Explaining Anomalous Events in Flight Data of UAV With Deep Attention-Based Multi-Instance Learning

Jie Yang, Diyin Tang, *Member, IEEE*, Jinsong Yu, *Member, IEEE*, Jian Zhang, and Haigang Liu

***Abstract*—Ensuring the safety and reliability of unmanned aerial vehicles (UAVs) has become a critical issue as they continue to advance. Analyzing anomalous events using event-based explanations is an effective approach to identifying key anomalous behaviors and mitigating potential risks. However, this task is challenging because anomalous events in flight data lack time-step labels in real-world UAV flight scenarios. To address this challenge, we propose a dual attention-based multi-instance learning (DA-DI-MIL) for pinpointing anomaly instances and automatically explaining anomalous events. Our MIL-based framework, from a weakly supervised learning perspective, treats a segment of flight data as a 'bag' with an available label, and its time steps as 'instances' without labels. The dual attention mechanism in DA-DI-MIL combines a temporal pseudo-label, predicted by temporal attention, and sensor variable importance obtained from delta feature attention to better pinpoint anomalies from instance-level labels. We use these labels in our proposed MIL-based framework to establish the relationships between anomalous events and anomalous behaviors. We conducted extensive experiments on real UAV flight data with engine failures to demonstrate the effectiveness and robustness of our proposed method compared to existing state-of-the-art methods. DA-DI-MIL achieves near 90% in evaluation metrics with less than 5 time steps' delay of anomalous event detection and around 85% in temporal anomaly detection. Additionally, we present an illustration of both global and local interpretations in the time domain and feature space, providing comprehensive insights into anomalous events. Our proposed method has efficiently and automatically explained anomalous events in UAV flight data, contributing to high-level aviation safety and reliability.**

***Index Terms*—Anomaly detection, anomalous events, attention mechanism, aviation safety, deep learning, multiple instance learning, unmanned aerial vehicle.**

I. Introduction

UNMANNED aerial vehicle (UAV) has achieved a wide application in civilian, military, and commercial fields, and mastered many complicated tasks, like agriculture and forestry, aerial shots in photography and cinematography, and rescue and traffic monitoring. However, without a pilot's real-time control, it is challenging to ensure the flight safety and reliability of a UAV which is a complex cyber-physical system with sophisticated structures. Anomaly detection of multivariate time series provides an efficient way to monitor the states of UAVs and automatically discover potential high-risk anomalies and faults based on flight data from various sensors. With the rapid improvement of intelligent algorithms and easy access to flight data from advanced sensor acquisition systems, anomalies can be detected by extracting useful information from various sensor readings of flight data [1], [2].

Research on anomaly detection of UAVs has primarily focused on unsupervised learning methods that rely on using intrinsic information from nominal data. Guo et al. [3] optimized a one-class support vector machine (OC-SVM) without negative samples to detect UAV sensor faults. Ji and Lee [4] used OC-SVM to detect anomalies for continuous value signals of hybrid electric vehicles. Zhang et al. [5] proposed a method that involves using probability density-based descriptions with dimension reduction techniques for each sub-window to identify anomalies. Similarly, other works use classical and popular methods, like Local Outlier Factor (LOF) [6], and Isolation Forest (iForest) [7], to conduct anomaly detection of UAV flight data. Deep structure methods have become state of the art in anomaly detection due to their capacity and flexibility in representing complex data, including high-dimensional, temporal, and spatial data [8], compared to the shallow structure methods mentioned earlier. Hundman et al. [9] used Long Shor-Term Memory (LSTM) with a nonparametric anomaly thresholding approach to deal with anomaly detection of multivariate time series data. Wang et al. [10] designed LSTM multivariate regression model, together with a residual filter, to extract spatial-temporal features of UAV flight data for anomaly detection and recovery. Zhong et al. [11] introduced Spatio-temporal correlations into an LSTM model to better express the complex Spatio-temporal relationship of high-dimensional flight data. The LSTM model is an efficient way to capture the temporal and spatial information of multivariate time series of flight data to enhance the performance of anomaly detection accuracy of UAVs.

However, most existing unsupervised learning-based anomaly detection methods often focus solely on detecting anomalies without providing detailed explanations of anomalous events. In the context of anomaly detection for multivariate time series, it is crucial not only to detect anomalies but also to provide explanations for the detected anomalous events in terms of the

Manuscript received 22 August 2022; revised 10 April 2023 and 31 July 2023; accepted 31 July 2023. Date of publication 3 August 2023; date of current version 17 January 2024. This work was supported in part by the National Key R&D Program of China under Grant No. 2022YFB3304600. The review of this article was coordinated by Prof. Ying-Dar Lin. *(Corresponding author: Jinsong Yu.)*

Jie Yang, Diyin Tang, and Jinsong Yu are with the School of Automation Science and Electrical Engineering, Beihang University, Beijing 100191, China (e-mail: megan_yj@buaa.edu.cn; tangdiyin@buaa.edu.cn; yujs@buaa.edu.cn).

Jian Zhang is with the Laboratory of Big Data Decision making for Green Development, Beijing 100192, China (e-mail: zhangjian@bistu.edu.cn).

Haigang Liu is with the Shenyang Aircraft Design and Research Institute, Shenyang 110035, China (e-mail: liuhaigang601@163.com).

Digital Object Identifier 10.1109/TVT.2023.3301678

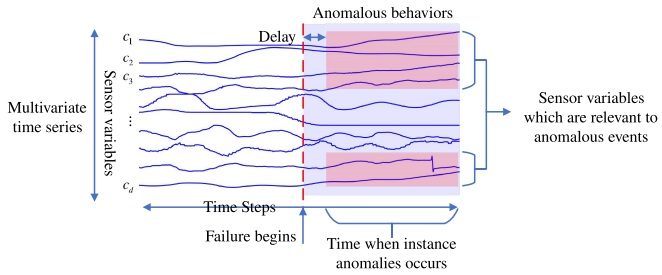

Fig. 1. Visualization of anomalous events and relevant behaviors.

most relevant sensor variables with their consecutive temporal anomalous segments. An illustrative example is presented in Fig. 1. As seen from Fig. 1, given multivariate time series, an anomalous event occurs, like a UAV engine failure, a group of sensor signals would present anomalous behaviors, such as airspeed abnormally increasing, the difference between commanded and measured altitude and pitch widening, etc. Should these anomalous behaviors be identified, the anomalous event would be found immediately the next time it occurs. Such explanations can offer valuable insights into the underlying anomalous behaviors or patterns and enable better understanding, localization, and prediction of anomalous events in future operations.

To achieve the explanation of the anomalous events, Deng and Hooi [12] used a learned graph that represents the relationship of sensors to localize and understand an anomaly. Liang et al. [13] utilized the reconstruction metrics deviation in the original space to locate anomalous dimensions. Although these works based on unsupervised learning methods provided more information for explaining anomalies, they focused on anomalies at the sensor level instead of establishing explicit relationships between anomalous events and the behaviors of relevant sensor variables. From an unsupervised learning perspective, it cannot offer a relationship between the anomalous behaviors from sensor variables and the specific anomalous event. As a result, they may lack the ability to identify and differentiate between different types of anomalous events, limiting their usefulness for future operation and anomaly prediction.

Supervised methods can establish relationships effectively, but their requirement for expensive time-step labels makes them impractical in real-world UAV flight scenarios. For complex UAV systems, defining time-step labels for anomalous events is challenging, but labeling for the occurrence of anomalous events or faults in a flight could be relatively more accessible based on domain-based rules or human annotators. In this context, the problem of anomaly detection with an explanation of anomalous events using limited labeled flight data can be transformed into a specific weakly supervised problem known as inexact supervision [14]. The inexact supervision scenario aligns well with our situation, where instance-level labels are not available, but segment-level labels exist. Given this setting, the MIL (Multi-Instance Learning) framework [15] emerges as a popular method for addressing inexact supervision, providing a natural fit to handle segment-level labels and make inferences at the instance level. MIL assumes that a set of instances is grouped into a bag with a label, but each instance inside the bag is unlabeled. This approach effectively identifies key instances to explain predictions and decision functions [16] and has been widely used in diverse areas, such as computer vision and document classification [17]. Furthermore, when addressing anomaly detection under a specific weakly supervised situation, the main challenge is to differentiate between anomaly time points and nominal time points without the availability of time-step labels.

