance has only been partially addressed. Existing studies mainly reformulate classification as regression (e.g., predicting continuous vertical acceleration instead of hard-landing labels) to alleviate imbalance [11], or restrict modeling to relatively balanced subsets, limiting generalization capability. Systematically integrating imbalance-aware learning into deep classification models for QAR time-series data remains largely unexplored.

In summary, integrating imbalance-aware learning into deep QAR-based time-series classification while incorporating aircraft domain knowledge to construct physically meaningful representations remains an open challenge and forms the primary motivation of this study.

3 Methodologies

3.1 Problem formulation

Let the dataset be defined as $\mathcal{D} = \{(X_i, z_i, y_i)\}_{i=1}^{N}$, where N denotes the number of flight samples. For the i-th sample, $X_i \in \mathbb{R}^{T\times C}$ denotes the multivariate time series recorded during the approach and landing phases, where T and C represent the numbers of time steps and observed variables, respectively. The contextual feature vector $z_i \in \mathcal{Z}$ contains flight-level metadata, including the departure airport, destination airport, departure runway, landing runway, and temporal information.

Let $\mathcal{Y} = \{\mathrm{NL}, \mathrm{SA}, \mathrm{TA}, \mathrm{CA}\}$ denote the landing outcome space, corresponding to normal landing (NL), speed anomaly (SA), trajectory anomaly (TA), and configuration anomaly (CA), respectively. The objective is to learn the following mapping:

$$f : \mathbb{R}^{T\times C} \times \mathbb{R}^{T\times 3} \times \mathcal{Z} \to \mathcal{Y}, \tag{1}$$

where the structured indicator sequence $I_i \in \mathbb{R}^{T\times 3}$ explicitly characterizes aircraft deviations in speed, trajectory, and configuration.

The task is characterized by two major challenges: class imbalance and temporal dependency. In operational aircraft data, normal landing samples outnumber anomalous ones, i.e.,

$$n_{\mathrm{NL}} \gg n_c, \qquad c \in \{\mathrm{SA}, \mathrm{TA}, \mathrm{CA}\}, \tag{2}$$

where n_c denotes the number of samples in class c. Meanwhile, landing anomalies arise from the continuous evolution of multiple flight variables, requiring the model to capture both long-term temporal dependencies and subtle state transitions.

Although the raw multivariate time series captures overall aircraft dynamics, safety-critical states are often implicitly embedded in complex inter-variable relationships, making discriminative learning under severe class imbalance particularly challenging. To address this issue, we introduce the structured indicator sequence I_i as an auxiliary input to explicitly encode deviations in speed, trajectory, and configuration. The construction of these indicators is presented in Section 3.3.

3.2 Model overview

To address the challenges of insufficient temporal dependency modeling, severe class imbalance, and limited physical interpretability in abnormal landing recognition under complex operational scenarios, a unified PRISM model is established, and the overall architecture of the proposed method is illustrated in Fig. 1.

The proposed model takes the raw multivariate time

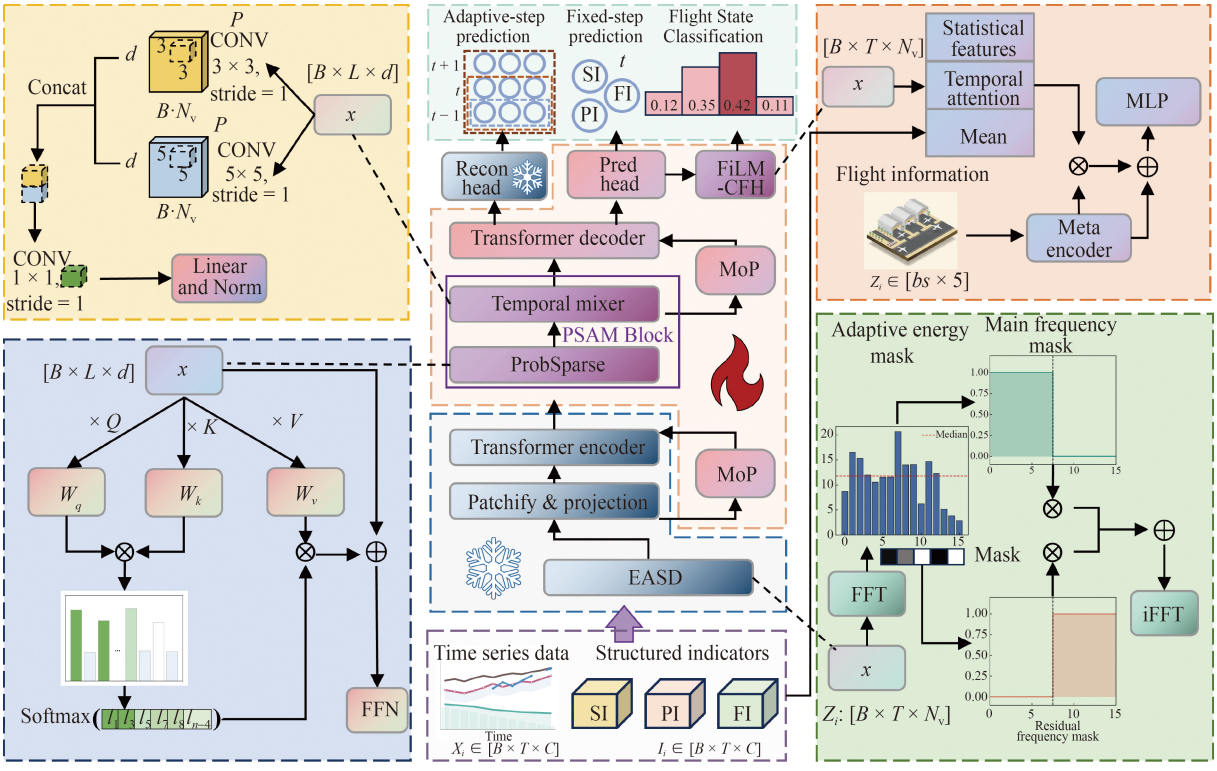


Fig. 1 Model structure.

series X_i and the structured indicator sequence I_i as inputs. They are first transformed into multi-scale representations through the energy-aware spectral decomposition (EASD) and Patchify-Projection modules, followed by a Transformer Encoder enhanced with a mixture-of-projections (MoP) mechanism for global feature extraction.

The encoded features are then processed by the proposed ProbSparse attention and Multi-Scale Mixing (PSAM) block (Section 3.4.1), which combines ProbSparse Self-Attention refinement with a Multi-Scale Temporal Mixer to jointly model sparse long-range dependencies and fine-grained local dynamics, emphasizing temporally critical flight states while suppressing redundant information.

The refined features are decoded by a MoP-enhanced Transformer Decoder with two auxiliary self-supervised objectives: a Reconstruction Head (Recon Head) for multi-scale structured-indicator forecasting and a Prediction Head (Pred Head) for fixed-horizon prediction. Together, these objectives regularize temporal representation learning and improve the modeling of flight-state evolution.

The prediction features generated by the Pred Head are then fused with the structured flight-mechanism indicators and aircraft operational context features z_i through the FiLM-based Conditioned Fusion Head (FiLM-CFH) (Section 3.4.2), where conditional modulation adaptively adjusts the decision boundaries across different operational environments before producing the final landing classification.

The model is optimized in two stages: the backbone is first pre-trained for general temporal representation learning, after which its parameters are frozen and only task-specific modules (e.g., FiLM-CFH) are fine-tuned. Details are provided in Section 3.5.

3.3 Indicator construction

The structured indicator sequence is constructed from three complementary dimensions—speed, trajectory, and configuration to explicitly characterize safety-critical aircraft states. For each sample, the indicator sequence is defined as

$$I_i = [SI_i, PI_i, FI_i] \in \mathbb{R}^{T\times 3}, \tag{3}$$

where SI, PI, and FI denote the speed, path, and configuration indicators, respectively.

3.3.1 Speed Indicator (SI)

Speed-related risks typically manifest as sustained speed deviations and thrust mismatches during approach. All variables are normalized using statistics computed from

normal landing samples:

$$l(x) = \frac{x - \mu_x}{\sigma_x + \epsilon}, \tag{4}$$

where μ_x and σ_x denote the mean and standard deviation of variable x, respectively, and ϵ is a small constant to avoid division by zero.

The SI is then defined as:

$$SI_t = \alpha_1 l(v)_t + \alpha_2 \overline{l(v)}_t + \alpha_3 l(N1)_t, \tag{5}$$

where the causal sliding-window average is defined as:

$$\overline{l(v)}_t = \begin{cases} \dfrac{1}{t+1} \displaystyle\sum_{s=0}^{t} l(v)_s, & t < W, \\ \dfrac{1}{W} \displaystyle\sum_{s=t-W+1}^{t} l(v)_s, & t \geqslant W, \end{cases} \tag{6}$$

where $l(v)_t$ and $l(N1)_t$ denote the standardized airspeed and N1 at time t, respectively, measuring their deviations from the nominal landing condition, while $\overline{l(v)}_t$ captures its recent persistence over the preceding W time steps. The causal formulation uses only current and past observations. The window length W and weighting coefficients $\{\alpha_1, \alpha_2, \alpha_3\}$ are treated as hyperparameters.

3.3.2 Path Indicator (PI)

Trajectory-related risks primarily arise from Glideslope and localizer deviations, as well as instabilities in attitude and altitude control. The PI is constructed as:

$$PI_t = \beta_1 l(\Delta_{gs})_t + \beta_2 l(h^+)_t + \beta_3 l(|\Delta_{\text{loc}}|)_t + \beta_4 \tilde{l}(a_{\text{aoa}})_t, \tag{7}$$

where Δ_{gs} and $|\Delta_{\text{loc}}|$ denote the Glideslope deviation and the absolute localizer deviation, respectively; h^+ denotes the positive altitude deviation relative to the normal altitude profile; and $\tilde{l}(a_{\text{aoa}}) = \text{clip}(l(a_{\text{aoa}}), -2, 2)$ clips the normalized angle of attack to suppress outlier sensitivity. The weighting coefficients $\{\beta_1, \beta_2, \beta_3, \beta_4\}$ control the relative contributions of different trajectory-related factors.

