Computers, Materials & Continua

Tech Science Press

https://doi.org/10.32604/cmc.2026.086522

ARTICLE

RSTD-KD: A Task-Cost-Aware Approach for UAV Communication Risk Warning via Knowledge Distillation

Caiyun Li[1,#], **Xiaowen Liu**[1,#], **Binhui Tang**[1,*], **Tingting Lu**[2], **Daibo Xiao**[3] and **Li Chen**[3]

[1]School of Artificial Intelligence, Hainan Normal University, Haikou, China
[2]College of Computer Science, Sichuan University, Chengdu, China
[3]Faculty of Humanities and Social Sciences, City University of Macau, Macau, China

*Corresponding Author: Binhui Tang. Email: tangbh@hainnu.edu.cn

[#]These authors contributed equally to this work

Received: 01 June 2026; Accepted: 01 July 2026; Published: 13 August 2026

ABSTRACT: Unmanned aerial vehicle (UAV) communication links in low-altitude operations support mission control, status feedback, and data transmission. Abnormal communication states may affect mission continuity and risk response. However, conventional anomaly detection usually determines only whether a communication anomaly exists, making it difficult to support risk-state assessment and alert-threshold decision-making. To address this problem, this paper proposes Risk Stratification and Task-cost-aware Decision via Knowledge Distillation (RSTD-KD), a task-cost-aware UAV communication risk warning approach that integrates risk stratification, probability calibration, and threshold decision-making. RSTD-KD aggregates fine-grained communication records into window-level behavior samples, constructs low-, medium-, and high-risk reference states, and introduces distillation-guided risk stratification to transfer binary anomaly-boundary information and soft attack-probability supervision into three-level risk-state learning. The estimated attack probabilities are then calibrated for threshold search and cost calculation. Cross-sea logistics and general aviation inspection are used as representative task-cost scenarios to examine the effect of false-positive and false-negative costs on alert-threshold selection. Experiments on the public ECU-IoFT dataset show that RSTD-KD achieves 93.75% Accuracy, 96.44% Balanced Accuracy, and 94.65% Macro-F1. Platt Scaling reduces the Brier Score from 0.0762 to 0.0113. In the additional external-dataset portability evaluation on UAVIDS-2025, RSTD-KD achieves 94.92% Accuracy and 94.38% Macro-F1 across five random seeds, providing further evidence of the portability of RSTD-KD under different UAV communication data granularity and attack taxonomy. These results show that RSTD-KD provides an effective path for converting UAV communication detection outputs into risk states, calibrated attack probabilities, and task-oriented alert decisions.

KEYWORDS: UAV communication intrusion detection; knowledge distillation; risk stratification; probability calibration; task-cost-aware threshold decision-making

1 Introduction

The development of the low-altitude economy is pushing unmanned aerial vehicles (UAVs) from isolated task tools toward continuous mission platforms [1]. In low-altitude tasks such as logistics transportation [2], general aviation inspection [3], maritime support [4], and emergency monitoring [5], UAV communication links support not only remote control and status feedback, but also mission data transmission, anomaly alerts, and operation monitoring [6]. When communication links are affected by interference,

Copyright © 2026 The Authors. Published by Tech Science Press. This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

attacks, or abnormal fluctuations, network-level anomalies may propagate to the mission level, leading to control delay, connection interruption, state distortion, or mission deviation [7,8].

During mission execution, communication anomalies are usually handled through alert mechanisms that may trigger responses such as verification, link switching, or mission adjustment. Whether an alert should be triggered depends on model outputs and threshold decision-making [9]. This threshold-based alert process may lead to two types of decision errors: false positives may trigger unnecessary responses, while false negatives may leave real anomalies undetected. Different low-altitude operations have different tolerance levels for these two types of errors. In this study, cross-sea logistics and general aviation inspection are selected as representative scenarios of two task preferences. Cross-sea logistics places more emphasis on mission continuity and dispatch stability, where excessive false positives may disrupt transportation processes. General aviation inspection places more emphasis on risk discovery and timely response, where false negatives may delay risk handling [10]. Therefore, UAV communication security research should not only determine whether a communication process is abnormal, but also characterize communication risk states and support threshold decision-making for alert generation and mission response.

Existing learning-based studies on UAV communication security mainly focus on communication anomaly detection and intrusion identification. These methods usually build models from communication traffic features, protocol fields, or message sequences to determine whether anomalous or intrusive behaviors occur during communication processes [11–14]. Public UAV communication security datasets are also mostly organized as fine-grained communication records, protocol message sequences, or network flow features. They can support the training and performance evaluation of detection models, but usually lack risk state labels and alert decision labels for mission processes [15,16]. Therefore, models trained on such data can produce communication anomaly probabilities or intrusion risk scores, but they cannot directly support graded alerts and task-oriented response decisions.

To use detection results for security warning during low-altitude missions, it is still necessary to construct risk-state representations and convert model probabilities into reliable inputs for threshold search and alert decision-making [17]. Some studies have started to consider UAV security situation assessment and response strategies, but the connection among detection probabilities, risk states, and cost-sensitive alert thresholds remains unclear [18]. How to convert communication anomaly probabilities into low-, medium-, and high-risk states, and how to determine alert thresholds according to task costs, remain open issues.

To address these issues, this paper proposes Risk Stratification and Task-cost-aware Decision via Knowledge Distillation (RSTD-KD), a knowledge distillation-based method for UAV communication risk-state stratification and threshold decision-making. RSTD-KD aggregates fine-grained communication records into window-level communication behavior samples, constructs low-, medium-, and high-risk reference states, and introduces binary communication-anomaly boundary information and soft attack-probability supervision into three-level risk-state learning through distillation-guided risk stratification. The output attack probabilities are then calibrated and combined with task-cost constraints for threshold search, thereby generating alert thresholds and alert decisions under different task preferences. Through this design, RSTD-KD extends UAV communication detection outputs from binary anomaly identification to risk states, calibrated attack probabilities, and task-oriented alert-decision evidence.

RSTD-KD is evaluated with ECU-IoFT as the primary benchmark dataset. The experiments cover window-level risk-state recognition, distillation-guided learning, probability calibration, and task-cost-aware threshold decision-making. The results show that RSTD-KD achieves effective risk-state recognition under the public ECU-IoFT benchmark setting; soft attack-probability supervision improves risk stratification performance; and probability calibration improves the reliability of attack probabilities as inputs for threshold search and cost calculation. Task-cost-based threshold scanning further shows that different task

preferences lead to differentiated alert thresholds, enabling the same model output to support task-oriented alert decisions under different low-altitude operations. In addition, UAVIDS-2025 is used as an additional external benchmark to examine the portability of RSTD-KD under different UAV communication data granularity and attack taxonomy. These results show that RSTD-KD provides a modeling path for converting public UAV communication records into risk states, calibrated attack probabilities, and task-oriented alert-decision evidence under representative low-altitude task preferences.

The main contributions of this paper are summarized as follows.

1. This paper develops a window-level communication behavior modeling process for UAV communication risk-state stratification. By aggregating fine-grained communication records into window-level communication behavior samples, RSTD-KD captures local behavior changes in continuous communication segments and provides structured inputs for communication anomaly detection and risk-state learning.
2. This paper proposes a distillation-guided risk-state stratification learning method that transfers binary communication-anomaly boundary information into low-, medium-, and high-risk state learning. By defining the model-level attack probability as the sum of medium- and high-risk state probabilities, the method aligns binary anomaly-probability supervision with three-level risk-state stratification, improving risk-state recognition while maintaining low inference complexity.
3. This paper designs a probability calibration and task-cost-aware threshold decision-making mechanism for UAV communication warning. Based on calibrated attack probabilities and task-specific false-positive (FP) and false-negative (FN) costs, the mechanism determines alert thresholds under different task preferences. Extensive evaluation on ECU-IoFT and additional external-dataset portability analysis on UAVIDS-2025 further examine the effectiveness and portability of the proposed framework across different UAV communication data granularities and attack taxonomies.

The rest of this paper is organized as follows. Section 2 reviews related work. Section 3 presents the RSTD-KD framework and its key methods. Section 4 reports the experimental setup and results. Section 5 concludes the paper and discusses future research directions.

2 Related Work

2.1 UAV Communication Security and Link Protection

Research on UAV communication security first focuses on the security of communication links and network structures. UAV systems usually involve UAV-to-ground station communication, multi-UAV cooperative communication, and flying ad hoc network (FANET) communication. Different connection patterns may expose the system to eavesdropping, jamming, spoofing, replay attacks, malicious nodes, and routing attacks [19]. Existing surveys have classified UAV security issues from the perspectives of hardware, software, communication, and sensors. They also show that communication links remain one of the most vulnerable components in UAV systems [20]. In terms of protection mechanisms, existing studies mainly investigate physical-layer security, cryptographic authentication, secure transmission, lightweight key management, and FANET authentication mechanisms [21–23]. These studies provide an important basis for UAV communication link protection. However, most of them focus on attack surface analysis and protection mechanism design. They pay less attention to how detection results can be converted into risk states and task-oriented alert decisions during mission execution.

2.2 Learning-Based Anomaly Detection and Communication Behavior Modeling

With the growth of UAV communication data and public security datasets, learning-based anomaly detection has become an important direction in UAV communication security. Existing studies usually extract features from traffic statistics, protocol fields, flow-level features, or UAV operational states. They then use machine learning, deep learning, or ensemble models for anomaly detection and attack classification. Mohammed and Fourati systematically analyzed intrusion detection system (IDS) datasets for intra- and inter-UAV communication systems. They showed that data sources, attack coverage, and communication scenarios directly affect the applicability of intelligent detection methods [24]. Miao et al. proposed a deep-meta-heuristic UAV intrusion detection system to improve detection performance in complex communication scenarios [25]. Later studies further explored artificial intelligence (AI)-enhanced detection, stratified sampling, and ensemble learning to improve the stability of UAV network intrusion detection [26,27].

Recent UAV/drone security studies have further introduced graph-based, attention-based, and Transformer-based modeling paradigms. Mughal et al. [28] proposed a topology-aware graph neural network (GNN)-based IDS for UAV swarms by exploiting spatial relationships among UAVs. Aldossary et al. [29] developed a Cross-Layer Convolutional Attention Network (CLCAN) for drone-to-drone and drone-to-base-station communications, using multi-scale convolution and attention-based fusion to capture complex communication abnormalities. Zhao et al. [18] proposed TransReSE for UAV swarm network security situation assessment by integrating Transformer-based dependency modeling and multi-scale deep feature extraction. These works show that UAV/drone security modeling has moved beyond conventional tree-based or shallow neural classifiers toward topology-aware, attention-enhanced, and Transformer-based deep architectures.

In addition to centralized or single-site learning-based UAV IDS methods, federated-learning-based UAV IDS has also attracted attention for distributed UAV and FANET environments. Recent studies have used federated training to support privacy-preserving intrusion detection, client-side data isolation, incremental model updating, or collaborative learning across multiple UAV nodes [14,30]. These studies highlight the importance of distributed training and deployment protocols for UAV network security. However, the present work focuses on centralized window-level risk-state stratification and task-cost-aware threshold decision-making under a unified feature space and data split.

Collectively, these studies demonstrate that learning-based UAV IDS has evolved from conventional machine-learning and ensemble classifiers toward topology-aware, attention-enhanced, Transformer-based, and federated/distributed modeling paradigms. However, most methods still take detection accuracy or attack classification as the final objective. Few of them explicitly connect window-level communication behavior modeling with risk stratification and subsequent threshold decision-making.