In current related research, only a few related contributions in event-level detection can be found in the literature for anomaly detection of multivariate time series. The MIL-based method was first used in time-series anomaly detection by Janakiraman et al. [18]. They proposed deep temporal multi-instance learning (DT-MIL) to capture the temporal behavior of time series to identify correlated sensors with adverse events. However, the DT-MIL could not find related anomalous behaviors from sensor variables automatically. To find explanations for events, Laine et al. [19] combined the MIL framework with Multi-Head Convolutional Neural Network-Recurrent Neural Network (MHCNN-RNN) architecture to identify the sensors variables to explain the occurrence of anomalous events. But this architecture needed to build individual sub-models for each sensor to complete the task of identifying relevant sensors, which was quite time-consuming. Moreover, it caused inconsistent evaluation between sensors and lost the correlations between sensors, which affects the performance of anomaly detection. A summary of recent research on UAVs' anomaly detection can be viewed in Table I.

This article proposes a deep attention-based multi-instance learning approach (DA-DI-MIL) to detect anomalies in UAV flight data and also identify key anomalous behaviors from relevant sensor variables as event-based explanations. The anomaly detection problem is formulated under a MIL-based framework to overcome the challenge of a lack of time-step labels for anomalous events in flight data. In contrast to existing works, we propose a dual attention mechanism, together with an addictive approximate integral loss function, to automatically identify relevant anomalous behaviors to explain anomalous events. The anomalous behaviors can be represented by the most relevant sensor variables with their temporal anomaly instances instance anomalies. Also, in this mechanism, a temporal pseudo-label predicted using temporal attention and sensor variables' importance obtained from delta feature attention are combined to better identify anomalies from instance-level labels. Then, these labels are used in the proposed MIL-based framework to establish relationships between anomalous events and anomalous behaviors. Experiments conducted on real UAV datasets and comparisons with other methods are performed to demonstrate the effectiveness and robustness of the proposed method.

The main contributions of this article are summarized below.

1) We propose a deep attention-based multi-instances learning (DA-DI-MIL) framework to address the anomaly detection problem with an event-based explanation of anomalous events as a weakly supervised learning problem, in a real-world UAV flight scenario. Our DA-DI-MIL method offers an efficient way to detect anomalous events with instance-level labels, together with automatically explaining anomalous events.

TABLE I
SUMMARY FOR ANOMALY DETECTION METHODS

References	Learning Techniques	Anomaly Detection Methods	Input	Output	Constrain
Guo et al. [3], Ji and Lee [4]	Unsupervised shallow structure model	[3] [4] OC-SVM optimized without negative samples	Multivariate time series, solely nominal data for training	Time-step instance labels	Hard to represent complex data, not present temporal information
Zhang et al. [5], Oehmcke et al. [6], Liu et al. [7]	Unsupervised shallow structure model	[5] PDD descriptions with sliding windows, [6] LOF, [7] iForest	Multivariate time series, solely nominal data for training	Time-step instance labels	Hard to represent complex data, not present temporal information
Hundman et al. [9], Wang et al. [10], Zhong et al. [11]	Unsupervised deep structure model	LSTM with [9] dynamic anomaly thresholding, [10] a residual filter, [11] spatio-temporal correlations	Multivariate time series, only nominal data for training	Time-step instance labels	Focus solely on detecting anomalies, Lack event-based explanation
Deng and Hooi [12]	Unsupervised deep structure model	Use learned graph to represent sensors' relationship to localize and understand an anomaly	Multivariate time series, solely nominal data for training	Time-step instance labels	Offer limited information of anomalies, Lack event-based explanation
Liang et al. [13]	Unsupervised deep structure model	Utilize the reconstruction metrics deviation in the original space to locate anomalous dimensions	Multivariate time series, solely nominal data for training	Time-step instance labels, anomalous dimensions	Lack event-based explanation
Janakiraman et al. [18]	Weakly supervised deep structure model	DT-MIL, capture temporal behaviors of time series	Multivariate time series, both nominal and adverse data with segment-level labels for training	Time-step instance labels, segment-level labels	Not capable of automatically identifying related anomalous behaviors from sensor variables, need both nominal and adverse data
Laine et al. [19]	Weakly supervised deep structure model	MHCNNRNN, build individual sub-models to identify sensors variables	Multivariate time series both nominal and adverse data with segment-level labels for training	Time-step instance labels, segment-level labels, key sensor variables	Time consuming, cause inconsistent evaluation, Lost the correlations between sensors, need both nominal and adverse data

2) To better pinpoint anomaly instances from instance-level labels, a dual attention mechanism is employed, consisting of temporal attention and delta feature attention, together with an addictive approximate integral loss function. This mechanism is particularly effective for identifying continuous anomalies when compared to other methods.

3) In addition to time-step labels of anomaly instances, the available event-based explanation expresses how the DA-DI-MIL model behaves in the presence of anomalous events in a qualitative manner. Also, an illustration in global and local interpretation both in the time domain and feature space is presented to further validate the effectiveness and robustness of the DA-DI-MIL method.

II. PROBLEM OVERVIEW

This article proposes a dual attention-based multi-instance learning (DA-DI-MIL) approach to deal with anomaly detection with an event-based explanation of an anomalous event. The MIL-based framework is from a weakly supervised learning, typically assuming that a set of unlabeled instances is grouped in the form of a labeled bag. Additionally, by virtue of the low requirement of labeling and its capacity of making use of both nominal and adverse records, it could enhance the capacity of finding anomaly instances in complicated situations of UAV, a sophisticated cyber-physical system [20].

Given each flight record as a bag with a label and its time steps as instances without labels, the task of the MIL-based framework aims to detect an anomalous event and also discover anomalous behaviors which are relevant to the anomalous event. The anomalous behaviors are defined as a representation both in the time domain and feature space, that is most relevant sensor variables with their consecutive temporal anomaly instances.

Suppose the multivariate time series are represented by $X \in R^{(b\times t\times d)}$ with dimensions (b, t, d), where b is the number of flight records and t is the length of time series and d is the number of sensory variables, then An instance E_i in a bag $B = E_1, E_2, \ldots, E_t$ with $E_i \in R^{(1\times d)}$ is given by,

$$E_i = \{x_i^1, x_i^2, \ldots, x_i^d\} \tag{1}$$

where x_i^m is ith time step of mth sensory variable. Assuming that each bag's label is denoted by Y, $Y = 1$ states that the bag is positive and at least an anomaly instance occurs; otherwise, $Y = 0$, is negative and no anomaly instance exists. In the MIL

framework, each temporal instance in B is with an individual label, i.e., $y_1, y_2, \ldots, y_t$, but not available. Based on the standard MIL assumption, the relationship between the bag label and the instances labels is shown below:

$$Y = \begin{cases} 0, & \sum_{i=0}^{t} y_i = 0 \\ 1, & \text{otherwise} \end{cases} \tag{2}$$

From (2), when at least one instance is labeled as positive, the bag label is set as 1. So, given the input data X and the bag label Y, the goal of anomaly detection in time domain is to find a set of temporal anomaly instances in a bag or record for which,

$$P(y_i = 1|B) > \delta,\ Y = 1 \tag{3}$$

where δ is the threshold used to define a temporal anomaly instance. When an instance's probability exceeds δ, it is regarded as a temporal anomaly instance and its' label y_i is set as 1.

Furthermore, when most relevant sensor variables are identified when in a positive bag, the representation of anomalous behaviors $\mathcal{H}$ can be expressed as,

$$\mathcal{H} = \{x_{l_1}^{\{c_1,c_2,\ldots,c_C\}}, x_{l_2}^{\{c_1,c_2,\ldots,c_C\}}, \ldots, x_{l_L}^{\{c_1,c_2,\ldots,c_C\}}\} \tag{4}$$

where $x_l^{\{c_1,c_2,\ldots,c_C\}} = \{x_l^{c_1}, x_l^{c_2}, \ldots, x_l^{c_C}\}$ is a anomaly instance at lth time step, C is the number of most relevant sensor variables, and L is the number of anomaly instances. The set $I = \{l_1, l_2, \ldots, l_L\}$ with $1 \leq l \leq t$ states the time steps in which anomaly instances occur and the set $M = \{c_1, c_2, \ldots, c_C\}$ with $1 \leq c \leq d$ represent the index of most relevant sensor variables. Therefore, anomaly instances in I, M are regarded as a quantitative representation of anomalous behaviors, an even-based explanation of anomalous events.

III. Datasets Description and Preprocessing

In this work, we use the ALFA datasets for UAV fault and anomaly detection [21] to verify our proposed explainable MIL-based anomaly detection method. The dataset includes 47 autonomous flights of eight fault types, including the failure of the engine, rudder, elevator, and aileron. This article mainly focuses on high-risk engine failure with nearly 50% of the full dataset (23 flight tests) to verify the capacity of detecting anomalies and explaining the anomalous event of the fault.