3.3.3 Flap Indicator (FI)

Configuration-related risks are primarily manifested as delayed flap deployment and mismatches between flap settings and aircraft states. The expected flap position is estimated from normal landing samples through an empirical altitude-to-flap mapping:

$$f(h) = \text{median}\{flap_t \mid h_t \in \mathcal{H}_k,\ y = 0\}, \quad h \in \mathcal{H}_k, \tag{8}$$

where the altitude range is partitioned into 20 equal-frequency bins and linear interpolation is applied between adjacent bin centers.

The configuration lag is then defined as:

$$\Delta_{flap} = \max(0,\ f(h) - flap). \tag{9}$$

To reflect the elevated risk of configuration lag at low altitudes, an altitude-dependent weighting function is introduced:

$$w(h) = \text{clip}\left(1.5 - \tilde{h},\ 0.5,\ 1.5\right), \tag{10}$$

where $\tilde{h} \in [0, 1]$ is the per-sample min-max normalized altitude. The truncation operators are defined as:

$$l(v)^+ = \max(0,\ l(v)), \qquad l(flap)^- = \max(0,\ -l(flap)). \tag{11}$$

Then, the final FI is:

$$FI_t = \gamma_1 \cdot w(h_t) \cdot l(\Delta_{flap})_t + \gamma_2 \cdot l(v)_t^+ \cdot l(flap)_t^-, \tag{12}$$

where the first term captures altitude-weighted configuration lag and the second term encodes the compound risk of simultaneous excess speed and insufficient flap deployment. The coefficients γ_1 and γ_2 balance the contributions of configuration lag and the compound speed–configuration risk, respectively.

The above three indicators provide a structured, physics-informed characterization of anomalous aircraft state deviations, rendering risk patterns that are otherwise latent in high-dimensional observations explicitly accessible, thereby enhancing both feature interpretability and the discriminability of the structured input to subsequent temporal modeling.

3.4 Model architecture

Based on the above discussion, a multi-source collaborative model under the unified model will be established that jointly processes the raw multivariate time series $\boldsymbol{X} \in \mathbb{R}^{T \times C}$, the structured safety indicator sequence $\boldsymbol{I} \in \mathbb{R}^{T \times 3}$, and the static contextual feature vector $\boldsymbol{z} \in \mathbb{R}^5$.

Since $\boldsymbol{I}$ and $\boldsymbol{X}$ are temporally aligned, they are concatenated along the variable dimension:

$$\widetilde{\boldsymbol{X}} = [\boldsymbol{X} \| \boldsymbol{I}] \in \mathbb{R}^{T \times N_v}, \quad N_v = C + 3, \tag{13}$$

where $N_v = C + 3$. The concatenated sequence serves as the primary temporal input, while $\boldsymbol{z}$ is introduced at the decision stage for conditional modulation.

3.4.1 ProbSparse attention and PSAM block

(1) ProbSparse self-attention refinement module. After encoder processing, $\widetilde{\boldsymbol{X}}$ is partitioned into P non-overlapping patches of length L_p and projected into a token sequence of total length $L = N_v \times P$. Standard self-attention incurs $O(L^2)$ complexity and performs

redundant computations for many queries with nearly uniform attention distributions. To address this issue, a ProbSparse refinement module selectively computes full attention only for the most informative queries.

Given encoder output $\boldsymbol{F} \in \mathbb{R}^{L\times d}$, queries, keys, and values are obtained via linear projections:

$$\boldsymbol{Q} = \boldsymbol{F}\boldsymbol{W}_Q, \quad \boldsymbol{K} = \boldsymbol{F}\boldsymbol{W}_K, \quad \boldsymbol{V} = \boldsymbol{F}\boldsymbol{W}_V. \tag{14}$$

A sparsity measure based on random key sampling is defined to quantify Query informativeness:

$$\mathcal{M}(q_i, \boldsymbol{K}) = \max_j\left(\frac{q_i k_j^\top}{\sqrt{d_k}}\right) - \frac{1}{\hat{L}}\sum_{j=1}^{\hat{L}}\frac{q_i k_j^\top}{\sqrt{d_k}}, \tag{15}$$

where $\hat{L} = \min(c\cdot\ln L, L)$ is the number of randomly sampled keys and c is the sampling factor.d denotes the size of the hidden layer. A large $\mathcal{M}(q_i, \boldsymbol{K})$ indicates that q_i is sharply focused on a few key positions and thus carries rich discriminative information.

The top-$u = \lceil c\cdot\ln L\rceil$ queries by score form the active set $\mathcal{S}$ and undergo full scaled dot-product attention; the remaining queries are filled with the global mean of the value matrix:

$$\boldsymbol{O}_i = \begin{cases}\mathrm{Softmax}\left(\dfrac{q_i\boldsymbol{K}^\top}{\sqrt{d_k}}\right)\boldsymbol{V}, & i\in\mathcal{S},\\ \bar{\boldsymbol{v}}, & i\notin\mathcal{S},\end{cases} \tag{16}$$

where $\bar{\boldsymbol{v}} = \dfrac{1}{L}\sum_{j=1}^{L} v_j$. This design reduces attention complexity from $O(L^2)$ to $O(L\ln L)$. The output is further enhanced via a residual connection and a feed-forward network, yielding a refined global feature representation $\boldsymbol{F}' \in \mathbb{R}^{L\times d}$.

(2) Multi-scale temporal mixer. While ProbSparse captures global dependencies, local landing dynamics occur at multiple temporal scales. Therefore, a Multi-Scale Temporal Mixer with parallel depthwise convolutions and cross-variable mixing is introduced.

Given the ProbSparse output $\boldsymbol{F}' \in \mathbb{R}^{(N_v\cdot P)\times d}$, it is first reshaped into $\boldsymbol{H} \in \mathbb{R}^{N_v\times d\times P}$ to apply temporal convolutions independently over each variable's patch sequence. Two parallel depthwise separable convolutions then extract multi-scale local features:

$$\boldsymbol{H}_k = \mathrm{DWConv}_k(\boldsymbol{H}), \quad k\in\{3,5\}, \tag{17}$$

where $k=3$ captures short-range abrupt changes and $k=5$ models medium-range gradual trends. Depthwise separable convolutions are adopted for computational efficiency. The above two branches are concatenated along the channel dimension and fused via a 1×1 convolution back to the original dimension:

$$\boldsymbol{H}_{\mathrm{ms}} = \mathrm{Conv}_{1\times1}([\boldsymbol{H}_3 \,\|\, \boldsymbol{H}_5]) \in \mathbb{R}^{N_v\times d\times P}. \tag{18}$$

A residual connection and layer normalization then refine the temporal features:

$$\boldsymbol{F}'' = \mathrm{LayerNorm}(\boldsymbol{F}' + \mathrm{Drop}(\mathrm{reshape}(\boldsymbol{H}_{\mathrm{ms}}))) \in \mathbb{R}^{(N_v\cdot P)\times d}. \tag{19}$$

Cross-variable linear mixing is subsequently applied to capture inter-sensor co-variation. After reshaping $\boldsymbol{F}''$ to $\mathbb{R}^{P\times d\times N_v}$, a learnable linear transformation is applied:

$$\boldsymbol{F}''' = \mathrm{LayerNorm}(\boldsymbol{F}'' + \mathrm{Drop}(\boldsymbol{F}''\boldsymbol{W}_{\mathrm{var}})), \quad \boldsymbol{W}_{\mathrm{var}} \in \mathbb{R}^{N_v\times N_v}, \tag{20}$$

where $\boldsymbol{W}_{var}$ is initialized as the identity matrix to preserve pre-trained representations at the beginning of fine-tuning while allowing cross-variable relationships to be learned. Together, the two modules provide efficient global-local temporal modeling with an overall complexity of $O(L\ln L + N_v P d)$.

3.4.2 FiLM-CFH

After the PSAM Block, the refined features are decoded by the MoP-enhanced Transformer Decoder. The Prediction Head then generates a future structured indicator sequence:

$$\widehat{\boldsymbol{I}} = \mathrm{PredHead}(\boldsymbol{F}''') \in \mathbb{R}^{T'\times3}, \tag{21}$$

where T' denotes the prediction horizon. Instead of being used only as an auxiliary pre-training objective, $\widehat{\boldsymbol{I}}$ is further retained as the primary temporal input to FiLM-CFH, providing forward-looking risk information for downstream classification.

Temporal features alone cannot capture operational variability across airports, runways, and procedures. Therefore, aircraft metadata are introduced for context-aware decision-making.

To address this problem, the FiLM-CFH is proposed, a context-adaptive discriminative module that takes $\widehat{\boldsymbol{I}}$ as the primary input and conditions it on both the observed structured indicator sequence $\boldsymbol{I}$ and the static contextual vector $\boldsymbol{z}$ for dynamic modulation and multi-source fusion.

(1) Multi-source feature extraction. Three complementary feature pathways are constructed to characterize aircraft states at multiple granularities.

Pathway 1: statistical Features. Global statistical features are extracted from $\widehat{\boldsymbol{I}}$ using first- and second-order moments:

$$\boldsymbol{h}_{\mathrm{stat}} = [\mu(\widehat{\boldsymbol{I}}),\ \sigma(\widehat{\boldsymbol{I}}),\ \max(\widehat{\boldsymbol{I}}),\ \min(\widehat{\boldsymbol{I}})] \in \mathbb{R}^{12}, \tag{22}$$

where all statistics are computed along the temporal axis T'.

Pathway 2: temporal attention features. A temporal

attention module aggregates future indicator features:

$$\alpha_t = \frac{\exp\left(f(\hat{i}_t)\right)}{\sum_{t'=1}^{T'} \exp\left(f(\hat{i}_{t'})\right)}, \qquad \boldsymbol{h}_{\text{attn}} = \sum_{t=1}^{T'} \alpha_t \hat{i}_t \in \mathbb{R}^3, \tag{23}$$

where $f(\cdot)$ is a two-layer MLP and $\hat{i}_t \in \mathbb{R}^3$ is the predicted indicator vector at future step t.

Pathway 3: observed indicator features. The observed indicator sequence is averaged to summarize historical deviations:

$$\boldsymbol{h}_{\text{ind}} = \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{I}_t \in \mathbb{R}^3. \tag{24}$$

Since $\boldsymbol{I}$ is deterministically constructed from domain knowledge, gradients are detached to preserve it as a fixed prior. Combined with the predicted indicator sequence $\widehat{\boldsymbol{I}}$, the classifier exploits complementary historical and future risk information.