2.3 Risk Assessment, Risk Stratification, and Probability Reliability

Some studies have moved from simple detection to risk assessment and security situation analysis. For example, Miao and Pan developed a UAV cyber range risk assessment model based on Bayesian–Nash equilibrium. Their model was used to analyze attack-defense strategies and risk changes [31]. Sun et al. modeled UAV operational risk factors and their dependencies from accident data. Their work shows that UAV risk analysis needs to consider the coupling among multiple risk factors [32]. These studies highlight the importance of risk modeling. Recent UAV swarm security situation assessment studies have further attempted to estimate the overall network security state and threat level from traffic and threat data. For example, TransReSE integrates Transformer, ResNeXt, and squeeze-and-excitation modules for UAV swarm network security situation assessment [18]. Nevertheless, these methods mainly focus on system-level

risk, operational safety risk, or global network threat assessment, rather than calibrated window-level risk probabilities and task-specific alert thresholds for UAV mission execution.

They do not directly address how communication detection probabilities should be mapped into low-, medium-, and high-risk states. At the same time, probability calibration studies have shown that classifier probabilities are not naturally reliable probability estimates. Calibrated probabilities are more suitable for uncertainty representation, cost-sensitive decision-making, and threshold decision-making [33]. Therefore, UAV communication risk stratification requires not only an anomaly detection model, but also reliable attack-probability estimates and well-defined risk-state boundaries.

2.4 Cost-Sensitive Learning and Threshold Decision-Making

In anomaly detection and classification tasks, a fixed threshold cannot always meet practical decision requirements. Leevy et al. showed that output-threshold adjustment can improve model performance in imbalanced classification without changing the training data, and that the decision threshold can be selected according to different evaluation metrics [34]. Gutiérrez-López et al. further discussed decision thresholds for rebalanced classification ensembles from a Bayesian threshold perspective [35]. These studies indicate that a threshold selection process is still needed between model probability outputs and final alert decisions.

For low-altitude UAV tasks, threshold selection depends not only on model discrimination performance but also on task-specific cost preferences. For example, mission-continuity-oriented tasks may require stricter false-alarm control, whereas risk-discovery-oriented tasks may place greater emphasis on reducing missed alarms. Existing studies on threshold selection and decision-making provide a useful methodological basis for this work. However, their application context is usually not UAV communication security, and they rarely integrate detection probabilities, risk states, and task costs into a unified task-oriented risk warning process.

Overall, existing studies have provided useful foundations for UAV communication security, learning-based anomaly detection, risk assessment, and threshold decision-making. Recent UAV/drone security methods have further explored topology-aware GNNs, convolutional-attention architectures, and Transformer-based security situation assessment, which strengthen the modeling capability for complex UAV/drone communication behaviors. However, most of them focus separately on communication anomaly identification, system-level risk modeling, or classification threshold selection. A complete modeling process from fine-grained communication records to three-level risk states and then to cost-sensitive alert thresholds is still lacking. To address this methodological disconnect, this paper proposes RSTD-KD, which integrates window-level communication behavior modeling, knowledge distillation-guided risk stratification, probability calibration, and task-cost-aware threshold decision-making. Through this design, UAV communication detection outputs are further converted into risk-state representations, calibrated probabilities, and alert-decision evidence for low-altitude mission processes.

3 Methodology

This section presents RSTD-KD, a knowledge distillation-based method for risk-state stratification and threshold decision-making in low-altitude UAV communication security. RSTD-KD takes UAV communication behavior samples as input and performs communication-anomaly detection, risk-state learning, attack-probability estimation, probability calibration, and threshold decision-making. For packet-level communication records, behavior samples are constructed by organizing raw records into fixed-length communication windows. For flow-level communication records, each flow can be treated as a behavior sample after dataset-specific feature encoding and normalization. In this paper, ECU-IoFT is used as the

primary packet-level benchmark to instantiate the full window-based modeling process, while UAVIDS-2025 is used as an additional flow-level external benchmark for external-dataset portability analysis.

RSTD-KD consists of four main stages, as illustrated in Fig. 1. First, raw communication records are processed through field normalization, chronological ordering, window or flow partitioning, and feature encoding and aggregation to form communication behavior samples. Second, reference risk-state labels are constructed as the Low, Medium, and High labels according to the potential impact of different communication behaviors on communication continuity, authentication or control security, routing integrity, and service degradation. Third, the risk stratification learning module adopts a distillation-guided mechanism that transfers binary communication-anomaly boundary information into three-level risk-state learning. Under the joint constraint of reference risk-state labels and soft attack-probability supervision, the risk stratification model learns low-, medium-, and high-risk reference states and outputs the attack probabilities required for subsequent probability calibration and threshold decision-making. Finally, the calibrated attack probabilities are used for threshold search and task-cost analysis, thereby generating task-cost-aware alert thresholds and task-oriented alert decisions for different task-cost scenarios.

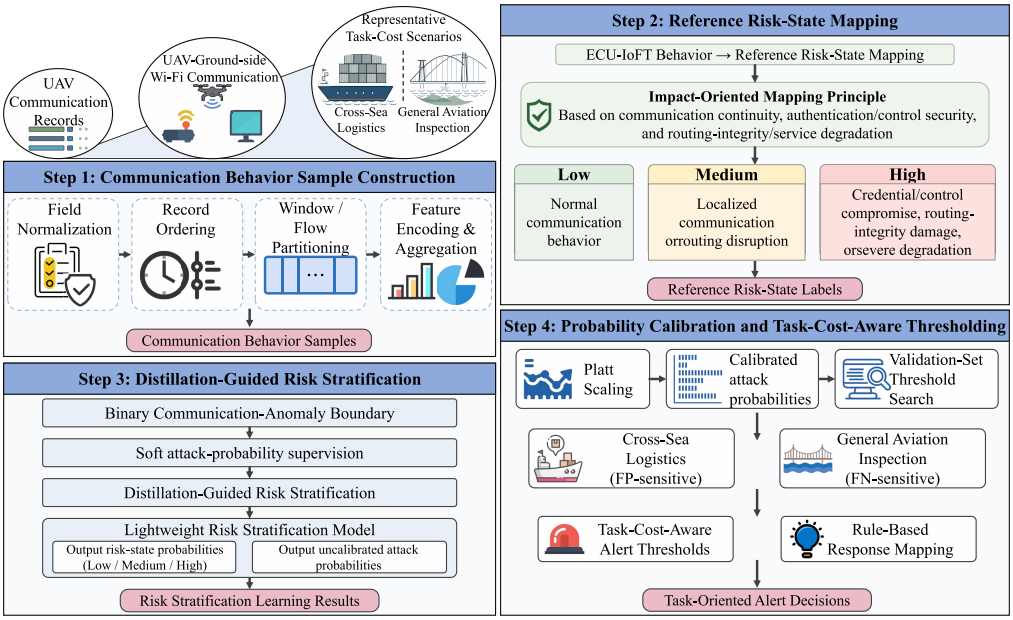

Figure 1: Overall framework of RSTD-KD. The framework converts UAV communication behavior samples into risk states, calibrated attack probabilities, and task-oriented alert decisions through behavior-sample construction, impact-oriented reference risk-state mapping, distillation-guided risk-state stratification, probability calibration, and task-cost-aware threshold decision-making.

3.1 Problem Definition and Overall Framework of RSTD-KD

Given a set of UAV communication behavior samples, the objective of this study is not only to estimate whether a sample belongs to the communication-anomaly class, but also to generate risk states and alert thresholds for task-oriented warning. Let the input feature vector of the i-th behavior sample be defined as

$$\mathbf{x}_i = [x_{i1}, x_{i2}, \ldots, x_{id}], \tag{1}$$

where d denotes the feature dimension. For packet-level UAV communication records, $\mathbf{x}_i$ is obtained by aggregating consecutive records within a communication window. For flow-level UAV communication

records, $\mathbf{x}_i$ can be directly constructed from one flow record after feature encoding and normalization. This distinction ensures that the sample-construction protocol is matched to the data granularity of each dataset. The corresponding binary label is defined as

$$y_i \in \{0, 1\}, \tag{2}$$

where $y_i = 0$ indicates a normal communication sample and $y_i = 1$ indicates a communication-anomaly or attack sample. The reference risk-state label is defined as

$$s_i \in \{\text{Low}, \text{Medium}, \text{High}\}. \tag{3}$$

In this study, the attack probability denotes the probability that a behavior sample contains communication anomalies or attack behavior. It represents the binary communication-anomaly likelihood, whereas the Low, Medium, and High labels denote reference risk states constructed from the operational impact of communication behaviors. RSTD-KD produces three types of outputs: the attack probability, the predicted risk state, and the task-oriented alert decision determined by the selected threshold. Unlike conventional binary communication-anomaly detection methods, RSTD-KD does not stop at attack identification. Instead, it further uses the estimated attack probability for risk-state stratification, probability calibration, and task-cost-aware threshold decision-making.

3.2 Communication Behavior Sample Construction and Risk-State Mapping

The raw ECU-IoFT data are organized as fine-grained communication records. A single record can reflect an instantaneous communication state, but it is insufficient to characterize the temporal variation of attack behavior within a local communication segment. To obtain a more stable behavior representation, fixed-length non-overlapping windows are constructed. Each window contains w consecutive communication records, and the stride is also set to w. In the experiments, $w = 32$. This setting avoids information overlap between adjacent windows while preserving local communication behavior patterns.

Before window construction, raw records are processed through field normalization, timestamp parsing, label unification, and missing value handling. Window-level features are then extracted from fields such as Source, Destination, Protocol, Info, and Length. These features include packet length statistics, Info field length, address diversity, communication endpoint patterns, protocol composition ratios, and Info keyword ratios. They are used to describe communication behavior patterns within a window, rather than relying on the instantaneous state of a single record.

Risk state construction converts original attack labels into trainable reference risk states. The mapping rules are defined as follows:

$$\begin{aligned} \text{No Attack} &\rightarrow \text{Low}, \\ \text{Wi-Fi Deauthentication Attack} &\rightarrow \text{Medium}, \\ \text{WPA2-PSK Wi-Fi Cracking Attack} &\rightarrow \text{High}, \\ \text{Tello API Exploit} &\rightarrow \text{High}. \end{aligned}$$

These labels serve as reference risk states for supervised learning. The mapping is constructed according to the expected impact of each behavior on communication continuity, authentication security, and control-interface security. No Attack is mapped to the Low-risk state because it corresponds to normal communication behavior and is used as the normal-risk reference. Wi-Fi Deauthentication Attack is mapped to the Medium-risk state because it mainly disrupts communication availability by forcing disconnection

from the wireless link. Although it may interrupt command transmission or status feedback, it does not directly indicate credential compromise or control-interface exploitation. WPA2-PSK Wi-Fi Cracking Attack is mapped to the High-risk state because successful cracking may expose authentication credentials and enable unauthorized access to the UAV communication network. Tello API Exploit is also mapped to the High-risk state because it targets the control interface and may directly affect UAV command execution or mission behavior. Therefore, Medium denotes an attack state that requires verification and alert confirmation, whereas High denotes a high-risk attack state that requires prioritized handling.