The experiment consists of 23 flights of engine failure (or adverse data) and 8 flights of no failure (or nominal data) with ground truth (the time when the failure happens) but no ground truth for which sensor variables are most relevant to the failure. For flights of no failure, each record of the time series is cut into several bags with 30 seconds recordings of UAV. For flights of engine failure, only the last 30 seconds' recordings are selected and we regard the duration of engine failure as the time that anomalies occur. So, we have 23 adverse samples and 13 nominal samples in total. Note that 2 fights of no failure are not used due to its length of less than 30-second recordings. The anomalies in the flight data are collective anomalies with variable lengths, which happen around the last 15 seconds. Each bag is labeled but instances of the bag are not. We choose a subset of 35 sensory variables based on domain knowledge, including pitch angle, roll angle, yaw angle, velocity along 3 axes, airspeed, altitude, latitude, longitude, atmospheric pressure, wind estimation, throttle, climb, and other variables. Note that we also add the errors of the command and measured values of pitch, roll, and yaw angles into the input variables.

Next, data preprocessing is conducted before feeding into a model. First, we resample the selected variables at the same frequency due to their different sampling rates. Then, the non-stationary multivariate time series of UAVs will bring difficulty in conducting anomaly detection with original data. So, we lighten the influence by obtaining their first-order difference to make sure that the time series are near the stationarity and concentrate on the incremental changes, which is helpful to enhance the performance of anomaly detection [22].

Besides, Gaussian-based data augmentation [23] is performed by using addictive Gaussian noise (AWGN). Given a signal-to-noise ratio (SNR), the target output $x(k)$ with AWGN is expressed as below,

$$x[k] = x_0[k] + n[k], n[k] \sim \mathcal{N}(0, \sigma_{noise}) \tag{5}$$

$$SNR = 20\log_{10}\frac{\sigma_{signal}}{\sigma_{noise}} \tag{6}$$

where σ_{signal} and σ_{noise} are standard deviation of original signal $x_0[k]$ and noise $n[k]$. Then for each variable in time series, σ_{noise} is set as,

$$\sigma_{noise} = \frac{1}{(10)^{\frac{SNR}{20}}} \tag{7}$$

The different Gaussian distribution is used according to the different standard deviations of each variable. Applying data augmentation in a training process not only ameliorates the difference in each flight but also makes the algorithm robust to noise in case of overfitting of the model [24]. Note that other methods for data augmentation can also be applied, such as the SMOTE-based method [25] and adversarial attack generation techniques [26].

IV. Model Architecture

In this section, we present the methodology for addressing the aforementioned issues in MIL-based anomaly detection with an event-based explanation. Fig. 2 illustrates the architecture of our proposed deep attention-based MIL model, which primarily comprises the LSTM neural network, dual attention mechanism, and MIL aggregation.

First, the LSTM neural network is designed to capture the temporal information of instances and obtain new transformed representations of each sensor variable. Next, we introduce a dual mechanism to automatically identify anomalous behaviors to offer an explanation of anomalous events. In the mechanism, we propose a temporal pseudo-label predicted by temporal attention to defining nominal and adverse areas. Additionally, we obtain sensor variables' importance from delta feature attention based on these two areas. Subsequently, we obtain a deep representation of input variables based on temporal and feature attention, which is then fed into a logistic layer to obtain instance labels. Finally, the bag label is obtained after MIL aggregation.

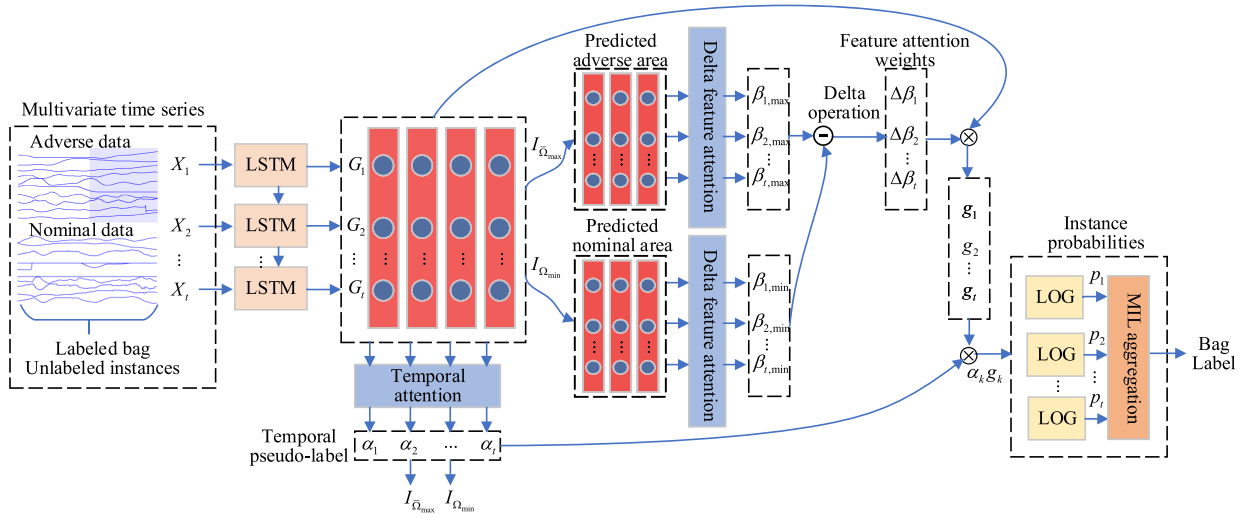

Fig. 2. Model architecture for the proposed DA-DI-MIL method.

During training, the model outputs anomalous event labels given multi-variate time series as inputs. During prediction, the model performs forward and back-forward processes, including labeling anomalous events and offering anomalous behaviors' representation, respectively. After the training and prediction phases, anomalous events detection and their explanation could be obtained. As mentioned earlier in Section II, anomalous behaviors, as an explanation of anomalous events, consists of relevant sensor variables and their temporal anomaly instances. Here, temporal anomaly instances can be defined by instance labels and relevant sensor variables can be identified from delta feature attention which states the impact level of which sensory variables affect the anomalous event.

A. Attention Mechanism

Given the LSTM network's capability of capturing temporal behaviors of time series, the inputs are firstly processed by the network, outputting a transformed representation of input variables. Next, to identify anomalous behaviors of an anomalous event in both the time domain and feature space, we adopt an attention mechanism to address the issue, mainly referring to the deep multi-instance contrastive learning approach with dual attention (MCDA) proposed by Xu et.al [27]. Based on the MCDA method, we further develop a dual attention mechanism to automatically pinpoint the anomalous behaviors for the anomalous event. In contrast to existing dual-attention methods, it needs to obtain a temporal pseudo-label to define the predicted nominal and adverse areas, later used in delta feature attention. In addition, the delta operation used in delta feature attention helps with identifying the most relevant sensor variables to anomalous events. It enlarges the feature weights difference between the most and less relevant sensors by obtaining the difference of weights between nominal and adverse areas.

Suppose the transformed representation after the LSTM layer is denoted as $G \in R^{(b\times t\times d)}$, we use the gated attention mechanism together with $tanh(\cdot)$ to define the temporal weights α. The selection of the gated attention mechanism considers that merely using $tanh(\cdot)$ function may not be able to express complex relations in multi-variate time series [28]. In this mechanism, the temporal weight $\alpha \in R^{(b\times t\times 1)}$ are rewritten as,

$$\alpha = \sigma\{(relu(G \cdot V_T) \bigodot \sigma(G \cdot U_T)) \cdot w_T\} \tag{8}$$

where $U_T \in R^{(d\times s_T)}$, $V_T \in R^{(d\times s_T)}$, $w_T \in R^{(s_T\times 1)}$ are the trainable parameters of temporal attention mechanism, and s_T is a hyper-parameter of $\{U_T, V_T, w_T\}$. $\bigodot$ is an element-wise multiplication and $\sigma(\cdot)$ is the sigmoid non-linearity. U_T, V_T, w_T controls what information should flow through the hierarchy of layers in the gated attention mechanism [29]. Here, we adopt the $\sigma(\cdot)$ instead of the $softmax(\cdot)$. It is because the temporal attention weights need to state how likely a temporal anomaly instance will happen with respect to time steps. In this case, it is not necessary to normalize the weights and make the sum of weights equal to one.

What's more, according to α, we can obtain temporal pseudo-label $\bar{p}$ which can be used to roughly define the predicted nominal Ω and adverse areas $\bar{\Omega}$ in a bag. If the length of the adverse area is $r = \xi * t$ where ξ is a value between 0 and 0.5, then $\bar{\Omega}$ is the area with the r largest temporal weights in α, while Ω is the area with the r lowest temporal weights in α.