The three pathways are concatenated into a base representation:

$$\boldsymbol{h}_x = [\boldsymbol{h}_{\text{stat}}, \boldsymbol{h}_{\text{attn}}, \boldsymbol{h}_{\text{ind}}] \in \mathbb{R}^{d_{\text{base}}}, \quad d_{\text{base}} = 18. \tag{25}$$

(2) Metadata conditional modulation. The base feature $\boldsymbol{h}_x$ is then conditioned on the aircraft metadata $\boldsymbol{z} \in \mathbb{R}^5$, which encodes departure airport, destination airport, departure runway, landing runway, and temporal information, all static within a given aircraft. A two-layer MLP encoder maps $\boldsymbol{z}$ to a continuous conditioning vector:

$$\boldsymbol{h}_z = \phi(\boldsymbol{z}) \in \mathbb{R}^d, \tag{26}$$

where $\phi(\cdot)$ consists of two fully connected layers with LayerNorm and Gaussian error linear unit (GELU) activation.

FiLM is then applied, generating element-wise scale and shift parameters from the conditioning vector:

$$\gamma = 2 \cdot \sigma(\boldsymbol{W}_\gamma \boldsymbol{h}_z), \qquad \boldsymbol{\beta} = \boldsymbol{W}_\beta \boldsymbol{h}_z, \tag{27}$$

where $\sigma(\cdot)$ denotes the Sigmoid function, constraining the scale factor to $(0, 2)$ to prevent gradient explosion while preserving bidirectional modulation capacity. The modulated feature is:

$$\tilde{\boldsymbol{h}}_x = \gamma \odot \boldsymbol{h}_x + \boldsymbol{\beta}, \tag{28}$$

(3) Classification. The modulated feature and the metadata encoding are concatenated and passed through a three-layer classifier to produce the final risk category prediction:

$$\hat{y} = \text{MLP}\left([\tilde{\boldsymbol{h}}_x, \boldsymbol{h}_z]\right) \in \mathbb{R}^4. \tag{29}$$

The secondary injection of $\boldsymbol{h}_z$ at the decision layer provides a direct operational context signal that is complementary to the upstream FiLM modulation, forming a dual conditioning pathway. The classifier consists of alternating linear layers, LayerNorm, and GELU activations, with Dropout applied after each layer for regularization.

FiLM-CFH integrates future risk prediction, historical safety indicators, and operational metadata into a unified context-aware classifier for landing risk identification.

3.5 Training Strategy

3.5.1 Pre-training

The pre-training stage aims to learn generalized temporal representations without manual annotations. Supervisory signals are generated online from raw aircraft parameters using the domain-knowledge formulas in Section 3.3, preserving the self-supervised nature of pre-training. Accordingly, the encoder and decoder are jointly optimized using two complementary objectives: multi-scale reconstruction and fixed-horizon prediction.

(1) Multi-scale variable-length prediction loss. The Recon Head predicts structured indicator sequences at multiple temporal horizons. Given the input $\widetilde{\mathbf{X}}$, it outputs $\{\widehat{\boldsymbol{I}}^{(i)}\}_{i=1}^{K}$, where $\widehat{\boldsymbol{I}}^{(i)} \in \mathbb{R}^{h_i \times 3}$. The reconstruction loss is:

$$\mathcal{L}_{\text{recon}} = \sum_{i=1}^{K} w_i \cdot \frac{1}{h_i \cdot 3} \left\| \widehat{\boldsymbol{I}}^{(i)} - \boldsymbol{I}_{-h_i:} \right\|_F^2. \tag{30}$$

A short-horizon-priority weighting strategy ($w_1 > w_2 > \ldots > w_K$) is adopted to stabilize optimization in the early training stage while gradually encouraging long-term temporal modeling. The multi-scale design further improves representation generalization across different prediction horizons.

(2) Single-scale fixed-horizon prediction loss. To model future safety evolution, the Pred Head predicts the future structured indicator sequence $\widehat{\boldsymbol{I}} \in \mathbb{R}^{T' \times 3}$ from the observed sequence. The prediction loss is

$$\mathcal{L}_{\text{pred}} = \frac{1}{T' \cdot 3} \left\| \widehat{\boldsymbol{I}} - \boldsymbol{I}_{\text{future}} \right\|_F^2. \tag{31}$$

During inference, $\widehat{\boldsymbol{I}}$ is directly fed into FiLM-CFH, enabling the predictive capability learned during pre-training to be transferred to downstream classification.

(3) Joint pre-training objective. Combining the two subtasks above, the overall loss function during pre-training is defined as:

$$\mathcal{L}_{\text{pretrain}} = \lambda_r \mathcal{L}_{\text{recon}} + \lambda_p \mathcal{L}_{\text{pred}}, \tag{32}$$

where λ_r and λ_p balance reconstruction and prediction losses, respectively.

3.5.2 Fine-tuning

During fine-tuning, the encoder backbone and Recon Head are frozen to preserve the generalized temporal representations learned during pre-training, while the remaining task-specific modules are optimized for landing risk classification.

Given the severe class imbalance, the fine-tuning objective jointly optimizes classification accuracy, risk discrimination, and physical consistency:

$$\mathcal{L}_{\text{fine}} = \mathcal{L}_{\text{ce}} + \lambda_{\text{rank}} \mathcal{L}_{\text{rank}} + \lambda_{\text{smooth}} \mathcal{L}_{\text{smooth}}, \tag{33}$$

where λ_{rank} and λ_{smooth} control the contributions of the ranking loss and the smoothness regularization term, respectively.

(1) Log-smoothed weighted cross-entropy loss. Logarithmically smoothed class weights are adopted to alleviate class imbalance:

$$w_c = \text{clamp}\left(\log \frac{N}{n_c},\ w_{\min},\ w_{\max}\right), \tag{34}$$

where $N = \sum_c n_c$ denotes the total number of samples, n_c represents the number of samples belonging to class c, and $w_{\min}$ and $w_{\max}$ are lower and upper clipping bounds, respectively. The weighted cross-entropy loss is formulated as:

$$\mathcal{L}_{\text{ce}} = -\sum_{c=0}^{3} w_c \mathbf{1}[y = c] \log p_c, \tag{35}$$

where p_c denotes the predicted probability for class c, and y is the ground-truth label.

(2) Safety-indicator ranking loss. To enforce consistency with aircraft safety mechanisms, a margin-based ranking loss is imposed on the three safety indicators:

$$\mathcal{L}_{\text{rank}} = \frac{1}{3} \sum_{k\in\{\text{SI,PI,FI}\}} \Big[\mathbb{E}_{\text{pos}}[\text{ReLU}(\hat{\tau}_k - s_k + m)] + \mathbb{E}_{\text{neg}}[\text{ReLU}(s_k - \hat{\tau}_k + m)]\Big], \tag{36}$$

where s_k denotes the corresponding ground-truth safety indicator value, m is the margin parameter, and and "pos" represents high-risk positive samples corresponding to abnormal classes, and "neg" denotes negative samples from the normal class (Class 0).

(3) Threshold smoothness loss. To encourage physically plausible threshold evolution, a temporal smoothness regularization is imposed on adjacent threshold vectors:

$$\mathcal{L}_{\text{smooth}} = \frac{1}{(T'-1)\cdot 3} \sum_{t=1}^{T'-1} \|\hat{\boldsymbol{\tau}}_t - \hat{\boldsymbol{\tau}}_{t-1}\|_2^2, \tag{37}$$

where $\hat{\boldsymbol{\tau}}_t$ is obtained from $\widehat{\boldsymbol{I}}$ through a lightweight linear layer within FiLM-CFH. This regularization suppresses unrealistic high-frequency fluctuations while preserving classification performance.

Together, the three objectives jointly improve statistical discrimination, safety-semantic consistency, and temporal plausibility during fine-tuning.

4 Experiments

4.1 Experimental setup

4.1.1 Dataset

This study conducts experiments on the Curated 4-Class Anomaly Detection Dataset [34], which was released on the NASA DASHlink platform. The dataset is derived from real-world onboard aircraft data collected over three years of commercial operations of a regional aircraft and is currently one of the few publicly available landing safety datasets that simultaneously provide multi-class annotations and large-scale sample coverage. It covers diverse operational scenarios involving 205 distinct airports and 617 different runways, providing substantial environmental and operational variability for evaluating the generalization capability of landing recognition models.

Each sample corresponds to a complete approach procedure and consists of a multivariate time series covering the final 160 seconds prior to landing. The dataset contains 20 flight parameters collected from onboard sensors, jointly characterizing aircraft operational states during the approach and landing phases. These variables cover multiple operational dimensions, including control surface states, aircraft attitudes, motion states, navigation deviations, propulsion-related variables, and environmental conditions. Detailed parameter definitions are summarized in Table 1.

In addition, each sample is associated with aircraft-level operational metadata, including departure and destination airports, takeoff and landing timestamps, departure runway, and landing runway information. These metadata describe the operational environment and mission context at a macroscopic level and can be further utilized to analyze the potential influence of different operational conditions, such as runway configurations and route structures, on approach stability and deviation risk.

The complete dataset contains 99,837 samples and exhibits severe class imbalance. The sample distribution of each category is summarized in Table 2.

Table 1 Flight parameters used in the dataset.

Feature ID	Parameter	Category
F0	Aileron LH	Control Surface State
F1	Aileron RH	Control Surface State
F2	Angle of Attack (AoA)	Aircraft Attitude
F3	Altitude	Motion State
F4	Airspeed	Motion State
F5	Selected Course	Navigation Variable
F6	Drift Angle	Aircraft Attitude
F7	Elevator	Control Surface State
F8	Flap Position	Control Surface State
F9	Glideslope Deviation	Navigation Deviation
F10	Selected Heading	Navigation Variable
F11	Localizer Deviation	Navigation Deviation
F12	N1 Average	Propulsion Variable
F13	Total Pressure	Propulsion Variable
F14	Pitch Angle	Aircraft Attitude
F15	Roll Angle	Aircraft Attitude
F16	Rudder	Control Surface State
F17	True Heading	Navigation Variable
F18	Vertical Acceleration	Motion State
F19	Wind Speed	Environmental Condition

Table 2 Distribution of dataset categories.