This rule-based mapping is not intended to define a universal severity standard for all UAV systems. Instead, it provides a consistent reference-label construction strategy for the ECU-IoFT benchmark setting. More generally, RSTD-KD follows an impact-oriented mapping principle: normal communication behavior is mapped to Low, localized communication or routing disruption is mapped to Medium, and attacks that may cause severe service degradation, authentication or identity compromise, control-interface abuse, or routing-integrity damage are mapped to High. When this principle is applied to UAVIDS-2025, dataset-specific traffic categories are mapped into the same Low/Medium/High reference-state structure according to their communication impact. The same reference risk-state labels are used for all compared models within each dataset, ensuring that the evaluation focuses on risk-state recognition capability rather than different label definitions.

3.3 Risk Stratification Learning and Attack Probability Estimation

Risk stratification requires learning both binary communication-anomaly detection information and the boundaries among low-, medium-, and high-risk reference states. Fig. 2 shows the distillation-guided risk stratification module in RSTD-KD. This module consists of input and supervision, a Teacher model for binary communication-anomaly estimation, a distillation constraint, and a lightweight risk stratification model. During training, both the Teacher and the Student receive the same window-level communication behavior features as input. The teacher model learns the binary communication-anomaly boundary and outputs soft attack probabilities as probability supervision. The lightweight risk stratification model learns the three-level risk state boundary under the joint supervision of reference risk state labels and soft attack-probability supervision. It also outputs the attack probabilities required for subsequent probability calibration and threshold decision-making. During inference, only the lightweight risk stratification model is retained to reduce deployment complexity. The key design of this module is that the Teacher and the Student are not trained for the same output space. The teacher model learns a binary communication-anomaly boundary, whereas the Student learns three-level risk states. By defining the model-level attack probability as the sum of the Medium and High risk-state probabilities, RSTD-KD transfers binary anomaly-boundary information into three-level risk stratification learning. Therefore, the distillation term is used not only for model compression, but also for boundary regularization between binary anomaly detection and risk state recognition.

In the implementation, the Teacher model is instantiated by HistGradientBoosting Classifier. Given $\mathbf{x}_i$, the Teacher outputs the probability that the sample belongs to the communication-anomaly class:

$$p_i^T = P_T(y_i = 1 \mid \mathbf{x}_i), \tag{4}$$

where $y_i = 1$ denotes a window containing communication anomalies or attack behavior, $P_T(\cdot)$ denotes the probability output function of the Teacher, and p_i^T denotes the attack probability produced by the Teacher. The Teacher is not used for direct three-level risk state prediction. Its role is to provide the risk stratification model with a smoother binary probability boundary.

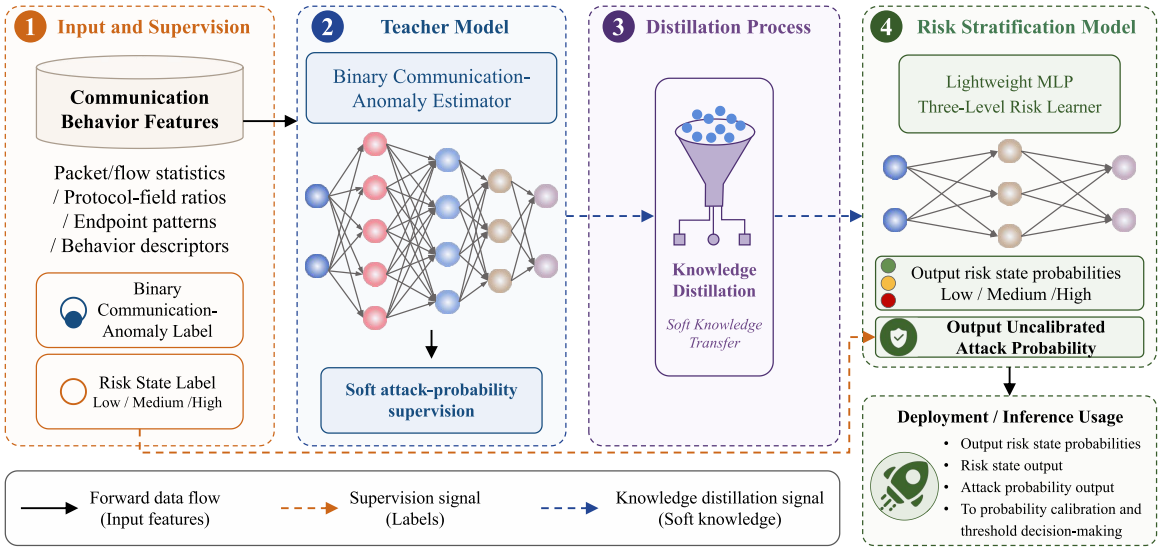

Figure 2: Distillation-guided risk-state stratification module of RSTD-KD. The binary communication-anomaly estimator provides soft attack-probability supervision, while the lightweight risk stratification model learns low-, medium-, and high-risk reference states and outputs uncalibrated attack probabilities for subsequent calibration and threshold decision-making.

The lightweight risk stratification model uses a multi-layer perceptron (MLP) structure and outputs a risk state probability vector:

$$\boldsymbol{\pi}_i = \left[\pi_{i,\mathrm{Low}}, \pi_{i,\mathrm{Medium}}, \pi_{i,\mathrm{High}}\right], \tag{5}$$

where $\pi_{i,\mathrm{Low}}$, $\pi_{i,\mathrm{Medium}}$, and $\pi_{i,\mathrm{High}}$ denote the probabilities that sample i belongs to the three risk states. Since both Medium and High indicate attack-risk states, the model-level attack probability is defined as

$$p_i^S = \pi_{i,\mathrm{Medium}} + \pi_{i,\mathrm{High}}. \tag{6}$$

The non-distilled variant, denoted as RSTD-KD w/o Distillation in the ablation study, is trained only with reference risk state labels and binary communication-anomaly labels. Its base loss is defined as

$$\mathcal{L}_{base} = \lambda_s \mathcal{L}_{CE}(s_i, \boldsymbol{\pi}_i) + \lambda_y \mathcal{L}_{BCE}(y_i, p_i^S), \tag{7}$$

where $\mathcal{L}_{CE}$ denotes the cross-entropy loss for risk state classification, $\mathcal{L}_{BCE}$ denotes the binary cross-entropy loss for binary communication-anomaly detection, and λ_s and λ_y are the weights of the two loss terms.

With distillation, the complete RSTD-KD model further aligns the model-level attack probability with the soft attack probability produced by the Teacher under the joint supervision of reference risk state labels, binary communication-anomaly labels, and soft attack probabilities. The distillation constraint is defined by the mean squared error:

$$\mathcal{L}_{distill} = \mathcal{L}_{MSE}(p_i^T, p_i^S). \tag{8}$$

This distillation term encourages the lightweight risk stratification model to approximate the soft probability boundary provided by the Teacher, enabling more stable attack probability estimation while keeping the model lightweight.

The final loss is defined as

$$\mathcal{L} = \lambda_s \mathcal{L}_{CE}(s_i, \boldsymbol{\pi}_i) + \lambda_y \mathcal{L}_{BCE}(y_i, p_i^S) + \lambda_d \mathcal{L}_{MSE}(p_i^T, p_i^S), \tag{9}$$

where λ_d denotes the weight of the distillation term. This objective allows the complete RSTD-KD model to jointly learn reference risk state labels, binary communication-anomaly detection information, and the soft attack-probability boundary transferred from the Teacher. The resulting model output can therefore be used for risk state prediction, probability calibration, and subsequent threshold decision-making.

The predicted risk state is defined as

$$\hat{s}_i = \arg \max_{k \in \{\text{Low,Medium,High}\}} \pi_{i,k}, \tag{10}$$

where $\hat{s}_i$ denotes the predicted risk state of the i-th window sample, and $\pi_{i,k}$ denotes the corresponding predicted probability for state k.

3.4 Probability Calibration and Task-Cost-Aware Thresholding

In task-oriented UAV communication warning, the model output probability is not only used for binary communication-anomaly detection, but also directly affects alert-threshold selection and subsequent response decisions. However, the probability produced by a learning model may be over-confident or insufficiently calibrated. Therefore, this study calibrates the attack probability produced by the trained RSTD-KD model before threshold search, so that it can be used as a more reliable input for task-cost calculation.

Let p_i^S denote the raw attack probability output by the trained RSTD-KD risk stratification model for the i-th behavior sample. The calibrated attack probability is defined as

$$\tilde{p}_i = C(p_i^S), \tag{11}$$

where $C(\cdot)$ denotes the probability calibration function, and $\tilde{p}_i$ denotes the calibrated attack probability. In this study, the calibration function is learned on the validation set and then fixed for test-set evaluation. The calibrated probability $\tilde{p}_i$ is used as the unified input for subsequent alert-threshold search and task-cost calculation.

Given an alert threshold θ, the binary alert decision is defined as

$$\hat{a}_i(\theta) = \begin{cases} 1, & \tilde{p}_i \geq \theta, \\ 0, & \tilde{p}_i < \theta, \end{cases} \tag{12}$$

where $\hat{a}_i(\theta) = 1$ indicates that an alert for a communication anomaly or attack behavior is triggered, and $\hat{a}_i(\theta) = 0$ indicates that no alert is triggered.

For a given threshold θ, the false positive rate and false negative rate are defined as

$$\text{FPR}(\theta) = \frac{\text{FP}(\theta)}{\text{FP}(\theta) + \text{TN}(\theta)}, \tag{13}$$

$$\text{FNR}(\theta) = \frac{\text{FN}(\theta)}{\text{FN}(\theta) + \text{TP}(\theta)}. \tag{14}$$

Here, $\text{FPR}(\theta)$ measures the proportion of non-anomalous windows incorrectly assigned to the alert class, while $\text{FNR}(\theta)$ measures the proportion of anomalous windows incorrectly assigned to the non-alert class. FP, FN, TP, and TN denote false positive, false negative, true positive, and true negative, respectively.

Let the cross-sea logistics scenario be denoted by $\mathcal{L}$, and the general aviation inspection scenario be denoted by $\mathcal{I}$. For a task scenario $q \in \{\mathcal{L}, \mathcal{I}\}$, the task-cost function is defined as

$$\mathrm{Cost}_q(\theta) = c_{\mathrm{FP}}^{(q)} \cdot \mathrm{FPR}(\theta) + c_{\mathrm{FN}}^{(q)} \cdot \mathrm{FNR}(\theta), \tag{15}$$

where $c_{\mathrm{FP}}^{(q)}$ and $c_{\mathrm{FN}}^{(q)}$ denote the false-positive and false-negative cost weights under task scenario q, respectively.

To avoid arbitrary threshold selection when multiple candidate thresholds achieve the same minimum validation cost, this study defines a task-preference-based tie-breaking rule. Let the minimum validation cost under task scenario q be

$$\mathrm{Cost}_q^{\min} = \min_{\theta \in \Theta} \mathrm{Cost}_q(\theta). \tag{16}$$

The minimum-cost threshold set is defined as

$$\Theta_q^* = \left\{ \theta \in \Theta \mid \mathrm{Cost}_q(\theta) = \mathrm{Cost}_q^{\min} \right\}, \tag{17}$$

where Θ denotes the candidate threshold set. In the experiments, Θ is defined as a discrete threshold grid from 0.01 to 0.99 with a step size of 0.01; the threshold candidates are searched on the validation set, and the final threshold is selected from Θ_q^* according to the task preference.