Similarly to temporal attention, the feature weights $\Delta\beta \in R^{(b\times d\times 1)}$ with delta feature attention for sensory variables in flight data are calculated by,

$$\Delta\beta = softmax(\beta_{\max} - \beta_{\min}) \tag{9}$$

$$\beta_{\max} = (tanh(G^\top(I_{\bar{\Omega}_{\max}}) \cdot V_D) \bigodot \sigma(G^\top(I_{\bar{\Omega}_{\max}}) \cdot U_D)) \cdot w_D \tag{10}$$

$$\beta_{\min} = (tanh(G^\top(I_{\Omega_{\min}}) \cdot V_D) \bigodot \sigma(G^\top(I_{\Omega_{\min}}) \cdot U_D)) \cdot w_D \tag{11}$$

where $U_D \in R^{(t\times s_D)}$, $V_D \in R^{(t\times s_D)}$, $w_D \in R^{(s_D\times 1)}$ are the trainable parameters, and s_T is a hyper-parameter of $\{U_D, V_D, w_D\}$. The $\beta_{\max} \in R^{(b\times d\times 1)}$ and $\beta_{\min} \in R^{(b\times d\times 1)}$ are the local feature weights of adverse areas $\bar{\Omega}$ and nominal areas Ω,

respectively. $G^{\top}(I_{\bar{\Omega}_{\max}})$ is the LSTM's partial output according to $\bar{\Omega}$, whereas $G^{\top}(I_{\Omega_{\min}})$ is the LSTM's partial output according to Ω. The $\Delta\beta$ describes the feature importance which shows the impact level of each variable on the anomalous event.

In delta feature attention, obtaining the difference of weights between nominal and adverse areas instead of only $\bar{\Omega}$ aims to enlarge the feature weights difference between the most and less relevant sensors. The delta operation in feature attention helps with identifying the most relevant sensor variables to anomalous events. To be more specific, in general, more relevant sensor variables always tend to have a more apparent difference in behaviors in nominal and adverse areas. Under this circumstance, the delta feature operation makes use of the discrepancy between most and less relevant sensor variables' behaviors when an anomalous event shows up. So, it is reasonable and helpful to use the delta operation in feature attention.

Therefore, we can obtain the set $M = \{c_1, c_2, \ldots, c_C\}$ by selecting the C sensory variables with most C largest weights of $\Delta\beta$ to explain the anomalous event from the perspective of feature space. The higher the feature's weight value is, the better the corresponding variables explain the anomalous event.

B. MIL Aggregation

Given the temporal attention α, feature weights $\Delta\beta$, and transformed representation G, we can obtain the instance-level label and bag-level label with probability using MIL aggregation with the max-pooling operation. The max-pooling operation is a good choice for the instance-level approach we adopt in this article. Besides, the probabilities of instances can be used to find all-possible anomalies in the bag.

First, we calculate the element-wise product between the transformed representation G' and the feature attention $\Delta\beta$ to obtain the deep representation of inputs,

$$G' = tanh(G \bigodot \Delta\beta^{\top}), G' \in R^{(b\times t\times d)} \tag{12}$$

Then we calculate the sum of weighted transformed representation G' along the axis of the time-step,

$$G'' = (g_1, g_2, \ldots, g_t), g_i \in R^{(b\times 1)} \tag{13}$$

After the transformation: $G' \in R^{(b\times t\times d)} \Rightarrow G'' \in R^{(b\times t\times 1)}$, the instances probabilities are obtained by fed into a logistic layer (LOG),

$$(p_1, p_2, \ldots, p_t) = LOG(\alpha_1 g_1, \alpha_2 g_2, \ldots, \alpha_t g_t) \tag{14}$$

Then, the representation of bag probability can be derived by max pooling operation,

$$y = P(Y = 1|X) = \max_{k=1,2,\ldots,t}\{(p_1, p_2, \ldots, p_t)\} \tag{15}$$

If $y > \delta$, then it means at least an anomaly happens in the bag with the probability of y; otherwise, it states that no anomaly occurs in the bag. Besides, for each instance E_i, if $p_i > \delta$, it means that an anomaly occurs at time i. Then we can find all possible time steps of anomalies based on the instances probabilities and the threshold. Then we can find temporal anomalous segments defined by the set $I = \{l_1, l_2, \ldots, l_L\}$ with all possible time steps of anomalies.

C. Objective Function

To train the model, our objective function consists of two parts, a binary cross-entropy loss function, and an addictive approximate integral loss function. The binary cross-entropy loss function is commonly used in the MIL-based framework which aims to improve the performance of the classification of bag detection. However, multivariate time series in the flight of UAV always refer to collective anomalies which are a sequence of anomalous values. The training model only with the loss cannot meet the requirement of our task for detecting consecutive anomalies in UAVs. In this case, we propose an addictive approximate integral loss function to address the issue of consecutive anomalies occurring in multivariate time series data of UAVs. Therefore, the total loss function in a weighted way is shown below,

$$L = e * L_{bina}(y, \tilde{y}) + (1 - e) * L_{int}((p_1, p_2, p_3, \ldots, p_t)) \tag{16}$$

$$L_{bina}(y, \tilde{y}) = -[y \log(\tilde{y}) + (1 - y) \log(1 - \tilde{y})] \tag{17}$$

$$L_{int} = -\log\left(\frac{\int_0^t p_\tau d\tau}{t}\right) \approx -\log\left[\frac{\left(\sum_i^{int(\frac{t}{\Delta t})} p_{i*\Delta t}\Delta t\right)}{t}\right] \tag{18}$$

where y and $\tilde{y}$ are the true and predicted bag probabilities. Δt is the unit of time step used for estimating the approximate integral loss. As mentioned in the model, the max-pooling operation is used to build the relationship between the bag and instances. It may be good for discovering point anomalies but ignoring the collective anomalies happening in UAV's flight data. The reason is that when at least one anomaly shows up, the bag label will be set to one. Under this circumstance, not all anomalies will be found. What's more, due to similar behaviors in collective anomalies, it is challenging to discover the full sequence of anomalous segments with variable lengths. So, we introduce an integral loss function to restrain the optimization direction of the model in order to discover as many temporal anomaly instances as possible with integral impact. We add up all the probabilities of instances to ensure that more anomalies will be found with integral impact. As seen from the function, when the probabilities of all instances are set at 1, the integral loss will be zero. This extreme phenomenon seems counter-intuitive, which can be solved by adding a small weight to it, as seen in (16).

V. Experiments and Discussion

A. Baseline Methods and Model Evaluation

We implement the proposed method (named DA-DI-MIL) with TensorFlow-GPU 2.4.0. The model is trained in an end-to-end way for 50 epochs with a Nadam optimizer and the learning rate is set as 0.001. Besides, we conduct 9-fold stratified cross-validation on the ALFA dataset and repeat the test 10 times with continuous seeds to validate the effectiveness and robustness of the proposed model.

We conduct experiments on three tasks: Task 1 is anomalous event detection which aims to determine whether our proposed method can identify the event in multi-variate time-series flight data. Task 2 is temporal anomaly instances detection which pinpoints the time of anomalies in an anomalous event and Task 3 is to select the most relevant sensor variables related to the event. Task 2 and Task 3 aim to obtain the two components of anomalous behaviors for explaining the anomalous event. In the experiments, we do not use the point adjust approach [30] for the evaluation strategy of temporal anomaly detection. The point adjust approach considers any point within the ground truth anomalous segments as correctly detected anomalies. In contrast, in our article, an instance is considered an anomalous instance only when its probability exceeds the threshold δ.

Moreover, we compare our proposed method with five state-of-art methods in anomaly detection, Local Outlier Factor (LOF) [6], Isolation Forest (iForest) [7], One-Class Support Vector Machines (OC-SVM) [31] and DT-MIL [18] and MHC-NNRNN [19]. Meanwhile, to evaluate of key components in the DA-DI-MIL, and its variants are conducted. DA-MIL is the variant without the delta operation of feature attention and the introduced integral loss. DA-D-MIL does not include the integral loss in the loss function, while DA-I-MIL does not use the delta operation when doing feature attention. Note that shallow structure methods, like LOF, Forest, and OC-SVM, in anomalous event detection are omitted, which are selected due to their high performance in anomaly detection. To maintain simplicity and clarity, we will refer to the delta operation and integral loss as the D and I modules, respectively, throughout the article.