	Nominal	Speed High	Path High	Flaps Late
Training set (63,895)	57384	4488	1413	610
Validation set (15,974)	14346	1122	353	153
Test set (19,968)	17933	1403	441	191
Total(99,837)	89663	7013	2207	954

The dataset is divided into training, validation, and test sets with a ratio of 64%, 16% and 20%, respectively. Stratified sampling is adopted during the splitting process to ensure that the class distribution in each subset remains consistent with that of the entire dataset. All experiments are conducted on a server equipped with an NVIDIA Tesla T4 GPU and an Intel Xeon 5215 CPU.

4.1.2 Indicator hyperparameter settings

The proposed model includes several manually specified hyperparameters, including indicator construction parameters, loss weights, class-weight clipping bounds, and the ranking margin. These values are determined empirically based on preliminary experiments and domain knowledge to balance robustness, discriminative capability, and optimization stability. The adopted settings are summarized in Table 3, and all subsequent experiments use these settings unless otherwise specified.

4.2 Discriminative analysis of the PI

To evaluate the effectiveness of the proposed indicators in distinguishing landing states, SI, PI, and FI are analyzed from the perspectives of statistical significance, discriminative capability, and distribution characteristics.

Table 3 Hyperparameter settings used in the proposed model.

Parameter	Value	Description
W	20	Causal sliding-window length
α_1	0.30	Instantaneous airspeed weight
α_2	0.50	Window-averaged airspeed weight
α_3	0.20	Engine fan speed ($N1$) weight
β_1	0.45	Glideslope deviation weight
β_2	0.25	Positive altitude deviation weight
β_3	0.20	Localizer deviation weight
β_4	0.10	Angle-of-attack weight
γ_1	0.75	Configuration lag weight
γ_2	0.25	Speed–flap interaction weight
λ_r	0.40	Reconstruction loss weight
λ_p	0.60	Prediction loss weight
λ_{rank}	0.30	Ranking loss weight
λ_{smooth}	0.05	Smoothness regularization weight
w_{min}	0.30	Minimum class weight
w_{max}	2.50	Maximum class weight
m	0.30	Ranking margin
d	0.1	Dropout
d_{model}	128	Dimension of model
p	10	Early stopping patience
b	32	Batch size of training input data
l	0.00001	Optimizer learning rate

The corresponding results are presented in Tables 4–5 and Figs. 1–3.

Table 4 Statistical and discriminative performance of the indicators.

Indicator	H Statistic	P-value	Target class	AUC
SI	9644.7	$< 1 \times 10^{-6}$	Speed High	0.797
PI	3325.0	$< 1 \times 10^{-6}$	Path High	0.852
FI	7326.7	$< 1 \times 10^{-6}$	Flaps Late	0.953

Table 5 Average values of SI, PI, and FI across different categories.

Category	SI Mean	PI Mean	FI Mean
Nominal	0.076	0.191	0.334
Speed High	0.585	0.213	1.059
Path High	0.470	1.196	1.441
Flaps Late	1.038	0.376	2.426

As shown in Table 4, all three indicators exhibit significant inter-class differences. The Kruskal–Wallis statistics are consistently high (e.g., 9644.7 for SI and 7326.7 for FI), with all P-values below 1×10^{-6}, indicating that the indicator distributions differ across landing categories and characterize their behavioral differences.

In terms of discriminative capability, binary ROC curves are constructed using "target anomaly class versus normal class" settings to quantitatively evaluate the classification ability of each indicator. As illustrated in Fig. 2, the AUC values of SI, PI, and FI reach 0.797, 0.852, and 0.953, respectively, all exceeding the commonly accepted threshold of 0.7 for effective discrimination. Among them, FI achieves the best performance, indicating that configuration-related anomalies exhibit a clearer separation boundary in the

https://doi.org/10.2738/ENGAI.2026.0007

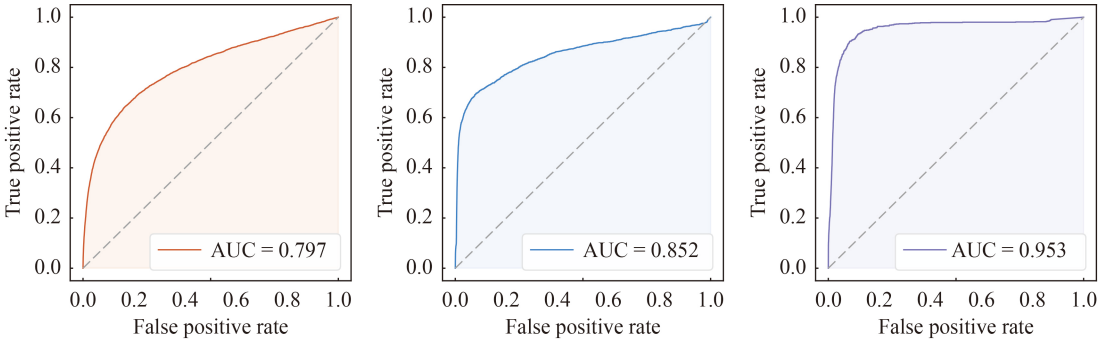


Fig. 2 ROC curve: Abnormal vs. Normal classification.

indicator space. PI follows closely and demonstrates strong capability in distinguishing trajectory-related anomalies. In contrast, SI shows relatively weaker discriminative performance, although it still maintains acceptable classification capability, reflecting the relatively larger overlap between speed-related anomalies and normal states.

Further observations from the box plots in Fig. 3 and the statistical mean values in Table 5 reveal clear distribution structures among different classes in the indicator space. Overall, each indicator exhibits a pronounced increase within its corresponding anomaly category. For example, the mean value of SI increases from 0.076 under normal conditions to 0.585 in the Speed High category; PI reaches 1.196 in the Path High category; and FI further rises to 2.426 in the Flaps Late category. Meanwhile, abnormal classes generally exhibit wider distribution ranges, reflecting the higher instability of abnormal aircraft states. Although partial overlap still exists between certain adjacent categories, the overall distributions remain well separable.

As shown in Fig. 4, 100 representative samples are randomly selected from each landing category to visualize the temporal evolution of the proposed indicators during the approach phase, where each curve corresponds to the indicator trajectory of an individual flight. For normal landings, SI, PI, and FI remain close to zero with only minor fluctuations, indicating that the aircraft generally operates within the nominal safety envelope. In contrast, abnormal landings exhibit distinct category-specific temporal patterns. Specifically, speed-related anomalies are characterized by persistently elevated SI values, trajectory-related anomalies show larger fluctuations and deviations in PI, and configuration-related anomalies present a pronounced increase in FI during the middle-to-late approach phase, reflecting the cumulative effect of delayed flap deployment. Despite the variability among individual flights, consistent evolution patterns can be observed within each anomaly category, demonstrating that the proposed indicators characterize the temporal progression of different landing risks.

To further examine whether the discriminative capability of the proposed indicators depends strongly on a specific weighting configuration, a hyperparameter sensitivity analysis is conducted. For each indicator, the weighting coefficients are locally perturbed around the baseline configuration by adjusting individual coefficients by 0.05 while maintaining their normalized sum. The baseline configurations are $(0.30, 0.50, 0.20)$ for SI, $(0.45, 0.25, 0.20, 0.10)$ for PI, and $(0.75, 0.25)$ for FI. For each indicator, only its corresponding anomaly category is compared with the Nominal category to isolate the effect of the weighting configuration. The

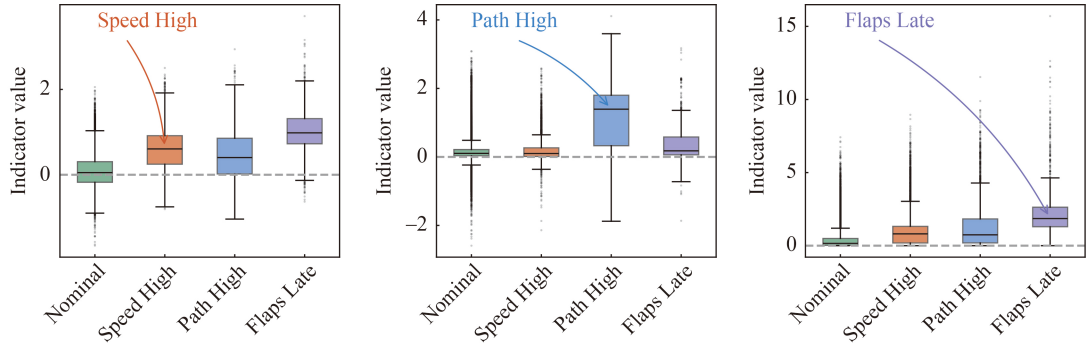


Fig. 3 SI, PI, and FI box plot.

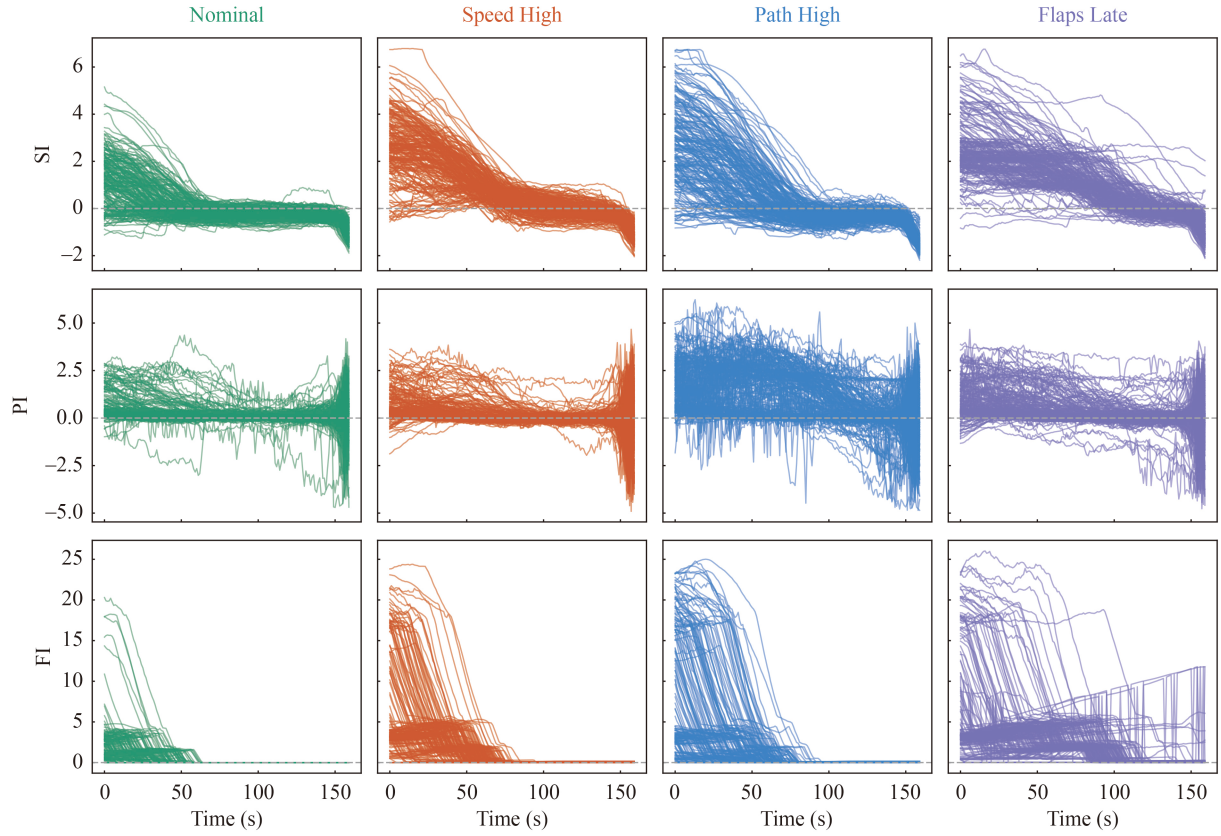


Fig. 4 The temporal evolution of each indicator in different categories.

resulting AUC values are summarized in Table 6.