For the cross-sea logistics scenario, false alarms may cause unnecessary task verification, link-switching assessment, or mission rescheduling. Therefore, this scenario is treated as false-positive-sensitive, and the upper bound of the minimum-cost threshold set is selected:

$$\theta_{\mathcal{L}}^* = \max \Theta_{\mathcal{L}}^*. \tag{18}$$

For the general aviation inspection scenario, missed alarms may delay the discovery and handling of real communication anomalies. Therefore, this scenario is treated as false-negative-sensitive, and the lower bound of the minimum-cost threshold set is selected:

$$\theta_{\mathcal{I}}^* = \min \Theta_{\mathcal{I}}^*. \tag{19}$$

For the balanced setting, where false-positive and false-negative costs are equal, the default decision point is retained when it belongs to the minimum-cost threshold set. Otherwise, the threshold closest to 0.50 is selected:

$$\theta_{\mathcal{B}}^* = \arg\min_{\theta \in \Theta_{\mathcal{B}}^*} |\theta - 0.50|. \tag{20}$$

This rule ensures that threshold selection is determined by validation-set cost minimization and predefined task preferences, rather than by post-hoc manual selection.

After probability calibration and task-cost-aware thresholding, the model output is converted from an attack probability into task-oriented alert decisions. The alert decision is further associated with the predicted risk state through a rule-based response mapping: low-risk windows correspond to continued monitoring, medium-risk windows correspond to link verification or alert confirmation, and high-risk windows correspond to prioritized inspection and mission safety assessment. In operational UAV systems, these warning outputs can be integrated with link switching, return-to-base assessment, or mission restriction mechanisms.

4 Experimental Evaluation

This section evaluates the effectiveness of RSTD-KD from three aspects. First, it examines whether communication behavior sample construction can provide stable inputs for UAV communication anomaly detection and risk-state stratification, including window-size sensitivity, evaluation-protocol stability, attack-type behavior, and external-dataset portability. Second, it evaluates whether distillation-guided risk-state stratification learning and probability calibration can improve the reliability of low-, medium-, and high-risk state recognition, including component ablation and Medium/High risk-state separability. Third, it analyzes whether task-cost-aware threshold decision-making can generate differentiated alert thresholds and task-oriented alert decisions under different task-cost settings, including cost-ratio sensitivity and simulated communication-interference robustness analysis. Accordingly, the experiments are organized around the following research questions (RQs):

RQ1 Can the constructed communication behavior samples support UAV communication anomaly detection and risk-state stratification under different window sizes, evaluation protocols, attack types, and dataset distributions?

RQ2 Can distillation-guided risk stratification improve low-, medium-, and high-risk state recognition, and can probability calibration improve the reliability of attack probabilities for threshold decision-making?

RQ3 Can task-cost-aware threshold decision-making generate differentiated alert thresholds and task-oriented alert decisions under different task costs and simulated deployment disturbances?

To answer these RQs, the experiments analyze ECU-IoFT-based risk-state recognition performance, component ablation, probability calibration reliability, window-size sensitivity, external-dataset portability on UAVIDS-2025, Medium/High risk-state separability, task-cost-aware threshold selection, cost-ratio sensitivity, and robustness under simulated communication interference.

To avoid protocol inconsistency among supplementary analyses, the window-size sensitivity analysis, cost-ratio sensitivity analysis, and simulated communication-interference robustness analysis are conducted under a unified attack-stratified split. The window-size and robustness analyses focus on fixed risk-state inference, whereas the cost-ratio sensitivity analysis uses calibrated attack probabilities and validation-set threshold search to determine task-specific alert thresholds.

4.1 Dataset and Preprocessing

ECU-IoFT is used as the primary benchmark dataset for the main evaluation, while UAVIDS-2025 is used as an additional external benchmark for external-dataset portability analysis. The two datasets differ in data granularity, communication scenario, attack taxonomy, and sample construction strategy, which enables RSTD-KD to be evaluated under both packet-level window-based communication records and flow-level FANET traffic records.

The ECU-IoFT dataset records normal communication and multiple attack behaviors between the UAV and the ground side, providing a suitable benchmark for evaluating UAV communication anomaly detection, risk-state construction, and task-oriented alert decision-making. Since ECU-IoFT is organized as fine-grained packet-level communication records, consecutive records are aggregated into window-level communication behavior samples to characterize local communication patterns within continuous communication segments. Based on the original communication labels, binary communication-anomaly labels, detailed attack-category labels, and low-, medium-, and high-risk reference-state labels are generated. The constructed ECU-IoFT dataset contains 1702 window-level communication behavior samples, including 996 Normal samples and 706 Attack samples. No Attack is mapped to the Low-risk state, Wi-Fi Deauthentication

Attack is mapped to the Medium-risk state, and WPA2-PSK Wi-Fi Cracking Attack and Tello API Exploit are mapped to the High-risk state according to their potential impact on communication continuity, authentication security, and control-interface security. This construction provides the empirical basis for evaluating window-level risk-state recognition, probability calibration, and task-cost-aware threshold decision-making under the ECU-IoFT benchmark setting.

UAVIDS-2025 is further adopted as a flow-level FANET intrusion detection dataset generated through NS-3 simulation. It contains 122,171 labeled flow records from five traffic categories: Normal Traffic, Blackhole Attack, Wormhole Attack, Sybil Attack, and Flooding Attack. Since UAVIDS-2025 is already organized as flow-level records, each flow is treated as one communication behavior sample after feature encoding and normalization, and the packet-level sliding-window construction used for ECU-IoFT is not applied. For UAVIDS-2025, the same impact-oriented three-level reference risk-state principle is used. Normal Traffic is mapped to the Low-risk state. Blackhole Attack is mapped to the Medium-risk state because it mainly causes localized routing disruption through packet dropping. Wormhole, Sybil, and Flooding attacks are mapped to the High-risk state because they may cause routing-integrity damage, identity manipulation, or severe service degradation in FANET communication.

The label distributions and reference risk-state mappings of the two datasets are summarized in Fig. 3. Panel (a) shows the constructed ECU-IoFT window-level samples, including binary communication-anomaly labels, detailed attack categories, mapped risk states, sample counts, and sample proportions. Panel (b) shows the UAVIDS-2025 flow-level traffic categories and their corresponding reference risk states.

(a) ECU-IoFT window-level label distribution and risk-state mapping

Binary label	Detailed category	Risk state	Count	Sample distribution
Normal 996 (58.52%)	Normal traffic	Low	996	58.52%
Attack 706 (41.48%)	WPA2-PSK Wi-Fi Cracking	High	545	32.02%
	Wi-Fi Deauthentication	Medium	155	9.11%
	Tello API Exploit	High	6	0.35%

(b) UAVIDS-2025 flow-level label distribution and risk-state mapping

Binary label	Detailed category	Risk state	Count	Sample distribution
Normal 26,172 (21.42%)	Normal traffic	Low	26,172	21.42%
Attack 95,999 (78.58%)	Blackhole Attack	Medium	26,110	21.37%
	Wormhole Attack	High	26,086	21.35%
	Sybil Attack	High	24,077	19.71%
	Flooding Attack	High	19,726	16.15%

Figure 3: Label distribution and reference risk-state mapping of the two datasets used in this study. Panel (**a**) reports the constructed ECU-IoFT window-level communication behavior samples. Panel (**b**) reports the UAVIDS-2025 flow-level traffic categories used for external-dataset portability analysis.

The ECU-IoFT data split follows a label-block-based chronological split strategy. This strategy preserves temporal order and maintains attack-type coverage across the training, validation, and test sets, making it suitable for a comprehensive benchmark-based evaluation of the proposed risk-state stratification and threshold decision-making process under the public ECU-IoFT setting. All ECU-IoFT-based compared models are trained and evaluated under the same window-level features, reference risk-state labels, data split, and evaluation metrics, ensuring a consistent experimental setting for performance comparison.

For UAVIDS-2025, a flow-level stratified evaluation protocol is adopted because the dataset does not contain packet-level temporal windows. In the repeated random-seed evaluation, flow records are split into training, validation, and test subsets in a stratified manner, so that the distributions of the original traffic categories and mapped risk states are preserved across different subsets. The validation subset is used for probability calibration and model-selection-related procedures, while the test subset is used only for final performance reporting. This process is repeated under five random seeds, and the mean and standard deviation are reported. In addition, a stratified 5-fold cross-validation experiment is conducted as a complementary stability check, where each fold preserves the class distribution as much as possible and is used once as the test fold.

The label-block-based chronological split is used as the main ECU-IoFT benchmark protocol for the overall performance comparison. Additional ECU-IoFT analyses that examine specific factors, such as window-size sensitivity, task-cost ratio sensitivity, and simulated communication-interference robustness, are reported separately in the corresponding experimental subsections. The UAVIDS-2025 experiment uses the flow-level stratified protocol described above, because its data granularity, attack taxonomy, and feature interface are different from those of ECU-IoFT.

4.2 Baselines

For the main ECU-IoFT benchmark comparison, all external baselines are evaluated under the same window-level communication behavior features, data split, and evaluation metrics. The compared baseline models are summarized in Table 1. DG-RF and DG-XGB are adapted from the machine-learning IDS paradigm of DroneGuard [36], and are implemented as Random Forest and XGBoost baselines, respectively. LGBM-IDS follows the LightGBM-based intrusion detection method [37], representing an efficient boosting-tree baseline for structured communication behavior features. CP-MLP is inspired by cyber-physical UAV IDS studies [12], and is implemented as a lightweight neural baseline using window-level communication behavior features. The baseline set also covers advanced UAV/drone security modeling paradigms. GNN-IDS is adapted from topology-aware UAV swarm intrusion detection studies [28]. CLCAN-IDS is adapted from convolutional-attention-based drone-network intrusion detection studies [29]. TransReSE-IDS is adapted from Transformer-based UAV swarm security situation assessment studies [18]. All external baselines are re-implemented as three-class risk-state classifiers under the same window-level features, reference risk-state labels, data split, and evaluation metrics. RSTD-KD denotes the proposed method and is evaluated as the final risk stratification and threshold decision-making model in the following experiments.

These baselines cover multiple representative UAV communication security modeling paradigms. DG-RF represents nonlinear bagging-based ensemble learning. DG-XGB and LGBM-IDS represent boosting-tree-based IDS models for structured communication behavior features. CP-MLP represents a lightweight neural classification baseline. GNN-IDS represents graph-based dependency modeling by constructing a feature-similarity graph over communication windows. CLCAN-IDS represents convolutional-attention-based feature extraction and fusion for drone communication intrusion detection. TransReSE-IDS represents Transformer-based dependency modeling and multi-scale deep feature extraction

for UAV security situation assessment. For fairness, all external baselines in the ECU-IoFT benchmark use the same input features and risk-state labels. Original dataset-specific preprocessing procedures, topology definitions, and external threat-level calculation rules are not directly transferred; instead, each baseline is adapted to the unified ECU-IoFT risk-state classification setting. RSTD-KD further introduces soft attack-probability supervision, probability calibration, and task-cost-aware threshold decision-making, which distinguishes it from the external classification baselines.

Table 1: External baseline models used for ECU-IoFT UAV communication risk-state classification comparison under the same window-level features, data split, and evaluation metrics.

Model	Description
DG-RF [36]	Random Forest baseline adapted from the DroneGuard machine-learning IDS paradigm for drone networks.
DG-XGB [36]	XGBoost baseline adapted from the DroneGuard framework for UAV intrusion detection.
LGBM-IDS [37]	LightGBM-based intrusion detection baseline using structured communication behavior features.
CP-MLP [12]	Lightweight neural baseline implemented using window-level communication behavior features.
GNN-IDS [28]	Graph neural network baseline adapted from recent topology-aware UAV swarm intrusion detection studies.
CLCAN-IDS [29]	Convolutional-attention baseline adapted from recent drone-network intrusion detection studies.
TransReSE-IDS [18]	Transformer-based UAV security situation assessment baseline adapted for communication risk-state classification.