This article adopts common evaluation metrics to illustrate anomalous events detection and its event-based explanation, i.e., Precision, Recall, F1 score, and Accuracy. Considering event detection in reality, we redefine the true positive used in the above evaluation metrics but remain it in temporal anomaly instances detection. It is because that in reality merely relying on original evaluation may not be applicable when some algorithms fail to accurately pinpoint event locations but correctly reports whether the event is in flight. Therefore, the true positive in event detection is corrected as one that outputs the right label from the right location belonging to the ground truth.

Meanwhile, the average detection delay (ADD) and standard deviation of feature attention weights $\Delta\beta$ (SFA) are raised. ADD shows the delay between the event happening and the detection. In a real application, a large detection delay is unacceptable, which would lead to irreversible circumstances and make UAVs out of control [21]. The detection time is presented as the first time hitting the threshold δ in the period of ground truth when the failure happens. Suppose the start time and the end time of the ground truth is T_{start} and T_{end}, and the first hitting time is T_1, then

$$ADD = \begin{cases} 0, & if\ 0 < T_{start} - T_1 < \zeta \\ T_1 - T_{start}, & if\ T_1 \geq T_{start} \\ T_{end} - T_{start}, & otherwise \end{cases} \tag{19}$$

where ζ represents a tolerance value used to determine the acceptable range for the difference between the true start time of the anomalous event. The purpose of using ζ in the equation is to account for a small time window within which we consider the detection to be accurate and correct. If the difference between T_{start} and T_1 falls within this threshold $(0, \zeta)$, we consider the detection to be perfect (i.e., $ADD = 0$). This condition ensures that even minor discrepancies in the detected start time are considered accurate detections within the specified tolerance range.

However, in cases where the algorithm fails to precisely pinpoint the start time of the anomalous event, we calculate ADD by taking the difference between the true start time and the true end time of the event ($T_{end} - T_{start}$). This scenario occurs when T_1 is either far before the true start time or after the true end time of the anomalous event. In such situations, we consider the entire duration of the ground truth anomaly as the return for ADD, capturing the full length of the anomalous event.

The evaluation metric ADD plays a crucial role in assessing the accuracy of the anomalous event detection performance, and it is essential for providing sufficient time to address and respond to anomalous events promptly. Note that ADD is measured in 'time steps' (Δt), where each time step corresponds to a fixed discrete interval of time, enabling precise quantification of the detection delay in the analysis.

The SFA reflects the performance of interpreting which sensory variables are involved in the anomalous event. The proposed SFA is shown as follows,

$$SFA = \sqrt{\frac{\sum_{i=0}^{d}(\Delta\beta_i - \Delta\bar{\beta})^2}{d}} \tag{20}$$

where $\Delta\bar{\beta}$ is mean of $\Delta\beta$ and d is the number of the sensor variables. For SFA, the higher value of it states the better performance in locating anomalous behaviors from sensory variables. To be more specific, not all selected sensory variables are relevant to the anomalous event, but only some parts of these variables. So, the weights of sensory variables more relevant should be much higher than the less relevant variables, which increases the standard deviation of $\Delta\beta$. Besides, from our mathematical derivation, the range of the SFA is $[0, \sqrt{d-1}/d]$.

B. Experiments Results and Analysis

1) Anomalous Event Detection: We compare our proposed method with other methods for detecting anomalous events in engine failure (Task 1). The corrected evaluation metrics are used, like precision, recall, F1, and accuracy, with redefined true positives. The results of corrected anomalous event detection can be viewed in Table II. Also, Fig. 3 shows the difference before and after correction in terms of evaluation metrics. The performance analysis of DA-DI-MIL and its variants reveals consistent results in evaluation metrics both before and after correction. This consistency indicates that these methods not only produce correct event labels but also demonstrate accurate event localization capabilities. In other words, they can effectively identify and locate anomalous events within the UAV flight data. In terms of comparison of DT-XX-MIL methods, DA-MIL performs the best in Task 1, outperforming DA-D-MIL and DA-I-MIL. The joint D and I modules in DA-DI-MIL does not seem to offer significant advantages over using an individual

TABLE II
CORRECTED ANOMALOUS EVENT DETECTION RESULTS

Methods	Precision	Recall	F1	Accuracy	ADD (Δt)	Trainable Variables	Model size (KB)
MH-CNN-RNN	0.6680±0.2360	0.5722±0.1916	0.5326±0.2491	0.5722±0.1916	220.6778±49.7814	66646	15088
DT-MIL	**0.9290±0.0469**	**0.9222±0.0539**	**0.9202±0.0569**	**0.9222±0.0539**	34.5435±92.8548	**8745**	**145**
DA-MIL	0.9236±0.0147	0.9139±0.0194	0.9107±0.0217	0.9139±0.0194	15.6913±39.1879	22204	316
DA-I-MIL	0.9150±0.0279	0.9000±0.03770	0.8945±0.0421	0.9000±0.0377	**2.6261±4.3241**	22204	317
DA-D-MIL	0.9164±0.0230	0.9056±0.0255	0.9017±0.0275	0.9056±0.0255	18.6522±40.6995	22204	316
DA-DI-MIL	0.9025±0.0250	0.8833±0.0347	0.8759±0.0394	0.8833±0.0347	4.6609±11.2050	22204	317

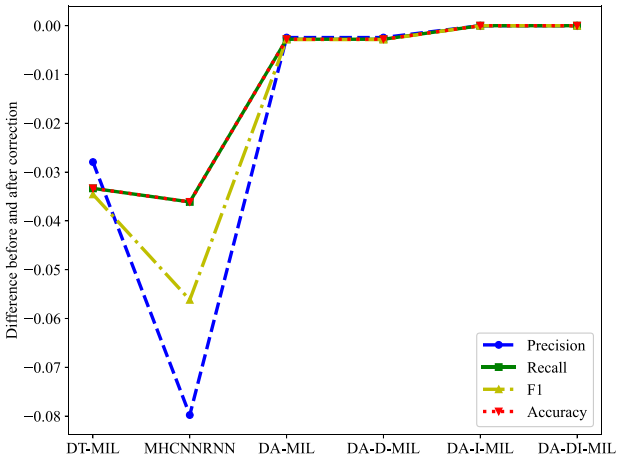

Fig. 3. Evaluation metrics results before and after correction.

module (DA-D-MIL or DA-I-MIL) for detecting anomalous events.

Furthermore, as discussed earlier in Section V, some algorithms may correctly identify the occurrence of anomalous events but struggle to precisely pinpoint their locations. In such cases, these algorithms may be considered as false positives, as they correctly report the presence of an anomaly but lack accuracy in localizing the event. This aspect is crucial for ensuring high-confidence in-time detection of anomalous events during UAV flights. On the other hand, DT-MIL, despite achieving the best results in Table II, exhibits a relatively larger difference between the original and corrected metrics. This discrepancy suggests that DT-MIL, along with MHCNNRNN, may face challenges in precisely pinpointing event locations in certain scenarios. While they achieve correct event labeling, their localization capabilities may be less accurate.

Additionally, in Table II, it is evident that both DA-DI-MIL and DA-I-MIL achieve lower ADD values (less than 5 time steps) compared to the other methods. This result indicates that they can detect anomalous events much earlier, showcasing their superior performance in in-time anomaly detection. A detailed explanation is given by the following Task 2, considering that the early detection capability of these methods is closely related to their effectiveness in temporal anomaly detection. Overall, our proposed method, DA-DI-MIL, exhibits strong performance in detecting anomalous events. Its ability to detect anomalies early on is a valuable feature, as it allows for timely responses and mitigates potential risks associated with anomalous events during UAV flights.

Task 2 involves detecting temporal anomaly instances, which is the specific location of anomalous instances in the time domain when an anomalous event occurs. Table III displays the results of temporal anomaly instances detection. DA-DI-MIL performs the best in Task 2, indicating that the combined use of D and I modules leads to improved performance in detecting temporal anomalies at the instance level. Also, the results show that DA-DI-MIL and its variants perform better than shallow structure models (OCSVM, LOF, and iForest) and other deep structure models (DT-MIL and MHCNNRNN). Compared to event detection, anomaly detection tends to be more complicated, especially with high-dimensional time series. This complexity explains shallow-structure methods' poor performance in Task 2. In this case, DT-MIL and MHCNNRNN as deep structure models behave slightly better than shallow structure models.