As shown in Table 6, SI exhibits very low sensitivity to moderate changes in its weighting coefficients. Its AUC varies only from 0.7971 to 0.7978, corresponding to a maximum absolute deviation of 0.0004 from the baseline value. FI also maintains consistently high discriminative capability, with AUC values ranging from 0.9472 to 0.9554 and an absolute deviation of no more than 0.0054 from the baseline. In contrast, PI exhibits a relatively stronger sensitivity to the weighting configuration, with AUC values ranging from 0.8417 to 0.8623. This variation suggests that the relative contributions of path-related variables have a greater influence on the discriminative capability of PI. Nevertheless, the baseline configuration achieves an AUC of 0.8519 and remains within the performance range of the investigated configurations.

The proposed indicator system (SI, PI, and FI) captures category-specific landing risk characteristics across both indicator and temporal domains. Demonstrating strong

Table 6 Hyperparameter sensitivity of SI, PI, and FI in terms of AUC.

Indicator	Configuration	Weights	AUC
SI	Baseline	(0.30, 0.50, 0.20)	0.7974
	Instantaneous-speed emphasis	(0.35, 0.45, 0.20)	0.7977
	Persistent-speed emphasis	(0.25, 0.55, 0.20)	0.7971
	N1 emphasis	(0.30, 0.45, 0.25)	0.7978
PI	Baseline	(0.45, 0.25, 0.20, 0.10)	0.8519
	Glideslope emphasis	(0.50, 0.20, 0.20, 0.10)	0.8417
	Altitude emphasis	(0.40, 0.30, 0.20, 0.10)	0.8623
	Localizer emphasis	(0.45, 0.20, 0.25, 0.10)	0.8526
FI	Baseline	(0.75, 0.25)	0.9526
	Compound-risk emphasis	(0.70, 0.30)	0.9498
	Flap-lag emphasis	(0.80, 0.20)	0.9554
	Balanced weighting	(0.65, 0.35)	0.9472

For each indicator, the weighting coefficients are normalized to sum to one. The baseline configuration corresponds to the parameter setting adopted in the proposed indicator construction.

class discrimination, clear physical interpretability, and high robustness to weight perturbations, it provides a solid foundation for downstream model learning.

4.3 Generalization of the Proposed Indicators across models

To evaluate the generalization capability of the proposed structured safety indicators and contextual features, they are incorporated into several representative time-series classification models, and the performance before and after feature augmentation is compared. The results are summarized in Table 7.

As shown in Table 7, incorporating the proposed structured indicators and contextual features generally improves the performance of all baseline models. Compared with using only raw time-series inputs and a standard multilayer perceptron classification head, most models achieve higher Accuracy, Precision, and F1-score, demonstrating that the introduced auxiliary information enhances the discrimination of abnormal landing patterns.

The performance gains are particularly evident for models with relatively limited temporal modeling capability. For example, the Accuracy of TCN increases from 79.147% to 86.969%, accompanied by improvements in Precision and F1-score, while Autoformer also shows consistent gains across all evaluation metrics. These results indicate that the proposed indicators compensate for insufficient temporal representation by introducing aircraft-domain priors. For stronger sequence models such as LSTM and Informer, the improvements are comparatively smaller but remain consistently positive. Notably, the Accuracy of Informer increases from 94.802% to 97.030%, suggesting that the proposed indicators provide complementary discriminative information even for advanced temporal models.

For BiGRU, the introduction of structured indicators slightly decreases Recall while improving Precision, resulting in a comparable overall F1-score. This trade-off reflects a more conservative decision boundary that reduces false positives at the cost of a slight increase in missed detections, which is common in highly imbalanced classification tasks.

Overall, the proposed structured indicators and contextual features demonstrate strong generalization across different time-series architectures. Without modifying the underlying model structures, they enhance abnormal landing recognition by providing physically meaningful aircraft-domain priors and complementary discriminative information.

4.4 Comparative evaluation with baseline methods

To evaluate the overall performance of PRISM, this section compares it with the aforementioned baseline methods under a unified setting where structured safety indicators and contextual features are incorporated into all models. All methods are trained using identical data splits and experimental configurations to ensure fair comparison. By providing all models with the same structured indicators and contextual features, this controlled comparison helps distinguish the performance gains associated with the proposed architecture from those arising solely from the enriched input information.

Fig. 5 summarizes the overall performance. PRISM achieves the best results across all evaluation metrics, reaching 97.316% Accuracy, 90.817% Precision, 91.194% Recall, and 90.929% F1-score. Compared with Informer, the strongest baseline, PRISM improves the F1-score by approximately 1.7% and Precision by approximately 3.9%, demonstrating superior discrimination of abnormal landing categories while reducing false alarms. Notably, this advantage is observed even when the same structured indicators and contextual features are provided to all models, suggesting that the superior performance of PRISM is not solely attributable to the enriched input representation.

Performance differences are also evident across model architectures. Sequential models (BiGRU and LSTM) achieve competitive Accuracy but relatively lower Precision, while convolution-based models [Convolutional Neural Network (CNN) and TCN] are limited by

Table 7 Performance comparison of different models with and without the proposed indicators and contextual features.

Model	Without indicators (%)				With indicators (%)			
	Accuracy	Precision	Recall	F1-score	Accuracy	Precision	Recall	F1-score
BiGRU	94.346	78.023	94.624	84.312	95.222	79.017	90.506	84.144
LSTM	95.608	79.569	92.715	85.354	95.638	82.656	93.066	87.271
TCN	79.147	35.938	46.568	37.172	86.969	66.841	88.538	73.664
TCN+Attention	94.842	77.095	88.377	81.600	95.413	80.606	93.221	85.952
CNN	92.753	73.467	91.911	80.635	92.919	75.285	92.962	82.205
Attention	90.410	66.140	90.391	74.939	94.807	77.993	91.774	83.888
Informer	94.802	82.269	90.739	85.824	97.030	86.909	92.168	89.231
Autoformer	82.642	47.033	64.016	52.510	92.203	67.715	78.867	72.033

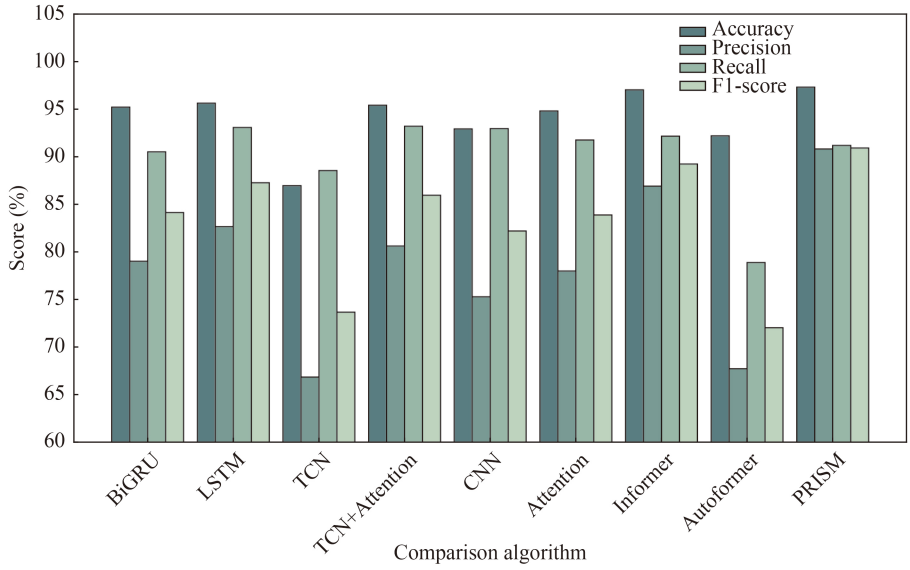


Fig. 5 Performance comparison of different models.

local receptive fields and therefore struggle to capture long-range temporal dependencies. Transformer-based models generally perform better owing to their global temporal modeling capability, with Informer achieving the strongest baseline performance. However, its classification head lacks contextual conditioning, limiting adaptability across different airports and runway configurations. In contrast, PRISM further improves performance by jointly exploiting structured safety indicators and context-aware conditional classification.

Since Informer provides the strongest baseline performance, it is selected for detailed comparison. As shown in Fig. 6, both models achieve comparable performance on the Normal category, whereas the largest difference is observed for the high-risk Path High category. Informer misclassifies 179 Path High samples as Normal (40.6%), while PRISM reduces this number to 47 (10.7%), corresponding to an error reduction of approximately 30%. Consequently, the Precision and F1-score of the Path High category increase from 66% to 86% and from 0.74 to 0.83, respectively.

The misclassification analysis indicates that Informer tends to classify certain high-risk approach samples as Normal, whereas PRISM reduces such errors through context-aware conditional classification. By incorporating operational metadata, such as airport and runway information, PRISM dynamically adapts its decision boundaries to different operational scenarios, leading to more reliable recognition of safety-critical landing states.