Because federated-learning-based UAV IDS methods require additional distributed-training protocol settings, such as client partitioning, non-IID data construction, communication rounds, aggregation strategies, and communication-overhead evaluation [14,30], they are not used as main centralized baselines under the present ECU-IoFT window-level risk-state classification protocol.

4.3 Evaluation Metrics

The evaluation metrics are divided into several groups according to the experimental objectives. For the overall risk-state classification comparison, we report Accuracy, Balanced Accuracy, Macro-F1, and High-state Recall. Accuracy measures the overall prediction correctness. Balanced Accuracy and Macro-F1 are used to reduce the influence of class imbalance on the overall evaluation. High-state Recall measures the capability of a model to correctly identify high-risk communication windows, which is particularly important for UAV mission safety because missed high-risk states may lead to delayed alerts.

High-state Recall is computed as the recall of the High-risk reference state, i.e., the proportion of true High-risk windows that are correctly predicted as High. Formally, it is defined as

$$\text{High-state Recall} = \frac{\text{TP}_{\text{High}}}{\text{TP}_{\text{High}} + \text{FN}_{\text{High}}} \tag{21}$$

where TP_{High} denotes the number of correctly identified High-risk windows, and FN_{High} denotes the number of High-risk windows incorrectly assigned to other risk states.

For the external-dataset portability analysis on UAVIDS-2025, Accuracy, Balanced Accuracy, and Macro-F1 are also reported under multiple random seeds and 5-fold cross-validation. Risk-level F1 scores and attack-type detection rates are additionally reported to examine whether the framework remains effective under a different UAV communication data granularity and attack taxonomy. For probability reliability evaluation, we report Brier Score. Brier Score evaluates the reliability of attack-probability estimates, where a lower Brier Score indicates better probability calibration.

For task-cost-aware threshold decision-making, we use false positive rate (FPR), false negative rate (FNR), Attack Recall, and the task-cost function Cost to analyze alert performance under different task-cost settings. Attack Recall measures the proportion of attack-related communication windows that are correctly assigned to the alert class. For cost-ratio sensitivity analysis, we further report the selected threshold, FPR, FNR, and expected cost under different false-positive and false-negative cost ratios. Probability calibration methods and alert thresholds are determined on the validation set and then fixed for test-set evaluation to avoid information leakage from the test set.

For Medium/High risk-state separability analysis, we report Medium-to-High and High-to-Medium cross-confusion counts, Direct M/H Sep. Ratio, probability margin, and prediction entropy. Direct M/H Sep. Ratio is used as a diagnostic statistic to examine whether the distillation objective weakens or preserves the direct separability between Medium- and High-risk states.

For stability and robustness analysis, we report mean and standard deviation over multiple random seeds or cross-validation folds when applicable. In the simulated communication-interference experiments, performance degradation and performance retention are calculated relative to the no-interference setting to quantify robustness under channel fading, electromagnetic noise, packet loss, and packet reordering.

4.4 Risk-State Recognition Performance

The comparative results of different models in terms of Accuracy, Balanced Accuracy, Macro-F1, and High-state Recall are shown in Fig. 4. Overall, DG-RF obtains relatively lower results, indicating that Random Forest can capture part of the nonlinear feature relationships, but it is less effective in representing the three-level risk-state boundaries. DG-XGB and LGBM-IDS clearly outperform DG-RF, which suggests that boosting-tree models are more suitable for modeling nonlinear relationships and feature interactions in window-level communication behavior features. CP-MLP further improves Balanced Accuracy and Macro-F1, showing that a lightweight neural model can learn more stable representations from window-level features.

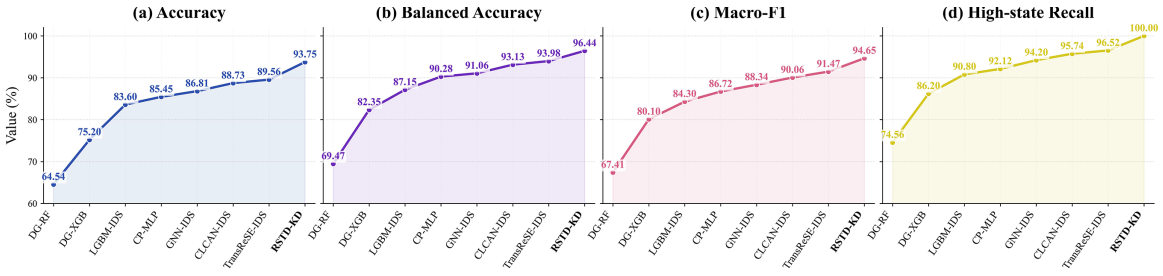

Figure 4: Performance comparison between external baseline models and RSTD-KD for window-level UAV communication risk-state recognition. Panels (**a**–**d**) report Accuracy, Balanced Accuracy, Macro-F1, and High-state Recall, respectively.

Among the external baselines, GNN-IDS, CLCAN-IDS, and TransReSE-IDS represent graph-based, convolutional-attention, and Transformer-based UAV/drone security modeling paradigms, respectively. GNN-IDS achieves 86.81% Accuracy, 91.06% Balanced Accuracy, 88.34% Macro-F1, and 94.20% High-state Recall, indicating that graph-based dependency modeling can provide useful relational information for communication risk-state recognition. CLCAN-IDS obtains 88.73% Accuracy, 93.13% Balanced Accuracy, 90.06% Macro-F1, and 95.74% High-state Recall, showing the effectiveness of convolutional-attention feature extraction for drone communication behavior modeling. TransReSE-IDS achieves the strongest performance among the external baselines, with 89.56% Accuracy, 93.98% Balanced Accuracy, 91.47% Macro-F1, and 96.52% High-state Recall, suggesting that Transformer-based dependency modeling and multi-scale feature extraction provide a competitive representation for UAV communication risk-state classification.

Among all compared methods, RSTD-KD achieves the best performance on all four metrics, with 93.75% Accuracy, 96.44% Balanced Accuracy, 94.65% Macro-F1, and 100.00% High-state Recall. Compared with TransReSE-IDS, which is the strongest external baseline in this experiment, RSTD-KD improves Accuracy, Balanced Accuracy, Macro-F1, and High-state Recall by 4.19, 2.46, 3.18, and 3.48 percentage points, respectively. These results show that the advantage of RSTD-KD is not limited to overall accuracy. It is more evident in class-balanced performance and high-risk state capture, which is consistent with the objective of risk stratification.

The 100.00% High-state Recall means that all test windows whose reference labels are High are correctly identified as High under the fixed ECU-IoFT test split. In other words, no High-risk reference window is misclassified as Low or Medium in this offline benchmark evaluation. This result is mainly attributed to the distillation-guided risk-state learning objective, which combines supervised High-risk reference labels with binary attack-probability supervision, thereby helping the model preserve high-risk attack evidence during risk-state classification.

It should be noted that this value is a test-set result under the current benchmark protocol, rather than a guarantee of zero missed High-risk states in all real-world UAV deployment scenarios. Therefore, the 100.00% High-state Recall is interpreted as evidence of strong sensitivity to High-risk windows in the ECU-IoFT evaluation, while further field validation is still required for deployment-level conclusions.

4.5 Distillation Ablation and Probability Calibration Reliability

Risk-state recognition depends not only on window-level communication behavior features, but also on whether the model can stably learn the relationship between binary communication-anomaly detection information and the boundaries among low-, medium-, and high-risk reference states. To analyze the effect of distillation-guided risk stratification, this section compares three method settings under the fixed-split evaluation protocol: Binary Probability Baseline, RSTD-KD w/o Distillation, and RSTD-KD. Binary Probability Baseline directly converts attack probabilities into three-level risk-state outputs. RSTD-KD w/o Distillation removes soft attack-probability supervision from the complete framework and is trained only with reference risk-state labels and binary communication-anomaly labels. RSTD-KD denotes the complete proposed method, which jointly uses reference risk-state labels, binary communication-anomaly labels, and soft attack-probability supervision for risk stratification learning.

For ablation comparison, Binary Probability Baseline is evaluated at the risk-state level by converting its binary attack probabilities into Low, Medium, and High states using two validation-set thresholds. Specifically, samples with probabilities below the lower threshold are assigned to Low, samples between the two thresholds are assigned to Medium, and samples above the upper threshold are assigned to High. This conversion is used only to examine whether binary attack probabilities alone can support three-level risk-state recognition; it is not used as the final inference setting of RSTD-KD.

The fixed-split ablation results show clear differences among the three method settings, as illustrated in Fig. 5. Binary Probability Baseline shows substantially lower performance in Macro-F1 and High-state Recall, indicating that binary attack probabilities alone cannot adequately characterize the boundaries among low-, medium-, and high-risk reference states. This result also suggests that directly converting binary attack probabilities into risk-state outputs is insufficient for reliable three-level risk recognition.

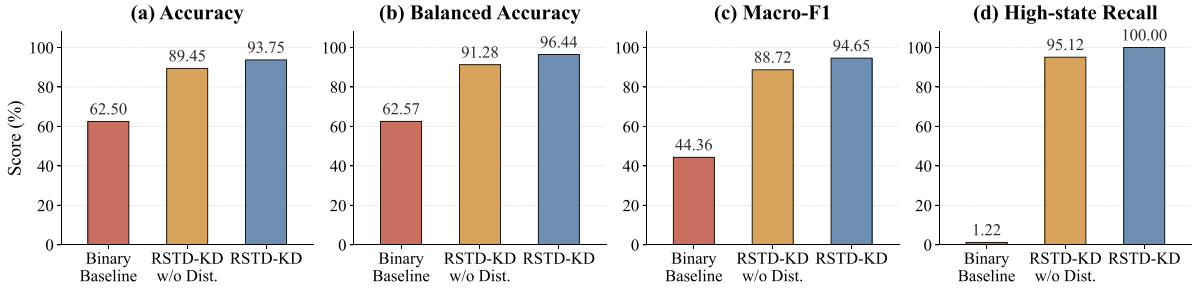

Figure 5: Ablation results of RSTD-KD for window-level UAV communication risk-state recognition under binary probability baseline, RSTD-KD w/o distillation, and complete RSTD-KD.

Compared with RSTD-KD w/o Distillation in the fixed-split evaluation, the complete RSTD-KD improves Accuracy from 89.45% to 93.75%, Balanced Accuracy from 91.28% to 96.44%, Macro-F1 from 88.72% to 94.65%, and High-state Recall from 95.12% to 100.00%. These numerical improvements provide a direct quantitative ablation analysis for the knowledge-distillation component under the main fixed-split benchmark protocol. This indicates that soft attack-probability supervision provides additional boundary information beyond hard reference risk labels and binary communication-anomaly labels. The distillation mechanism therefore transfers binary communication-anomaly detection information into the risk stratification learning process and improves the stability of the three-level risk-state boundaries. Because task-cost-aware thresholding relies on reliable attack-probability estimates, we further evaluate probability calibration using the Brier Score.

If the attack probability is poorly calibrated, subsequent threshold search may lead to unstable alert decisions even when risk-state prediction is accurate. Therefore, Platt Scaling is applied to the attack probabilities produced by Binary Probability Baseline and RSTD-KD, and Brier Score is used to measure probability prediction error.