The superior performance of DA-DI-MIL in temporal anomaly instances detection benefits from one of its modules, addictive integral loss in training. The DA-I-MIL performs better than DA-MIL without additive integral loss. The introduced integral loss optimizes the direction of finding anomalies, ensuring detecting more temporal anomaly instances effectively. Moreover, the ADD values mentioned in Task 1 are closely related to temporal anomaly instances detection performance. It is because the ADD depends on the first hitting time when an instance's probability exceeds a certain threshold. By leveraging the integral loss module, DA-DI-MIL achieves enhanced detection of temporal anomalies, which in turn increases the likelihood of identifying the earliest anomaly instance. Consequently, DA-DI-MIL demonstrates a lower ADD value, indicating its ability to identify anomalies at an earlier stage.

In Task 3, we use the SFA to illustrate the DA-DI-MIL's ability to identify the most relevant sensor variables to an anomalous event. To some extent, a higher SFA value indicates that the model is better at selecting the most relevant sensor variables, assuming that not all sensor variables are related to an anomalous event in reality. As shown in Table III, DA-D-MIL performs the best in Task 3, suggesting that the D module is particularly effective in identifying the most relevant sensor variables. However, DA-DI-MIL still performs relatively well, indicating that the combination of both D and I modules allows it to capture important sensor variables effectively. DA-D-MIL validates the effectiveness of its D module in identifying key sensor variables. Conversely, DA-MIL without the D module has a lower SFA value. Specifically, the delta operation within the module obtains the delta attention in nominal and adverse areas, which enlarges the difference between the importance of the most and less relevant variables regarding an anomalous event. In other words, the operation increases the statistical dispersion of sensor feature weights, resulting in higher SFA values.

TABLE III
TEMPORAL ANOMALY INSTANCES DETECTION RESULTS

Methods	Precision	Recall	F1	Accuracy	SFA	Trainable Variables	Model size(KB)
OCSVM	0.3423+0.0001	0.3486+0.0002	0.3327+0.0001	0.3486+0.0002	-	37	4064
MHCNNRNN	0.5387±0.1577	0.5394±0.023	0.3978±0.0411	0.5394±0.023	0.0026±0.0005	66646	65
LOF	0.5210+0.0000	0.4865+0.0000	0.3944+0.0000	0.4865+0.0000	-	**1**	684
iForest	0.3916+0.0162	0.4639+0.0022	0.3183+0.0023	0.4639+0.0022	-	1212644	427
DT-MIL	0.5121±0.0814	0.5441±0.0302	0.4268±0.0491	0.5441±0.0302	-	8745	**9**
DA-MIL	0.8419±0.1393	0.822±0.1159	0.796±0.155	0.822±0.1159	0.0167±0.011	22204	22
DA-I-MIL	0.8505±0.1386	0.8417±0.1232	0.8158±0.1611	0.8417±0.1232	0.0162±0.012	22204	22
DA-D-MIL	0.8211±0.1422	0.7950±0.1160	0.7684±0.1550	0.795±0.1160	**0.0222±0.0104**	22204	22
DA-DI-MIL	**0.8750±0.1216**	**0.8518±0.1181**	**0.8343±0.1501**	**0.8518±0.1181**	0.0218±0.0120	22204	22

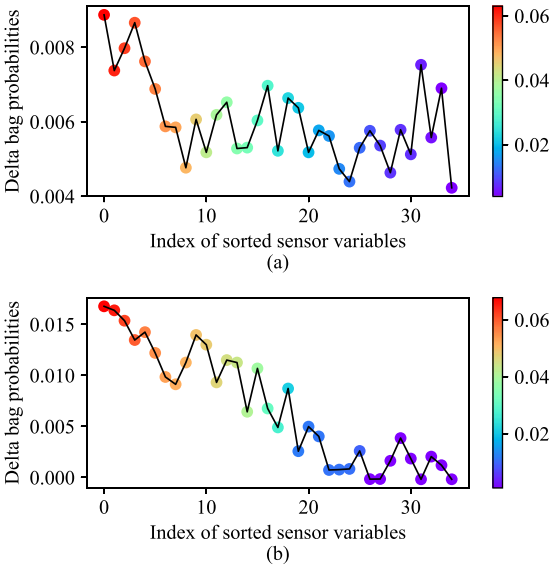

Fig. 4. Masking test for sensor identification verification. (a) Repetitive K-fold tests and (b) One K-fold test.

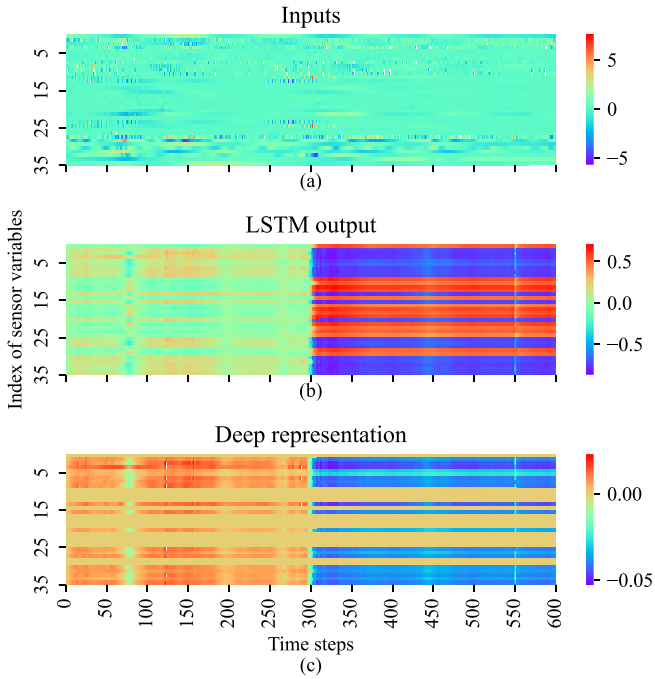

Fig. 5. Heatmaps of proposed model's layers' outputs in adverse sample #13.

Furthermore, a masking test is conducted to verify the rationale of sensor identifications in Task 3 above. Given a trained model, this involves masking one value in the delta feature weights ($\Delta\beta$) at a time, in descending order of weights, and calculating the difference between the current output of bag probability and the former output without masking. This approach highlights the corresponding sensor variable's impact on the anomalous event, with a more significant difference indicating a higher impact. The results of this test are shown in Fig. 4, where the colored points are plotted according to the feature weights of the variables. As seen from Fig. 4(a) and (b), the delta bag probabilities have an approximately linear relationship with the index of sorted sensor variables by weights, particularly for one K-fold test. In summary, the analysis demonstrates the selected sensor variables with the highest weights have a higher impact on the anomalous event.

2) Interpretation for Anomalous Event: The above experiments demonstrate quantitatively the effectiveness of our proposed method for detecting anomalous events and explaining them in both the time domain and feature space. To further validate the interpretation of anomalous events by our proposed method, we employ visualization techniques in the following section.

First, we use the adverse sample #13 as an example to illustrate how the model behaves in the time domain and feature space. We use heatmaps to present important layers of the DA-DI-MIL model in #13, shown in Fig. 5, including inputs X, LSTM outputs G, and deep representation G'. The ground truth when engine failure begins is at 308th time steps and the estimated is at 297th. In Fig. 5(b), from the temporal perspective, the image of LSTM outputs is clearly cut into two areas, one nominal area and one adverse area, which could be recognized by further instances probabilities in Fig. 6(a). The adverse area in the second area represents the time around 300th when an anomalous event occurs. In view of feature space, in Fig. 5(c) some sensor variables show distinct color differences between the normal and problematic areas, while other variables exhibit the same color in both areas. Therefore, the apparent difference in the time domain and feature space illustrates our proposed method has the capacity of identifying anomalous behaviors, presented by the time when anomalies occur and the most relevant sensor variables, for the purpose of explaining anomalous events.

Next, in terms of outputs in DA-DI-MIL, Fig. 6 shows that temporal anomaly instances' probability changes as time steps

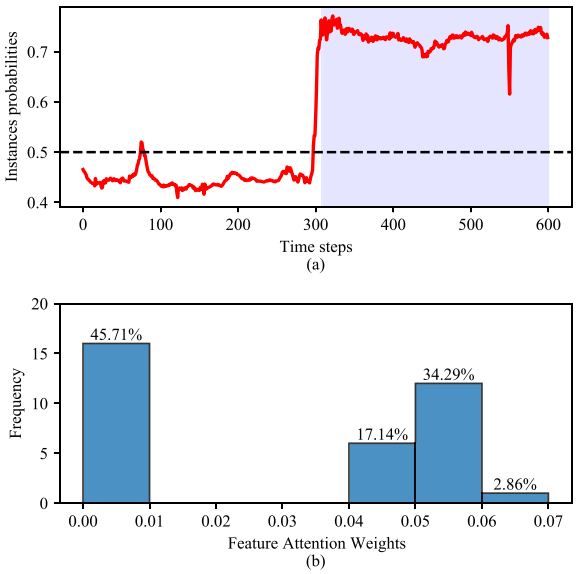

Fig. 6. Results for adverse sample #13. (a) Temporal anomaly instances' probabilities and (b) Distribution of feature attention weights of sensor variables.

increase and delta feature weights $\Delta\beta$ are varied in sensor variables in the #13 sample. The blue shaded area is the ground truth of the adverse area and the dotted black line is the threshold used to define temporal anomaly instances in Fig. 6(a). The estimated adverse area in #13 is identified by temporal anomaly instances' probability which presents how likely an instance is anomalous. If its probability is higher than the threshold of 0.5, it is an anomaly instance. From the red line and blue shaded area, our proposed method accurately pinpoints the adverse area including temporal anomaly instances. Also, when the engine failure begins, simultaneously we can see a sharp increase in the instances probabilities. It shows that the proposed method has a quick response when an anomalous event happens. In this case, the ADD value is zero which means there is almost no delay between the anomalous event happening and the detection.