Furthermore, the class-wise Miss Detection Rate (MDR) comparison among PRISM, Informer With Indicators, and Informer Without Indicators further demonstrates the effectiveness of PRISM. As shown in Table 8, PRISM achieves lower MDRs for the Nominal and FlapsLate categories while maintaining competitive performance across the remaining abnormal categories. Although the Informer variants achieve lower MDRs for certain individual categories, PRISM provides a better balance between missed detections and prediction reliability, as reflected by its higher Precision and F1-score. The comparison between the two Informer variants further confirms the complementary value of structured safety indicators. More importantly, PRISM achieves superior overall performance under the same indicator information, demonstrating that its advantages cannot be attributed solely to the additional inputs.

The radar chart further confirms PRISM's superior performance in Accuracy, macro-average Precision, and the safety-critical Path High category while maintaining comparable Recall. Overall, these results demonstrate that the joint design of structured indicators, global–local temporal modeling, and context-aware conditional classification enables more balanced and reliable landing risk identification.

4.5 Ablation study on model components

To further validate the effectiveness of each component in PRISM, ablation studies are conducted on four key modules: the ProbSparse attention refinement module (Pro), the Temporal Mixer module (ME), the FiLM-based Conditional Fusion Head (FF), and the pre-training strategy (Pt). The corresponding results are presented in Table 9, with the performance variations illustrated in Fig. 7.

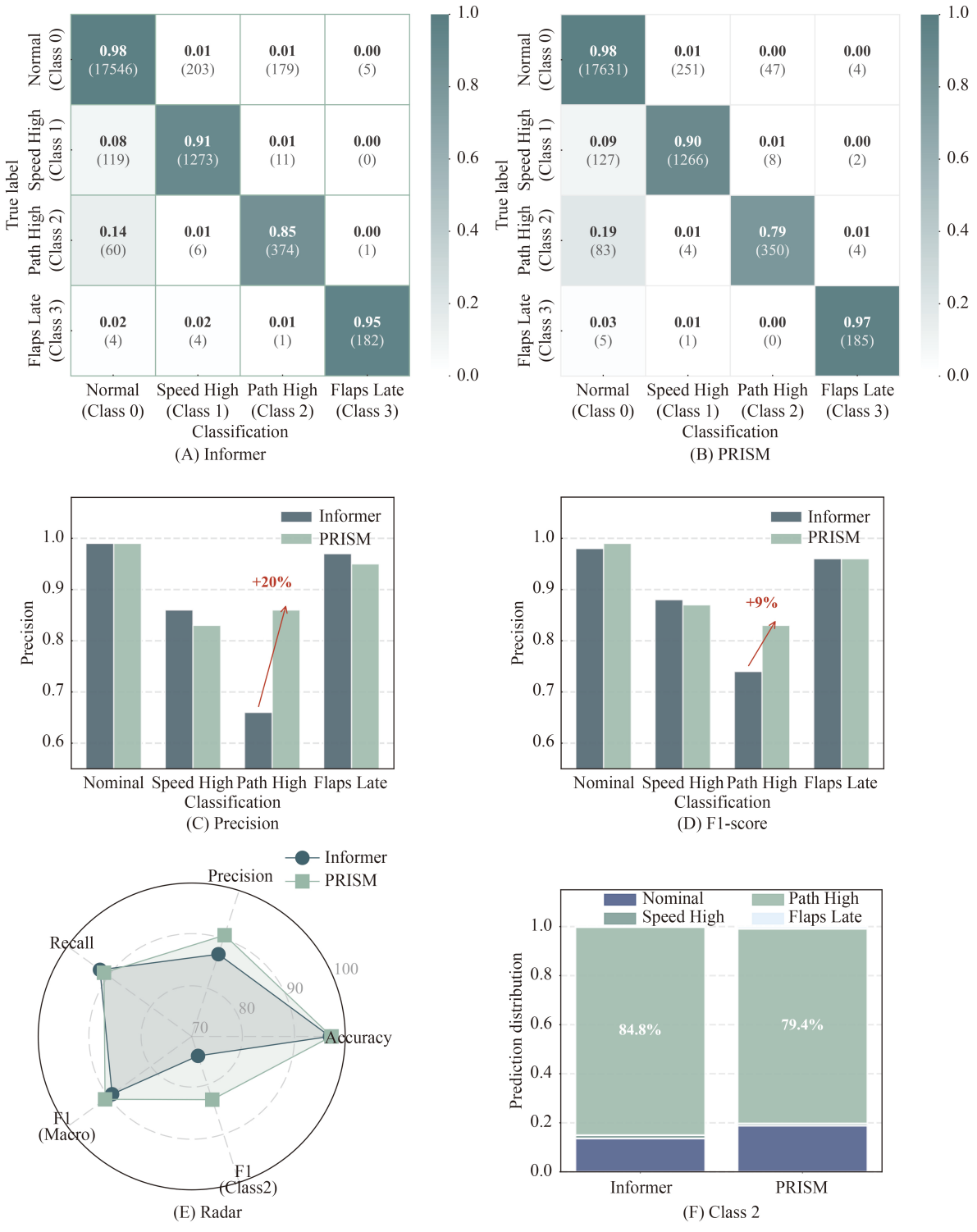


Fig. 6 PRISM vs. Informer comprehensive comparison.

As shown in Table 9, the complete PRISM achieves the best performance across all evaluation metrics, attaining 97.316% Accuracy, 90.817% Precision, 91.194% Recall, and 90.929% F1-score. Removing any component consistently degrades performance, confirming that each module contributes positively to the overall model.

Among all components, FiLM-CFH has the greatest impact. Removing this module ("-FF") reduces the F1-

Table 8 Class-wise MDR comparison among different models.

Model	Nominal	Speed High	Path High	Flaps Late
PRISM	**1.68**	9.76	20.63	**3.14**
Informer With Indicators	2.16	9.27	**15.19**	4.71
Informer Without Indicators	4.59	**8.70**	19.05	4.71

Table 9 Ablation study on the contribution of individual components.

Model	Accuracy	Precision	Recall	F1-score
Our method	97.316	90.817	91.194	90.929
-Pro	97.180	88.859	91.214	90.007
-ME	96.795	87.664	90.825	89.182
-FF	94.912	83.124	82.666	82.855
-Pt	96.860	89.140	90.567	89.804
-Pro-FF	95.142	82.388	86.944	84.521
-Pro-ME	97.150	89.539	90.378	89.954
-Pt-FF	94.702	83.737	84.598	84.132
-Pro-ME-FF	93.590	80.040	83.174	81.460

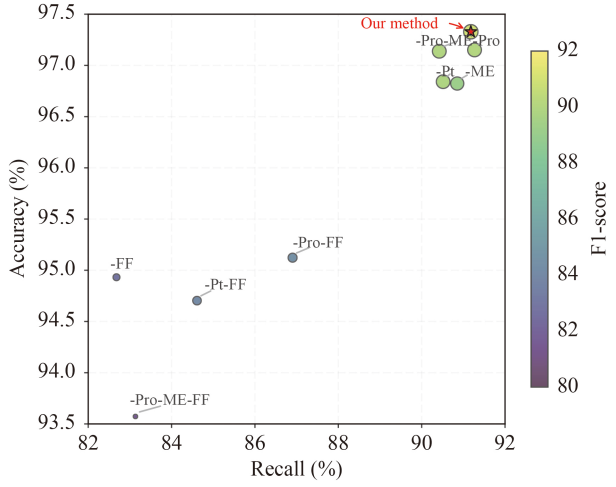


Fig. 7 Ablation study on model components.

score from 90.929% to 82.855%, accompanied by declines in both Precision and Recall. This result indicates that contextual conditioning is essential for adapting the model to different airports, runways, and operational environments.

The ProbSparse attention refinement module and the Temporal Mixer also provide consistent performance gains. Removing ProbSparse ("-Pro") decreases the F1-score by approximately 0.9%, demonstrating the importance of selectively modeling informative long-range temporal dependencies. Likewise, removing the Temporal Mixer ("-ME") degrades all evaluation metrics, with Precision decreasing from 90.817% to 87.664%, indicating that local temporal dynamics remain critical for abnormal landing recognition even in Transformer-based architectures.

The pre-training strategy further improves model generalization. Without pre-training ("-Pt"), both Accuracy and F1-score decline, suggesting that the learned temporal representations provide a stronger initialization for downstream fine-tuning, particularly under limited abnormal sample conditions. Furthermore, removing multiple modules results in larger performance degradation. For example, the "-Pro-ME-FF" configuration achieves an F1-score of only 81.460%, nearly 9.5% lower than the complete model, highlighting the complementary roles of the proposed modules.

Overall, the ablation results demonstrate that the performance improvement of PRISM is not solely attributable to the enriched input information. PSAM enhances the modeling of critical temporal dependencies, while FiLM-CFH further exploits contextual information through adaptive feature modulation. Together, these architectural components, along with the Temporal Mixer and pre-training strategy, contribute to the overall performance improvement of PRISM.

4.6 Early classification capability analysis for time-series

To further evaluate the early risk perception capability of PRISM, additional experiments are conducted under different prediction horizons by progressively increasing the forecasting length. The corresponding results are shown in Fig. 8.

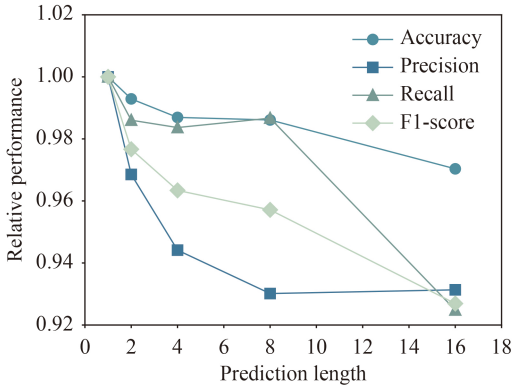


Fig. 8 Robustness analysis of classification performance relative to prediction lead time.

As the prediction horizon H increases from 1 to 16, all evaluation metrics exhibit only a slight decline, which is expected due to the increasing uncertainty associated with longer-term forecasting. Nevertheless, PRISM demonstrates strong robustness, maintaining an Accuracy above 94% and an F1-score above 84% even at $H = 16$. Compared with the short-term setting ($H = 1$), the overall performance degradation remains within approximately 6%, indicating stable long-horizon prediction capability.