The probability calibration results show that Platt Scaling substantially reduces the Brier Scores of both Binary Probability Baseline and RSTD-KD, as reported in Table 2. For Binary Probability Baseline, the Brier Score decreases from 0.1872 to 0.0307. For RSTD-KD, the Brier Score decreases from 0.0762 to 0.0113. This result confirms that Platt probability calibration improves the reliability of attack-probability estimates. Its contribution is mainly reflected in probability calibration rather than direct risk-state classification accuracy. Since the calibrated Brier Score of RSTD-KD is lower than that of Binary Probability Baseline, the calibrated attack probabilities produced by RSTD-KD are used as the main input for subsequent task-cost-aware threshold decision-making.

Table 2: Probability calibration results of binary probability baseline and RSTD-KD. Lower Brier Score indicates better probability calibration.

Method	Calibration Method	Raw Brier Score	Calibrated Brier Score
Binary Probability Baseline	Platt Scaling	0.1872	0.0307
RSTD-KD	Platt Scaling	0.0762	0.0113

4.6 Window-Size Sensitivity Analysis

The window length w controls the temporal granularity of window-level communication behavior modeling. A smaller window can generate more samples and preserve short-term communication changes, whereas a larger window can cover longer communication segments but may reduce the number of available samples and smooth short-term attack behaviors. Since the main experiments adopt $w = 32$, we conduct a supplementary window-size sensitivity analysis to examine whether the performance of RSTD-KD strongly depends on this setting. To ensure protocol consistency among supplementary ECU-IoFT analyses, this analysis uses a unified attack-stratified split and fixed risk-state inference protocol. The row with $w = 32$ is used as the reference setting, and ΔAccuracy is computed relative to this default configuration.

The results in Table 3 show that RSTD-KD is not highly sensitive to the window-size setting under the unified evaluation protocol. When w varies from 8 to 64, Accuracy remains within a narrow range from 93.02% to 94.71%, with a maximum difference of 1.69 percentage points. Macro-F1 remains between 93.45% and 95.91%, and High-state Recall remains between 98.21% and 100.00%. Compared with the default setting $w = 32$, the largest Accuracy improvement is only 0.96 percentage points at $w = 8$, while the largest decrease is 0.73 percentage points at $w = 64$. These results indicate that the main performance trend of RSTD-KD is stable across different temporal granularities, rather than being dependent on a single carefully tuned window size.

Table 3: Window-size sensitivity analysis of RSTD-KD under the unified attack-stratified split and fixed risk-state inference protocol.

w	Windows	Test	Accuracy (%)	ΔAccuracy	Macro-F1 (%)	High-State Recall (%)
8	6811	1021	94.71	+0.96	95.91	100.00
16	3405	510	94.12	+0.37	94.91	98.80
24	2270	342	94.15	+0.40	95.48	98.21
32	1702	256	93.75	0.00	94.65	100.00
48	1135	170	94.12	+0.37	94.67	100.00
64	851	129	93.02	−0.73	93.45	100.00

Although $w = 8$ achieves the highest Accuracy and Macro-F1 in this supplementary analysis, the improvement over $w = 32$ is limited. The default setting $w = 32$ is retained because it provides a practical balance between local communication context length, sample availability, class-balanced performance, and high-risk state sensitivity. Larger windows do not provide a clear advantage: $w = 48$ achieves performance close to $w = 32$ but uses fewer window samples and a coarser temporal granularity, whereas $w = 64$ leads to lower Accuracy and Macro-F1. Therefore, increasing the window length beyond $w = 32$ does not bring additional benefit under the current window-level statistical feature design.

We further examine adaptive multi-window aggregation by comparing fixed-window inference with majority voting, confidence-weighted voting, and maximum-confidence selection over $w \in \{8, 16, 24, 32, 48, 64\}$. These aggregation strategies do not improve Accuracy over the fixed $w = 32$ setting: all three obtain 93.75% Accuracy and 100.00% High-state Recall. They only slightly improve Macro-F1 from 94.65% to 95.16%, suggesting that multi-window aggregation brings limited additional benefit in the current setting. Therefore, the fixed $w = 32$ setting is not the primary bottleneck of RSTD-KD, although more advanced data-driven window construction remains a meaningful future direction for dynamic UAV traffic scenarios.

4.7 External-Dataset Portability Analysis

To examine whether RSTD-KD can be adapted beyond the primary ECU-IoFT benchmark, we evaluate it on UAVIDS-2025 using the flow-level stratified protocol described in the dataset preprocessing step. This experiment is positioned as an external-dataset portability evaluation, rather than as a zero-shot transfer experiment or a universal generalization claim across all UAV deployment scenarios.

The results reported in Table 4 show that RSTD-KD maintains stable performance on UAVIDS-2025. Under five random seeds, it achieves 94.92% Accuracy, 93.52% Balanced Accuracy, and 94.38% Macro-F1. Under 5-fold cross-validation, it obtains 94.84% Accuracy, 93.32% Balanced Accuracy, and 94.26% Macro-F1. The small standard deviations, all no larger than 0.25 percentage points, indicate that the results are stable with respect to random initialization and data partitioning on this external UAV communication benchmark.

Table 4: External-dataset portability results of RSTD-KD on UAVIDS-2025 under repeated random-seed evaluation and 5-fold cross-validation.

Evaluation Protocol	Accuracy (%)	Balanced Accuracy (%)	Macro-F1 (%)
5 Random seeds	94.92 ± 0.15	93.52 ± 0.16	94.38 ± 0.17
5-fold CV	94.84 ± 0.16	93.32 ± 0.25	94.26 ± 0.16

The detailed results in Table 5 further show that RSTD-KD maintains effective discrimination across different risk levels and attack categories. Specifically, RSTD-KD achieves 98.17% F1 for Normal traffic, 89.26% F1 for the Medium-risk class corresponding to Blackhole attack, and 95.69% F1 for the High-risk class covering Flooding, Sybil, and Wormhole attacks. At the attack-category level, the average recall rates are 99.90% for Blackhole, 99.97% for Flooding, 99.96% for Sybil, and 98.10% for Wormhole. These results suggest that RSTD-KD can preserve effective risk-state discrimination when the attack types shift from Wi-Fi protocol-layer attacks in ECU-IoFT to FANET routing-layer attacks in UAVIDS-2025.

Overall, the UAVIDS-2025 results provide additional evidence that the main modeling components of RSTD-KD can be adapted to UAV communication data with different granularity and attack taxonomy. Nevertheless, UAVIDS-2025 is generated through network simulation and cannot fully represent all real low-altitude UAV missions, communication frequencies, physical-layer effects, and deployment environments. Therefore, further validation on real-world UAV communication datasets and field-collected flight data remains necessary.

Table 5: Detailed risk-level and attack-category performance of RSTD-KD on UAVIDS-2025 under the repeated random-seed evaluation.

Level	Category	Average Value (%)
Risk-level F1	Normal traffic	98.17
Risk-level F1	Medium-risk class/Blackhole attack	89.26
Risk-level F1	High-risk class/Flooding, Sybil, and Wormhole attacks	95.69
Attack-category Recall	Blackhole Attack	99.90
Attack-category Recall	Flooding Attack	99.97
Attack-category Recall	Sybil Attack	99.96
Attack-category Recall	Wormhole Attack	98.10

4.8 Medium/High Risk-State Separability Analysis

The distillation objective in RSTD-KD aligns the teacher model's binary attack probability with the aggregated student probability $\pi_{i,\text{Medium}} + \pi_{i,\text{High}}$. Since both Medium and High belong to attack-related risk states, a potential concern is that this binary probability constraint may blur the boundary between the two states. To examine this issue, we conduct a pairwise Medium/High separability diagnostic under different KD configurations. The evaluation is restricted to samples whose reference labels are Medium or High, and the direct cross-confusion between these two states is explicitly counted.

Direct M/H Sep. Ratio is defined as the proportion of Medium- and High-risk samples that are not directly confused with each other:

$$\text{Direct M/H Sep. Ratio} = 1 - \frac{N_{\text{Medium}\to\text{High}} + N_{\text{High}\to\text{Medium}}}{N_{\text{Medium}} + N_{\text{High}}}. \tag{22}$$

For each diagnostic sample, the probability margin is defined as

$$\gamma_i = \pi_{i,s_i} - \max_{k\in\{\text{Low,Medium,High}\},\, k\neq s_i} \pi_{i,k}, \tag{23}$$

where s_i is the reference risk-state label. Med Margin and High Margin denote the average γ_i values for Medium- and High-risk samples, respectively. Prediction entropy is computed from the full three-class risk-state probability vector as

$$H_i = -\sum_{k\in\{\text{Low,Medium,High}\}} \pi_{i,k} \log(\pi_{i,k}). \tag{24}$$

As shown in Table 6, the direct Medium/High decision boundary is preserved under all KD configurations. The Direct M/H Sep. Ratio remains 1.000, and no Medium→High or High→Medium cross-confusion is observed. This result indicates that aligning the teacher model's binary attack probability with $\pi_{i,\text{Medium}} + \pi_{i,\text{High}}$ does not collapse Medium and High into a single attack-related class.

The probability-level indicators provide further support for this observation. The Medium and High margins remain above 0.85 in all configurations, indicating that the correct Medium or High class is separated from the closest competing risk state with a clear probability gap. The entropy values remain at a comparable level across KD configurations and do not increase under Full KD compared with No KD, indicating that Full KD does not introduce additional prediction uncertainty near the Medium/High boundary. Therefore,

the binary distillation constraint mainly regularizes the total attack-probability mass, while the supervised risk-state classification loss preserves the discrimination between Medium and High.

Table 6: Diagnostic analysis of Medium/High risk-state separability under different KD configurations. Direct M/H Sep. Ratio measures only the direct separability between Medium and High on a 0–1 scale, and is not equivalent to the full three-class risk-state accuracy.

KD Configuration	Direct M/H Sep. Ratio	Medium→ High	High→ Medium	Med Margin	High Margin	Entropy
Full KD	1.000	0	0	0.852	0.869	0.284
No KD	1.000	0	0	0.869	0.854	0.316
KD ×0.5	1.000	0	0	0.860	0.864	0.292
KD ×2	1.000	0	0	0.861	0.912	0.224
Risk KD Only	1.000	0	0	0.854	0.868	0.285
Bin+Atk KD Only	1.000	0	0	0.900	0.893	0.261

The diagnostic results reported in Table 6 should be interpreted as a pairwise Medium/High boundary analysis rather than as a replacement for the complete three-class risk-state evaluation. In the full evaluation, High-state Recall counts High-risk samples misclassified as either Low or Medium, whereas Direct M/H Sep. Ratio only measures direct cross-confusion between Medium and High. Therefore, this diagnostic analysis provides targeted evidence that the binary distillation constraint does not blur the Medium/High boundary.

4.9 Task-Cost-Aware Threshold Decision-Making Analysis

After obtaining the calibrated attack probabilities produced by RSTD-KD, this section examines how representative task-cost preferences affect alert-threshold selection. Cross-sea logistics and general aviation inspection are adopted to instantiate two representative cost preferences. Cross-sea logistics places greater emphasis on mission continuity and dispatch stability, where false alarms may lead to unnecessary task verification, link-switching assessment, or mission rescheduling. General aviation inspection places greater emphasis on risk discovery, where missed alarms may leave real communication anomalies untreated. Based on these differences, different false-positive and false-negative costs are assigned to the two scenarios to analyze how task preferences reshape the alert boundary.

The Fixed strategy uses the default threshold of 0.50 and recalculates the Test Cost under the corresponding scenario-specific cost weights. The Task-cost-aware strategy searches the validation set according to the cost function and selects the task-preferred alert threshold. Under this setting, the calibrated attack probabilities produced by RSTD-KD form a validation minimum-cost threshold interval of 0.17–0.68, as shown in Fig. 6a. This interval indicates that the calibrated probability outputs provide a stable decision region in which multiple thresholds achieve the same minimum validation cost.