In Fig. 6(b), the histogram of feature attention weights for variables reveals two distinct classes. Approximately 45.71% of the variables have weights lower than 0.01, while the majority of the remaining variables have higher weights, with a subtle difference between them. This pattern closely resembles the distribution of G' values shown in Fig. 5(c). The presence of these two distinct classes in the feature attention weights highlights the importance of certain variables in the analysis, while also suggesting a potential differentiation in their contributions to anomalous events.

What's more, regarding global interpretation in the model, we analyze all adverse samples in both the time domain and feature space. First, in Fig. 7, we provide a visual representation to demonstrate the effectiveness of our proposed DA-DI-MIL approach in detecting continuous anomalies in the time domain for all adverse samples. The figure consists of 9 sub-figures, each corresponding to adverse samples in a one-fold test. Each sub-figure represents the instances probabilities of being anomalous over time for the adverse samples. The cross marks in the sub-figure represent the true start time of the anomalous events, and the numbers indicate the index of the samples. The colors in the sub-figure range from blue to red, with blue indicating a lower probability of being an anomaly and red indicating a higher probability.

By analyzing Fig. 7, the proximity of the cross marks to the near-red points signifies that our proposed methods have a smaller delay in detecting anomalous events. This prompt detection of continuous anomalies showcases the effectiveness of our approach in capturing and responding to anomalous events more efficiently. Although our proposed method achieves the best overall performance in this task, we also observe some discontinuous segments, indicating that it fails to detect some temporal anomaly instances. This observation suggests that there is still room for improvement in our method.

Also, We obtain the total feature weights in one complete K-fold test, seen in Fig. 8. The results show that our proposed method has a consistent feature selection in adverse samples of engine failure. The subtle difference between samples is acceptable because it may be ascribed to variations in working conditions and inner states of UAVs in reality. In total, as shown in Figs. 7 and 8, our DA-DI-MIL approach demonstrates a consistently superior feature selection and outperforms other methods in detecting continuous anomalies in both time domain and feature space.

3) Parameter Sensitivity Test: In this section, we examine the impact of an important parameter ξ in our delta feature attention on the performance of our proposed method. Table IV presents the results, including precision, recall, F1, and accuracy for temporal anomaly instances, and ADD for event detection. The table shows that when ξ is set to 0.3, the proposed method performs the best in most cases. Furthermore, the performance of the proposed method is relatively insensitive to changes in ξ.

At the same time, we explore the impact of Gaussian-based data augmentation on modeling performance and its correlation with signal-to-noise levels. Specifically, we analyze the temporal anomaly instances detection results of SNR levels ranging from -20 to 100, as shown in Table V. The results indicate that the performance of DA-DI-MIL improves with decreasing SNR until it reaches around 60, indicating the effectiveness of Gaussian-based data augmentation within a certain range of SNR values. Moreover, augmentation can improve the robustness and accuracy of modeling.

Additionally, the ADD values demonstrate an unusual pattern where the trend is consistently low, below 10-time steps, until $SNR < 0$. Also, according to results in event detection, Additionally, the precision for event detection approaches 0.93 when $SNR = 0$. These unexpected patterns result from the different tolerance to noise between anomaly instance detection and event detection modeling. Event detection modeling is typically less sensitive to noise due to its simpler task. When the original signal dominates over the noise, Gaussian-based data augmentation can enhance the model's

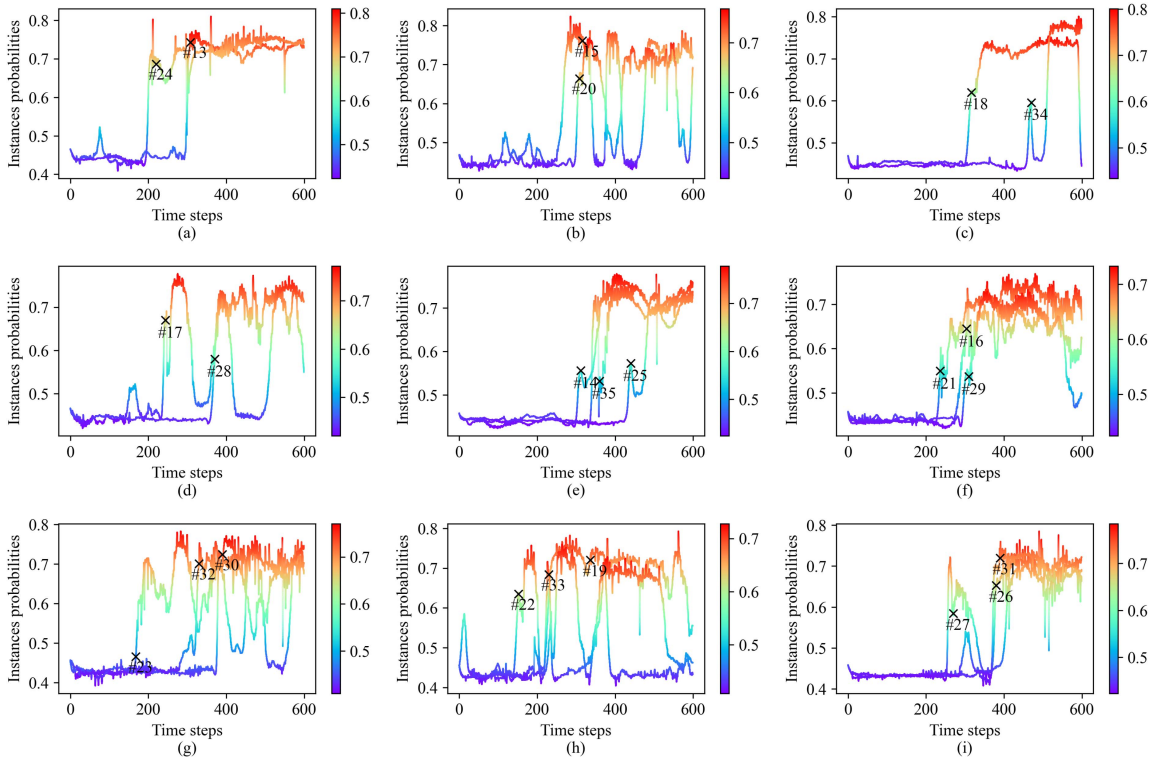

Fig. 7. Instance probabilities of a K-fold test.

TABLE IV
SENSITIVITY TEST

ξ	Precision	Recall	F1	Accuracy	SFA	ADD (Δt)
0.1	0.8328±0.1365	0.8092±0.1367	0.7803±0.1726	0.8092±0.1367	0.0162±0.0090	9.5957±32.3809
0.2	0.8519±0.1275	0.8327±0.1196	0.8106±0.1521	0.8327±0.1196	0.0187±0.0113	6.2522±16.9476
0.3	**0.8750±0.1216**	**0.8518±0.1181**	**0.8343±0.1501**	**0.8518±0.1181**	0.0218±0.0120	**4.6609±11.205**
0.4	0.8469±0.1099	0.8261±0.1067	0.8025±0.1383	0.8261±0.1067	0.0218±0.0124	7.8435±30.9436
0.5	0.8634±0.1153	0.8418±0.1128	0.8231±0.1434	0.8418±0.1128	**0.0222±0.0121**	5.3087±13.7105