From a normalized perspective, all evaluation metrics remain above 0.9 across different prediction horizons, demonstrating consistent discriminative performance. In

particular, Recall decreases more gradually over the intermediate range ($H \in [4,8]$), suggesting that PRISM captures latent risk patterns during the early stages of approach. Although Precision shows a slight decline with increasing horizon, the degradation remains limited and acceptable for operational applications.

These results demonstrate that PRISM is capable of identifying unstable landing risks multiple time steps in advance rather than merely recognizing the current flight state. Such early risk perception provides valuable warning time for pilots and decision-support systems to implement corrective actions before safety risks further develop, highlighting the practical applicability of PRISM in real-world aviation safety scenarios.

4.7 SHAP-Based interpretability analysis

To investigate how input variables contribute to PRISM predictions, SHAP is employed from three perspectives: feature importance, class-level contributions, and temporal contribution patterns. This analysis provides insights into model behavior and evaluates whether the identified influential features align with aircraft operational knowledge. Note that SHAP values reflect feature contributions to model predictions rather than causal relationships; therefore, the interpretation is limited to explaining model behavior rather than inferring causal mechanisms.

4.7.1 Feature-level analysis

To examine how the proposed structured indicators affect feature learning, we compare SHAP distributions before and after indicator augmentation. As shown in Fig. 9, the introduced indicators clearly redistribute feature importance. Before augmentation, the model relies heavily on a few dominant variables, particularly F11 (Localizer Deviation) and F6 (Drift Angle). After augmentation, their importance is moderated, while previously less influential features, including F9 (Glideslope Deviation), F4 (Airspeed), and F17 (True Heading), contribute more substantially.

These results indicate that the proposed structured indicators encourage PRISM to integrate information from multiple complementary flight parameters rather than relying on a small number of dominant variables. The resulting feature attribution is more balanced and better reflects the multi-factor decision process underlying aircraft operation, thereby improving both the robustness and interpretability of abnormal landing recognition.

4.7.2 Class-level analysis

To further investigate the discriminative basis for different abnormal categories, Fig. 10 illustrates the SHAP heatmap distributions in the joint “time–feature”

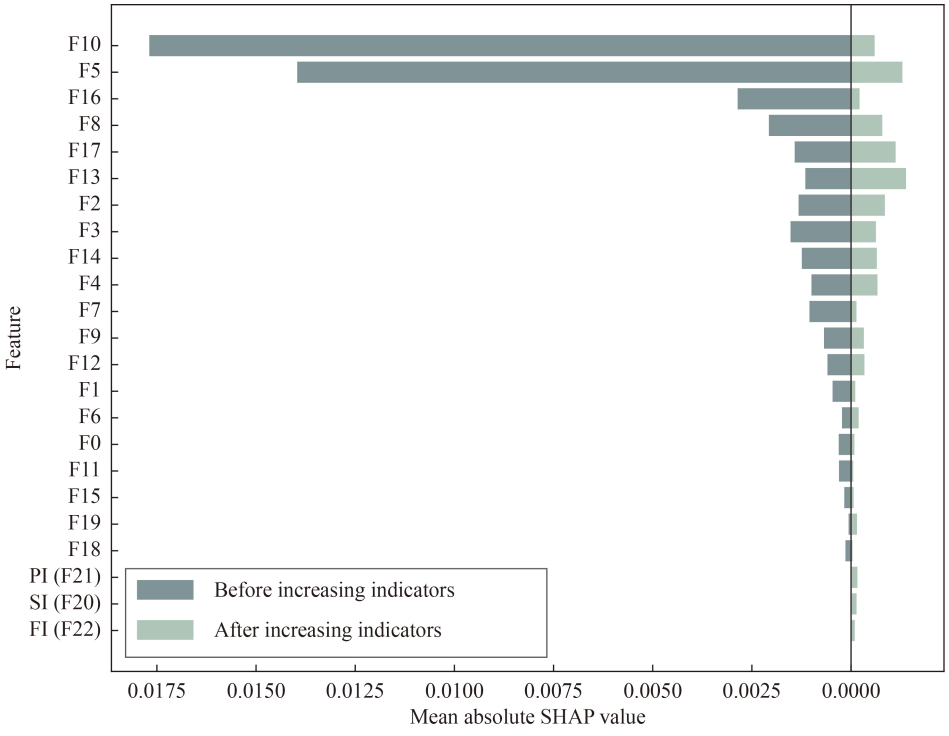


Fig. 9 SHAP-based feature importance analysis before and after incorporating structured safety indicators.

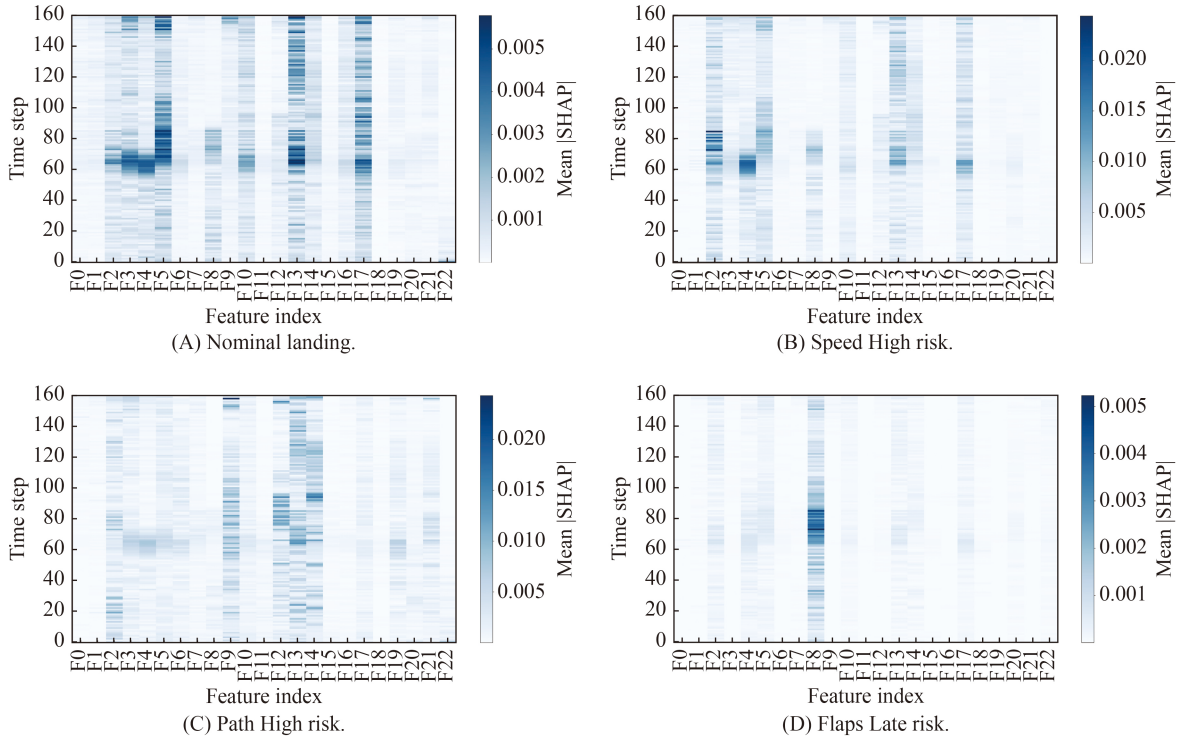


Fig. 10 SHAP heatmap visualization of temporal-feature contributions across different landing risk categories.

space for each category.

For the Nominal category (Class 0), the model mainly relies on the joint responses of F2 (Angle of Attack), F4 (Airspeed), and F5 (Selected Course) during steps 60–90, indicating that the coordinated stability of aircraft attitude, speed, and heading is the primary characteristic of normal approaches.

For the Speed High category (Class 1), F9 (Glideslope Deviation), F13 (Total Pressure), and F14 (Pitch Angle) dominate the middle-to-late approach phase (steps 60–100), with F14 remaining influential over an extended period. This suggests that the model captures the sustained trajectory deviations associated with excessive approach speed and insufficient pitch adjustment.

For the Path High category (Class 2), F8 (Flap Position) exhibits the strongest contribution, with the highest SHAP responses concentrated around steps 65–90. This indicates that abnormal flap configuration is a key discriminative characteristic of trajectory anomalies, consistent with operational practices in which configuration adjustments are often used to compensate for deviations from the desired approach path.

For the Flaps Late category (Class 3), the importance distribution becomes more dispersed. Besides F8, variables such as F2, F4, F5, F13, and F17 also contribute, with high-response regions extending from steps 60 to 160. This result suggests that configuration anomalies arise from the coupled evolution of aircraft configuration, speed, heading, and aerodynamic states, requiring the model to integrate multidimensional information over a longer temporal horizon.

4.7.3 Temporal analysis

Fig. 11 shows the distribution of mean absolute SHAP values across the approach sequence. Rather than being uniformly distributed, feature importance is concentrated at several critical stages of flight.

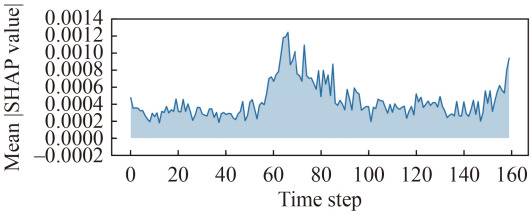


Fig. 11 Time step importance (averaged over features and classes).

The highest SHAP values occur around steps 60–75, corresponding to the stabilized approach phase, during which aircraft speed, trajectory, and configuration undergo coordinated adjustments. This indicates that PRISM primarily relies on the evolution of aircraft states within this critical interval to distinguish normal and

abnormal landing behaviors.

A second increase is observed near steps 155–160, immediately before touchdown, suggesting that final flight adjustments, such as attitude correction, thrust control, and configuration confirmation, provide additional discriminative evidence for model decisions.

Overall, the SHAP analysis demonstrates that PRISM captures physically meaningful decision patterns across temporal, feature, and category dimensions. The learned attribution patterns are well aligned with aircraft operational principles, indicating that the proposed structured indicators improve both the interpretability and discriminative capability of abnormal landing recognition.

5 Conclusions and Discussion

This paper presented PRISM, a unified model for unstable approach and landing recognition that integrates structured flight-mechanism indicators, sparse-mix temporal modeling, and context-aware conditional classification. Unlike conventional data-driven methods relying solely on raw flight sequences, PRISM explicitly incorporates flight safety knowledge by modeling speed, trajectory, and configuration deviations, enabling more discriminative and physically interpretable representations.