According to the predefined task-preference tie-breaking rule, the cross-sea logistics scenario selects the upper bound of the minimum-cost threshold set because it is false-positive-sensitive. Therefore, the final threshold is set to 0.68. In contrast, the general aviation inspection scenario selects the lower bound of the same minimum-cost threshold interval because it is false-negative-sensitive. Therefore, the final threshold is set to 0.17. This result shows that the same calibrated attack probabilities can support different alert boundaries when task-cost preferences change.

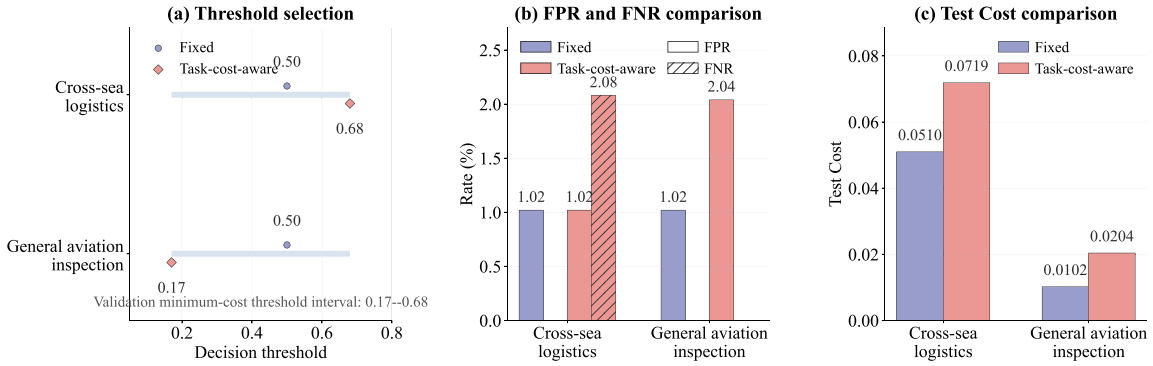

Figure 6: Threshold decision-making results under representative task-cost scenarios. Panel (**a**) shows the selected thresholds and the validation minimum-cost threshold interval. Panel (**b**) reports the FPR–FNR trade-off under the fixed and task-cost-aware strategies. Panel (**c**) compares the test cost under different task-cost scenarios.

The FPR–FNR results in Fig. 6b further illustrate the trade-off introduced by threshold adjustment. In the cross-sea logistics scenario, the higher task-cost-aware threshold keeps the FPR at 1.02%, while the FNR increases to 2.08%. This setting is consistent with a false-positive-sensitive preference, because it avoids expanding the alert scope unnecessarily. In the general aviation inspection scenario, the lower task-cost-aware threshold reduces the FNR to 0.00%, while the FPR increases to 2.04%. This setting is consistent with a false-negative-sensitive preference, because it prioritizes the discovery of attack-related communication windows. Therefore, the threshold adjustment produces interpretable alert behaviors that correspond to different operational priorities.

The Test Cost results in Fig. 6c further confirm that the selected thresholds reflect different task-cost preferences. The Fixed strategy yields low Test Cost values because the default threshold of 0.50 falls inside the validation minimum-cost interval. The task-cost-aware thresholds provide more explicit control over the alert boundary: the cross-sea logistics setting selects a stricter threshold to maintain false-alarm control, whereas the general aviation inspection setting selects a more sensitive threshold and achieves 0.00% FNR. These results indicate that the proposed thresholding module converts calibrated attack probabilities into task-oriented alert decisions by encoding predefined false-positive and false-negative cost preferences, rather than relying on a single fixed threshold for all low-altitude UAV communication tasks.

4.10 Cost-Ratio Sensitivity Analysis

The two representative task scenarios demonstrate that different false-positive and false-negative costs can lead to different alert-threshold choices. However, practical UAV missions may involve more diverse FP/FN cost ratios. To further analyze the sensitivity of the task-cost-aware thresholding module, we vary the cost ratio C_{FP}:C_{FN} from 5:1 to 1:5 using the same calibrated attack probabilities and validation-set threshold search protocol. The validation-set minimum-cost threshold interval remains 0.17–0.68. For false-positive-sensitive settings, the upper bound of this interval is selected; for false-negative-sensitive settings, the lower bound is selected; and for the balanced setting, the default threshold is retained when it lies inside the minimum-cost threshold interval. The test cost is computed as $C_{FP} \cdot \mathrm{FPR} + C_{FN} \cdot \mathrm{FNR}$.

As reported in Table 7, different FP/FN cost preferences lead to different threshold selections within the validation-set minimum-cost threshold interval. When false-positive cost is larger than false-negative cost, such as 5:1 and 2:1, the selected threshold is the upper bound 0.68 of the validation minimum-cost threshold interval. This setting avoids unnecessarily expanding the alert region, while allowing a small FNR of 2.08%

on the test set. When the two costs are balanced, the default threshold 0.50 is retained because it lies inside the validation minimum-cost threshold interval and achieves 1.02% FPR and 0.00% FNR. When false-negative cost is larger, such as 1:2 and 1:5, the selected threshold is the lower bound 0.17, which favors more sensitive alert triggering and produces 0.00% FNR on the current test split.

Table 7: Cost-ratio sensitivity analysis under different FP/FN cost ratios. FPR and FNR are reported in percentages, while Test Cost is computed using decimal rates as $C_{FP} \cdot FPR + C_{FN} \cdot FNR$.

C_{FP}:C_{FN}	Preference	θ^*	FPR (%)	FNR (%)	Test Cost
5:1	FP-sensitive	0.68	1.02	2.08	0.0719
2:1	FP-sensitive	0.68	1.02	2.08	0.0412
1:1	Balanced	0.50	1.02	0.00	0.0102
1:2	FN-sensitive	0.17	2.04	0.00	0.0204
1:5	FN-sensitive	0.17	2.04	0.00	0.0204

The repeated zero FNR values do not mean that RSTD-KD can guarantee zero missed attacks in all deployment environments. They indicate that, under the current ECU-IoFT test split and the selected thresholds, all attack-related windows are assigned to the alert class. Since the FNR term is zero for the balanced and false-negative-sensitive settings, increasing C_{FN} from 2 to 5 does not further change the Test Cost; the cost is then determined only by $C_{FP} \cdot FPR$. For example, under 1:5, the Test Cost remains $1 \times 0.0204 + 5 \times 0 = 0.0204$. Therefore, this analysis shows that the task-cost-aware module changes the threshold according to FP/FN preferences, while the final Test Cost depends jointly on the selected threshold, FPR, FNR, and the cost weights.

Overall, this cost-ratio sensitivity analysis quantitatively examines whether the proposed thresholding mechanism remains interpretable under different FP/FN cost ratios. The results show that different cost preferences select different positions within the same minimum-cost threshold interval: the upper bound for FP-sensitive tasks, the default threshold for the balanced setting, and the lower bound for FN-sensitive tasks.

4.11 Robustness under Simulated Communication Interference

Real UAV communication environments may involve channel attenuation, electromagnetic noise, packet loss, and packet reordering caused by mobility and topology variation. Although the ECU-IoFT benchmark is an offline dataset, these deployment-related disturbances can still be approximated through controlled perturbation of the window-level communication features. This robustness analysis focuses on risk-state recognition performance under perturbed communication features, rather than on recalibrated alert-threshold selection. Therefore, the clean and perturbed samples are evaluated using the same fixed risk-state inference protocol. The clean no-interference baseline Accuracy is 93.75% and is used as the reference for calculating performance retention under different perturbation settings.

Channel fading is simulated by applying amplitude attenuation to communication features. Electromagnetic interference is simulated by adding feature-space noise under different signal-to-noise ratio levels. Packet loss and packet reordering are approximated through feature-space dropout and feature-order perturbation, respectively. These perturbations are not intended to fully reproduce real physical-layer channel dynamics, but to provide a controlled feature-space robustness test for deployment-related communication disturbances. For each interference type, multiple severity levels are tested, and the worst-case result under the extreme setting is reported in Table 8. Accuracy drop is reported as the absolute decrease in percentage

points relative to the clean baseline, and retention is computed as the ratio between perturbed Accuracy and clean baseline Accuracy.

Table 8: Robustness of RSTD-KD under simulated communication-interference conditions. Accuracy drop is reported as the absolute decrease in percentage points relative to the clean no-interference baseline Accuracy of 93.75%. Retention is computed as perturbed Accuracy divided by the clean baseline Accuracy.

Interference Type	Extreme Setting	Worst Accuracy (%)	Accuracy Drop (pp)	Retention
Channel fading	$\alpha = 0.1$	84.84	8.91	90.5%
Electromagnetic noise	SNR = 0 dB	93.75	0.00	100.0%
Packet loss	70%	88.67	5.08	94.6%
Packet reordering	70%	58.98	34.77	62.9%
Average	–	–	–	87.00%

The robustness results reported in Table 8 indicate that RSTD-KD maintains stable performance under electromagnetic noise and packet-loss perturbations in the simulated feature-space setting. Under electromagnetic noise, the Accuracy remains 93.75% even at the strongest tested noise level, corresponding to 100.0% retention. Under 70% packet loss, the Accuracy remains 88.67%, with a retention rate of 94.6%, indicating that the model can tolerate a considerable degree of missing communication evidence.

Channel fading causes a moderate performance drop. When the attenuation coefficient decreases to $\alpha = 0.1$, the Accuracy decreases by 8.91 percentage points to 84.84%, corresponding to 90.50% retention. Packet reordering has the largest impact. When the reordering ratio reaches 70%, the Accuracy decreases by 34.77 percentage points to 58.98%, and the retention rate drops to 62.9%. This result is consistent with the design of window-level communication behavior modeling: severe order perturbation may distort the correspondence between window-level feature patterns and communication behavior states more strongly than additive noise or random feature loss.

Overall, the average retention rate under the four extreme disturbance settings is 87.00%. These results suggest that RSTD-KD has a certain degree of robustness to simulated communication disturbances, especially electromagnetic noise, packet loss, and moderate channel fading. Nevertheless, the degradation under severe packet reordering indicates that packet ordering protection, order-robust feature aggregation, or sequence-consistency checking should be considered in real UAV deployment. This analysis should still be interpreted as controlled simulation-based evidence rather than a substitute for real UAV flight tests with onboard communication devices.

5 Conclusion

This paper proposed RSTD-KD, a knowledge distillation-guided method for UAV communication risk stratification and task-cost-aware threshold decision-making. RSTD-KD converts fine-grained UAV communication records into window-level behavior samples, constructs Low, Medium, and High reference risk states, and transfers binary communication-anomaly boundary information into three-level risk state learning through soft attack-probability supervision. By further combining probability calibration with task-cost-aware threshold selection, RSTD-KD links communication detection outputs with calibrated attack probabilities and task-oriented alert decisions. Experiments on the public ECU-IoFT benchmark show that RSTD-KD achieves superior risk state recognition performance compared with external baseline models and its ablated variant without soft attack-probability supervision. Platt Scaling improves the reliability of attack probabilities for threshold search and task-cost calculation. The threshold decision-making results

further show that the same calibrated probabilities can support different alert boundaries under different task preferences, enabling false-alarm control for cross-sea logistics and missed-alarm reduction for general aviation inspection. Overall, RSTD-KD provides a complete and reproducible modeling process for UAV communication risk stratification and task-oriented alert decision-making. Future research can extend this paradigm to UAV swarm communication scenarios involving inter-UAV links, dynamic topology changes, and coordinated low-altitude missions. Beyond this swarm-oriented extension, future work may further validate RSTD-KD on real UAV platforms with onboard communication devices to examine real-time inference, communication stability, and deployment robustness. More heterogeneous UAV communication datasets involving different platforms, frequencies, flight environments, and attack types can also be used to evaluate the portability of the proposed risk-state mapping strategy. In addition, adaptive window construction and dynamic task-cost estimation may be explored to better support UAV missions with changing traffic patterns and mission priorities. These directions will further extend RSTD-KD from benchmark-based risk warning toward practical low-altitude UAV communication security applications.

Acknowledgement: Not applicable.

Funding Statement: This work was supported in part by the Major Science and Technology Special Project of Sichuan Province under Grant 2024ZDZX0044, the Regional Innovation Cooperation Project of Sichuan Province under Grant 2026YFHZ0121, the Education and Teaching Reform Research Project of Hainan Higher Education Institutions under Grant Hnjg2026ZC-29, the High-level Talent Research Start-up Fund of Hainan Normal University under Grant HSZK-KYQD-202610, and the Graduate Innovation Research Project of Hainan Province under Grant Hys2025-441. The funding was acquired by Tingting Lu, Caiyun Li, and Binhui Tang.

Author Contributions: The authors confirm contribution to the paper as follows: Conceptualization, Caiyun Li and Xiaowen Liu; methodology, Caiyun Li, Xiaowen Liu and Binhui Tang; software, Caiyun Li; validation, Caiyun Li, Xiaowen Liu and Tingting Lu; formal analysis, Caiyun Li and Xiaowen Liu; investigation, Caiyun Li and Tingting Lu; resources, Tingting Lu, Caiyun Li and Binhui Tang; data curation, Caiyun Li; writing—original draft preparation, Caiyun Li; writing—review and editing, Xiaowen Liu, Tingting Lu, Daibo Xiao and Li Chen; visualization, Caiyun Li; supervision, Binhui Tang; project administration, Binhui Tang; funding acquisition, Tingting Lu, Caiyun Li and Binhui Tang. Caiyun Li and Xiaowen Liu contributed equally to this work. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: This study used publicly available data from the ECU-IoFT dataset at https://github.com/iMohi/ECU-IoFT and the UAVIDS-2025 dataset at https://zenodo.org/records/15336998. The processed window-level samples and flow-level samples, trained-model configurations, and source scripts required to reproduce the reported results are publicly available at https://github.com/KK0567/RSTD-KD.git.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Fattori F, Cocuzza S. Tethered drones: a comprehensive review of technologies, challenges, and applications. Drones. 2025;9(6):425. doi:10.3390/drones9060425.
2. Naser HN, Hashim HA, Ahmadi M. Aerial assistive payload transportation using quadrotor UAVs with nonsingular fast terminal SMC for human physical interaction. Results Eng. 2025;25(5):103701. doi:10.1016/j.rineng.2024.103701.
3. McEnroe P, Wang S, Liyanage M. A survey on the convergence of edge computing and AI for UAVs: opportunities and challenges. IEEE Internet Things J. 2022;9(17):15435–59.

4. Han X, Lin B, Na Z, Li B, Zhang C, Zhang R. Spatial crowdsourcing-based task allocation for UAV-assisted maritime data collection. IEEE Trans Veh Technol. 2024;74(2):3375–88. doi:10.1109/tvt.2024.3483890.
5. Xiao J, Zhang R, Zhang Y, Feroskhan M. Vision-based learning for drones: a survey. IEEE Trans Neural Netw Learn Syst. 2025;36(9):15601–21. doi:10.1109/tnnls.2025.3564184.
6. Wang Q, Li X, Bhatia S, Liu Y, Alex LT, Khowaja SA, et al. UAV-enabled non-orthogonal multiple access networks for ground-air-ground communications. IEEE Trans Green Commun Netw. 2022;6(3):1340–54.
7. Qu C, Sorbelli FB, Singh R, Calyam P, Das SK. Environmentally-aware and energy-efficient multi-drone coordination and networking for disaster response. IEEE Trans Netw Serv Manag. 2023;20(2):1093–109. doi:10.1109/tnsm.2023.3243543.
8. Ceviz O, Sen S, Sadioglu P. A survey of security in UAVs and FANETs: issues, threats, analysis of attacks, and solutions. IEEE Commun Surv Tutor. 2025;27(5):3227–65. doi:10.1109/comst.2024.3515051.
9. Cha J, Jang J, Shin D, Shin D. Cost-sensitive threshold optimization for network intrusion detection: a per-class approach with XGBoost. Electronics. 2026;15(7):1542. doi:10.3390/electronics15071542.
10. Ahmad T, Hadi MU, Vassiliou V, Dimitriou L, Anwar A, Tran TA. Real-time anomaly detection in smart vehicle-to-UAV networks for disaster management. Trans Emerg Telecommun Technol. 2025;36(5):e70162. doi:10.1002/ett.70162.
11. Kurunathan H, Huang H, Li K, Ni W, Hossain E. Machine learning-aided operations and communications of unmanned aerial vehicles: a contemporary survey. IEEE Commun Surv Tutor. 2023;26(1):496–533. doi:10.1109/comst.2023.3312221.
12. Hassler SC, Mughal UA, Ismail M. Cyber-physical intrusion detection system for unmanned aerial vehicles. IEEE Trans Intell Transp Syst. 2023;25(6):6106–17. doi:10.1109/tits.2023.3339728.
13. Hadi HJ, Cao Y, Li S, Hu Y, Wang J, Wang S. Real-time collaborative intrusion detection system in UAV networks using deep learning. IEEE Internet Things J. 2024;11(20):33371–91. doi:10.1109/jiot.2024.3426511.
14. He X, Chen Q, Tang L, Wang W, Liu T, Li L, et al. Federated continuous learning based on stacked broad learning system assisted by digital twin networks: an incremental learning approach for intrusion detection in UAV networks. IEEE Internet Things J. 2023;10(22):19825–38. doi:10.1109/jiot.2023.3282648.
15. Do Yoo J, Kim H, Kim HK. GUIDE: gAN-based UAV IDS enhancement. Comput Secur. 2024;147:104073.
16. Hutchins C, Aniello L, Gerding E, Halak B. A flying ad-hoc network dataset for early time series classification of grey hole attacks. Sci Data. 2025;12(1):1431. doi:10.1038/s41597-025-05560-1.
17. Layman L, Roden W. A controlled experiment on the impact of intrusion detection false alarm rate on analyst performance. Proc Hum Factors Ergon Soc Annu Meet. 2023;67(1):220–5. doi:10.1177/21695067231192573.
18. Zhao D, Shen P, Han X, Zeng S. Security situation assessment in UAV swarm networks using TransReSE: a Transformer-ResNeXt-SE based approach. Veh Commun. 2024;50(16):100842. doi:10.1016/j.vehcom.2024.100842.
19. Tsao KY, Girdler T, Vassilakis VG. A survey of cyber security threats and solutions for UAV communications and flying ad-hoc networks. Ad Hoc Netw. 2022;133(3):102894. doi:10.1016/j.adhoc.2022.102894.
20. Mekdad Y, Aris A, Babun L, El Fergougui A, Conti M, Lazzeretti R, et al. A survey on security and privacy issues of UAVs. Comput Netw. 2023;224(4):109626. doi:10.1016/j.comnet.2023.109626.
21. Hadi HJ, Cao Y, Nisa KU, Jamil AM, Ni Q. A comprehensive survey on security, privacy issues and emerging defence technologies for UAVs. J Netw Comput Appl. 2023;213(1):103607. doi:10.1016/j.jnca.2023.103607.
22. Sen MA, Al-Rubaye S, Tsourdos A. Securing UAV flying ad hoc wireless networks: authentication development for robust communications. Sensors. 2025;25(4):1194.
23. Sarkar S, Shafaei S, Jones TS, Totaro MW. Secure communication in drone networks: a comprehensive survey of lightweight encryption and key management techniques. Drones. 2025;9(8):583.
24. Mohammed AB, Fourati LC. Investigation on datasets toward intelligent intrusion detection systems for intra- and inter-UAVs communication systems. Comput Secur. 2025;150(1):104215. doi:10.1016/j.cose.2024.104215.
25. Miao S, Pan Q, Zheng D, Mohi-ud-din G. Unmanned aerial vehicle intrusion detection: deep-meta-heuristic system. Veh Commun. 2024;46(17):100726. doi:10.1016/j.vehcom.2024.100726.
26. Sakibul IMD, Sharif MA, Sheltami TR. AI-enhanced intrusion detection for UAV systems: a taxonomy and comparative review. Drones. 2025;9(10):682. doi:10.3390/drones9100682.

27. Lin L, Ge H, Zhou Y, Shangguan R. UAV airborne network intrusion detection method based on improved stratified sampling and ensemble learning. Drones. 2025;9(9):604. doi:10.3390/drones9090604.
28. Mughal UA, Atat R, Ismail M. Graph neural network-based intrusion detection system for a swarm of UAVs. In: Proceedings of the MILCOM 2024—2024 IEEE Military Communications Conference (MILCOM); 2024 Oct 28–Nov 1; Washington, DC, USA. New York, NY, USA: IEEE; 2024. p. 578–83.
29. Aldossary M, Alzamil I, Almutairi J. Enhanced intrusion detection in drone networks: a cross-layer convolutional attention approach for drone-to-drone and drone-to-base-station communications. Drones. 2025;9(1):46. doi:10.3390/drones9010046.
30. Ceviz O, Sen S, Sadioglu P, Vassilakis VG. A novel federated learning-based IDS for enhancing UAVs privacy and security. Internet Things. 2025;31(3):101592. doi:10.1016/j.iot.2025.101592.
31. Miao S, Pan Q. Risk assessment of UAV cyber range based on Bayesian–Nash equilibrium. Drones. 2024;8(10):556. doi:10.3390/drones8100556.
32. Sun X, Hu Y, Qin Y, Zhang Y. Risk assessment of unmanned aerial vehicle accidents based on data-driven Bayesian networks. Reliab Eng Syst Saf. 2024;248:110185.
33. Silva Filho T, Song H, Perello-Nieto M, Santos-Rodriguez R, Kull M, Flach P. Classifier calibration: a survey on how to assess and improve predicted class probabilities. Mach Learn. 2023;112(9):3211–60.
34. Leevy JL, Johnson JM, Hancock J, Khoshgoftaar TM. Threshold optimization and random undersampling for imbalanced credit card data. J Big Data. 2023;10(1):58. doi:10.1186/s40537-023-00738-z.
35. Gutiérrez-López A, Gonzalez-Serrano FJ, Figueiras-Vidal AR. Optimum Bayesian thresholds for rebalanced classification problems using class-switching ensembles. Pattern Recogn. 2023;135(3):109158. doi:10.1016/j.patcog.2022.109158.
36. Ihekoronye VU, Ajakwe SO, Lee JM, Kim DS. DroneGuard: an explainable and efficient machine learning framework for intrusion detection in drone networks. IEEE Internet Things J. 2024;12(7):7708–22. doi:10.1109/jiot.2024.3519633.
37. Liu J, Gao Y, Hu F. A fast network intrusion detection system using adaptive synthetic oversampling and LightGBM. Comput Secur. 2021;106(99):102289. doi:10.1016/j.cose.2021.102289.