Extensive experiments show that PRISM consistently outperforms existing methods in accuracy and robustness. The results further demonstrate that combining mechanism-guided representations with temporal feature learning and contextual decision modeling improves generalization across diverse operational scenarios while providing interpretable decision evidence consistent with established flight operation principles.

Future work will incorporate additional operational factors, such as weather conditions, pilot actions, and air traffic control information, to improve adaptability in complex environments. Online abnormal landing recognition and foundation-model-assisted reasoning will also be explored to support real-time and intelligent aviation safety management.

Data availability No datasets were generated or analyzed for this work.

Acknowledgments This work was supported by the National Natural Science Foundation of China under Grants (U2570207, 62503506), the Key Research and Development Program of Hunan Province under Grant (2025AQ2027), and the Science and Technology Innovation Program of Hunan Province under Grant (2026RC1005). The authors declare that no generative artificial intelligence (AI) or AI-assisted technologies were used in the preparation, analysis, or writing of this manuscript.

Competing interests The authors declare no conflicts of interest.

References

1. International Air Transport Association. IATA releases 2024 safety report. Available online (accessed 22 December 2025)
2. International Air Transport Association. Interactive safety report. Available online (accessed 23 December 2025)
3. Odisho E V, Truong D, Joslin R E. Applying machine learning to enhance runway safety through runway excursion risk mitigation. Journal of Aerospace Information Systems, 2022, 19(2): 98−112
4. Qi X G, Liu C, Fan D M, et al. Causal modeling and risk assessment of runway excursion events via interpretable machine learning. Reliability Engineering & System Safety, 2026, 269: 112055
5. Wang W C, Chen S Y, Zhang Y C, et al. Multimodal data fusion for risk assessment of civil aviation hard landing events. Journal of Air Transport Management, 2026, 135: 103024
6. Lu Y, Zhu T. Pre-training of autoregressive model for aircraft hard landing prediction based on QAR data. In: 2019 IEEE 31st International Conference on Tools with Artificial Intelligence (ICTAI). Portland, USA, 2019, 1613–1617
7. Ismail Fawaz H, Forestier G, Weber J, et al. Deep learning for time series classification: a review. Data Mining and Knowledge Discovery, 2019, 33(4): 917−963
8. Shang J X, Li X Q, Zhang R X, et al. A dual two-stage attention-based model for interpretable hard landing prediction from flight data. Engineering Applications of Artificial Intelligence, 2025, 154: 110911
9. Li K P, Xu J J, Zhao H M, et al. A clustered borderline synthetic minority over-sampling technique for balancing quick access recorder data. Journal of Intelligent & Fuzzy Systems, 2023, 45(4): 6849−6862
10. Geng Y, Luo X Y. Cost-sensitive convolutional neural networks for imbalanced time series classification. Intelligent Data Analysis, 2019, 23(2): 357−370
11. Kong Y X. Deep learning models for aircraft landing safety assessment. Nashville: Vanderbilt University, 2022
12. Flight Safety Foundation. ALAR briefing note 7-1: Stabilized approach. Available online (accessed 23 December 2025)
13. International Air Transport Association. Examining unstable approaches: Risk mitigating efforts. Available online (accessed 23 December 2025)
14. Blajev T, Curtis W. Go-around decision-making and execution project. Alexandria: Flight Safety Foundation, 2017
15. Das S, Matthews B L, Lawrence R. Fleet level anomaly detection of aviation safety data. In: 2011 IEEE Conference on Prognostics and Health Management. Denver, CO, USA, 2011, 1–10
16. Turgut E T. A 25-year journey in quick access recorder (QAR) data: a thematic review of analytics for aircraft safety, efficiency, and health management. Progress in Aerospace Sciences, 2026, 162: 101204
17. Wang L, Wu C X, Sun R S. An analysis of flight quick access recorder (QAR) data and its applications in preventing landing incidents. Reliability Engineering & System Safety, 2014, 127:

86−96
18. Hu C, Zhou S H, Xie Y, et al. The study on hard landing prediction model with optimized parameter SVM method. In: 2016 35th Chinese Control Conference (CCC). Chengdu, China, 2016, 4283–4287
19. Qiao X D, Chang W B, Zhou S H, et al. A prediction model of hard landing based on RBF neural network with K-means clustering algorithm. In: 2016 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM). Bali, Indonesia, 2016, 462–465
20. Kang Z W, Shang J X, Feng Y, et al. A deep sequence-to-sequence method for accurate long landing prediction based on flight data. IET Intelligent Transport Systems, 2021, 15(8): 1028−1042
21. Zhang H C, Zhu T Y. Aircraft hard landing prediction using LSTM neural network. In: Proceedings of the 2nd International Symposium on Computer Science and Intelligent Control. Stockholm, Sweden, 2018, 28
22. Chen H D, Shang J X, Zheng L J, et al. SDTAN: scalable deep time-aware attention network for interpretable hard landing prediction. IEEE Transactions on Intelligent Transportation Systems, 2023, 24(9): 10211−10223
23. Li X, Shang J X, Zheng L J, et al. IMTCN: an interpretable flight safety analysis and prediction model based on multi-scale temporal convolutional networks. IEEE Transactions on Intelligent Transportation Systems, 2024, 25(1): 289−302
24. Dang X, Hua C C, Rong C T. Automated anomaly detection and causal analysis for civil aviation using QAR data. Applied Sciences, 2025, 15(5): 2250
25. Ismail Fawaz H, Lucas B, Forestier G, et al. InceptionTime: Finding AlexNet for time series classification. Data Mining and Knowledge Discovery, 2020, 34(6): 1936−1962
26. Dempster A, Petitjean F, Webb G I. ROCKET: exceptionally fast and accurate time series classification using random convolutional kernels. Data Mining and Knowledge Discovery, 2020, 34(5): 1454−1495
27. Foumani N M, Tan C W, Webb G I, et al. Improving position encoding of transformers for multivariate time series classification. Data Mining and Knowledge Discovery, 2024, 38(1): 22−48
28. Tang Y X, Wei Y X, Hu Y P, et al. Convolutional network integrated with frequency adaptive learning for multivariate time series classification. ACM Transactions on Knowledge Discovery from Data, 2025, 19(8): 1−23
29. Foumani N M, Miller L, Tan C W, et al. Deep learning for time series classification and extrinsic regression: a current survey. ACM Computing Surveys, 2024, 56(9): 1−45
30. Chawla N V, Bowyer K W, Hall L O, et al. SMOTE: synthetic minority over-sampling technique. Journal of Artificial Intelligence Research, 2002, 16: 321−357
31. Chen W X, Yang K X, Yu Z W, et al. A survey on imbalanced learning: latest research, applications and future directions. Artificial Intelligence Review, 2024, 57(6): 137
32. Ross T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection. In: Proceedings of the IEEE International Conference on Computer Vision. Venice, Italy, 2017, 2999–3007
33. Araf I, Idri A, Chairi I. Cost-sensitive learning for imbalanced medical data: a review. Artificial Intelligence Review, 2024, 57(4): 80
34. Matthews B. Curated 4-class anomaly detection data set. Available online (accessed 23 December 2025)



Fanbiao Li (Senior Member, *IEEE*) received the B.S. degree in Applied Mathematics from Mudanjiang Normal University, Mudanjiang, China, in 2008, the M.S. degree in Operational Research and Cybernetics from Heilongjiang University, Harbin, China, in 2012, and the Ph.D. degree in Control Theory and Control Engineering from the Harbin Institute of Technology, Harbin, in 2015. His research interests include control and design of aircraft brake system, sliding mode control, and fault diagnosis and identification. Prof. Li currently serves as an Associate Editor for a number of journals, including *IEEE Transactions on System, Man, and Cybernetics: Systems*, *IEEE Transactions on Fuzzy Systems*, etc.



Xiong Peng received the B.S. degree in Information and Computing Science from Hengyang Normal University, Hengyang, China, in 2022, and the M.S. degree in Computer Science and Technology from the Civil Aviation University of China, Tianjin, China, in 2026. He is currently working toward the Ph.D. degree in Control Science and Engineering with Central South University, Changsha, China. His research interests include aircraft landing, intelligent control, reinforcement learning.



Yiyun Zhao (Member, *IEEE*) received the B.E. and Ph.D. degrees in Electrical Engineering from Northwestern Polytechnical University, Xi'an, China, in 2017 and 2024, respectively. In 2024, he joined the Central South University, as a Lecturer. His research interests include more electric aircraft, nonlinear control, and mechatronic servo control.



Jose I. Leon (*IEEE* Fellow) was born in Cadiz, Spain. He received the B.S., M.S., and Ph.D. degrees in Telecommunications Engineering from the Universidad de Sevilla, Seville, Spain, in 1999, 2001, and 2006, respectively. He is currently an Associate Professor with the Universidad de Sevilla, ENGREEN. Since 2019, he has also been a Chair Professor with Harbin Institute of Technology, Harbin, China. His research interests include modulation and control of power converters for high-power

applications and renewable energy systems. Dr. Leon was the corecipient of the 2008, 2015, 2020, and 2022 Best Paper Award of IEEE Industrial Electronics Magazine, and the 2012 Best Paper Award of IEEE Transactions on Industrial Electronics and the recipient of the 2014 IEEE J. David Irwin Industrial Electronics Society Early Career Award, the 2017 IEEE Bimal K. Bose Energy Systems Award, and the 2017 Manuel Losada Villasante Award.



Weihua Gui (Member, *IEEE*) received the B.S. degree in Electrical Engineering and the M.S. degree in Automatic Control Engineering from Central South University in 1976 and 1981, respectively. From 1986 to 1988, he was a Visiting Scholar at University-GH-Duisburg, Germany. Since 1991, he has been a Full Professor at Central South University. Since 2013, he has been an Academician with the Chinese Academy of Engineering. His main research interests include modeling and optimal control of complex industrial processes, distributed robust control, and fault diagnosis.



Haonan Xu received M.S. degrees from Fudan University, Shanghai, China and Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany. He is currently a senior engineer and leads the Monte Carlo Analysis and Applications Group at the COMAC Innovation Valley, China. He was involved throughout the testing, flight testing, and airworthiness certification of the C919 automatic flight system and served as a core team member in addressing key technical challenges in automatic landing. His experience spans aircraft development, technology research, and technical project management. His research interests include uncertainty quantification, Monte Carlo analysis, and data science for aerospace applications.