TABLE V
SNR PERFORMANCE ANALYSIS RESULTS

SNR	Noise	Precision	Recall	F1	Accuracy	SFA	ADD (Δt)
-20	1000.000%	0.6082±0.1208	0.6067±0.0984	0.5356±0.1337	0.6067±0.0984	0.0071±0.0058	19.9652±53.4988
-10	316.228%	0.7146±0.1462	0.7113±0.1278	0.6504±0.1700	0.7113±0.1278	0.0132±0.0074	10.8478±42.7629
0	100.000%	0.8260±0.1527	0.8145±0.1404	0.7837±0.1811	0.8145±0.1404	0.0222±0.0125	5.7217±27.3839
5	56.234%	0.8398±0.1457	0.8192±0.1342	0.7912±0.1733	0.8192±0.1342	0.0221±0.0130	**5.1913±15.6492**
10	31.623%	0.8441±0.1280	0.8229±0.1223	0.7989±0.1564	0.8229±0.1223	0.0226±0.0122	7.5261±36.3731
15	17.783%	0.8541±0.1279	0.8349±0.1216	0.8127±0.1546	0.8349±0.1216	0.0217±0.0116	8.5913±37.6053
20	10.000%	0.8628±0.1224	0.8409±0.1188	0.8187±0.1522	0.8409±0.1188	0.0216±0.0120	10.9391±40.1961
25	5.623%	0.8395±0.1229	0.8234±0.1152	0.8008±0.1470	0.8234±0.1152	0.0228±0.0109	8.3391±30.1518
30	3.162%	0.8471±0.1276	0.8274±0.1257	0.8061±0.1587	0.8274±0.1257	0.0231±0.0109	7.6870±29.7440
40	1.000%	0.8566±0.1243	0.8329±0.1227	0.8116±0.1552	0.8329±0.1227	0.0223±0.0118	11.1913±41.0183
50	0.316%	0.8603±0.1177	0.8385±0.1161	0.8168±0.1483	0.8385±0.1161	0.0212±0.0116	9.5957±32.3809
60	0.100%	**0.8750±0.1216**	**0.8518±0.1181**	**0.8343±0.1501**	**0.8518±0.1181**	0.0218±0.0120	10.0087±36.2662
70	0.032%	0.8654±0.1182	0.8425±0.1142	0.8229±0.1450	0.8425±0.1142	**0.0219±0.012**	10.5043±40.2060
80	0.010%	0.8605±0.1141	0.8384±0.1089	0.8176±0.1403	0.8384±0.1089	0.0226±0.0121	7.3957±28.8814
90	0.003%	0.8606±0.1203	0.8367±0.1180	0.8153±0.1495	0.8367±0.1180	0.0209±0.0118	10.4391±33.7831
100	0.001%	0.8596±0.1220	0.8340±0.1190	0.8132±0.1501	0.8340±0.1190	0.0217±0.0119	8.1348±23.7835

performance. However, when the $SNR < 0$, and the noise level is three or ten times larger than the signal, the signal is nearly masked by the noise, leading to poor performance in event detection, as indicated by higher ADD values shown in Table V.

In summary, it shows that Gaussian-based data augmentation could enhance modeling performance in a certain range of SNR but with varying degrees of improvement for different tasks. Furthermore, SNR analysis provides a useful guide for selecting appropriate levels of augmentation in modeling.

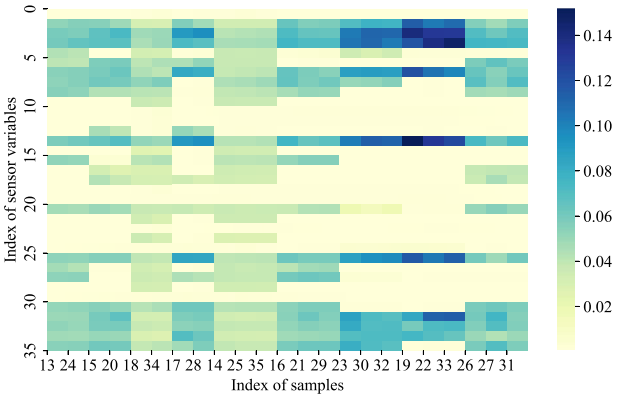

Fig. 8. Sensor variables' weights of a K-fold test.

VI. Conclusion

In this article, we proposed an attention-based MIL framework to identify anomalous behaviors to provide an event-based explanation for an anomalous event of a UAV's engine failure. To address the issue of the lack of time-step labels of the anomalous event, we regarded it as a weakly supervised learning and proposed a special dual attention mechanism, combining temporal pseudo-label and sensor variables' importance, to better pinpoint anomalies from instance-level labels. The introduced two modifications, i.e. D module-local delta feature attention and I module-approximate integral loss, have improved temporal anomaly instances detection and relevant feature identification. Analyzing the results, we observe that the D and I modules in the DA-DI-MIL framework exhibit varying impacts on different tasks. Particularly, the joint modules demonstrate superiority in detecting temporal anomaly instances (Task 2), while not providing substantial advantages over using individual modules in other tasks (Task 1 and Task 3). This suggests that the relationship between the D and I modules is complex and task-dependent, leading to performance variations across different tasks.

What's more, we used a visualization technique to show the identified anomalous behaviors and verify the effectiveness of our proposed method from local and global interpretations. We also use a masking test to state the rationale for identifying the key anomalous behaviors from sensor variables. The results show that sensor variables with high weights have high impacts on anomalous events. DA-DI-MIL achieves nearly 90% in evaluation metrics with no more than 5 time steps' delay of anomalous event detection and around 85% in temporal anomaly detection without point adjust methods. Overall, compared to other baseline methods, our proposed DA-DI-MIL method demonstrates promising results in detecting anomalous events and providing detailed event-based explanations for anomalous events of UAVs.

However, in our pursuit of achieving an event-based explanation in one single model, we ignored the trade-off between interpretability and accuracy in terms of anomalous event detection. Specifically, we tried to enhance the model's explanation both in the time domain and feature space at the expense of event detection's accuracy. In the future, we plan to build an individual explanation model to interpret complex black-box detection models so as to decrease the sensitivity level of the interpretability-accuracy trade-off. Further analysis and investigation of the interactions between the D and I modules are a valuable addition, as it could shed more light on the performance variations observed in the DA-DI-MIL method. Furthermore, the representation of anomalous behaviors to anomalous events could be applied to future event prediction for high-level aviation safety.

References

[1] K. Sheridan, T. G. Puranik, E. Mangortey, O. J. Pinon-Fischer, M. Kirby, and D. N. Mavris, "An application of dbscan clustering for flight anomaly detection during the approach phase," in *Proc. AIAA Scitech Forum*, 2020, Art. no. 1851.

[2] T. G. Puranik and D. N. Mavris, "Identification of instantaneous anomalies in general aviation operations using energy metrics," *J. Aerosp. Inf. Syst.*, vol. 17, no. 1, pp. 51–65, 2020.

[3] K. Guo, L. Liu, S. Shi, D. Liu, and X. Peng, "UAV sensor fault detection using a classifier without negative samples: A local density regulated optimization algorithm," *Sensors*, vol. 19, no. 4, 2019, Art. no. 771.

[4] Y. Ji and H. Lee, "Event-based anomaly detection using a one-class SVM for a hybrid electric vehicle," *IEEE Trans. Veh. Technol.*, vol. 71, no. 6, pp. 6032–6043, Jun. 2022.

[5] L. Zhang, J. Zhao, and W. Li, "Online and unsupervised anomaly detection for streaming data using an array of sliding windows and PDDs," *IEEE Trans. Cybern.*, vol. 51, no. 4, pp. 2284–2289, Apr. 2019.

[6] S. Oehmcke, O. Zielinski, and O. Kramer, "Event detection in marine time series data," in *Proc. Joint German/Austrian Conf. Artif. Intell.*, 2015, pp. 279–286.

[7] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation forest," in *Proc. IEEE 8th Int. Conf. Data Mining*, 2008, pp. 413–422.

[8] G. Pang, C. Shen, L. Cao, and A. V. D. Hengel, "Deep learning for anomaly detection: A review," *ACM Comput. Surv.*, vol. 54, no. 2, pp. 1–38, Mar. 2021, doi: 10.1145/3439950.

[9] K. Hundman, V. Constantinou, C. Laporte, I. Colwell, and T. Soderstrom, "Detecting spacecraft anomalies using LSTMs and nonparametric dynamic thresholding," in *Proc. 24th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2018, pp. 387–395.

[10] B. Wang, D. Liu, Y. Peng, and X. Peng, "Multivariate regression-based fault detection and recovery of UAV flight data," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 6, pp. 3527–3537, Jun. 2020.

[11] J. Zhong, Y. Zhang, J. Wang, C. Luo, and Q. Miao, "Unmanned aerial vehicle flight data anomaly detection and recovery prediction based on spatio-temporal correlation," *IEEE Trans. Rel.*, vol. 71, no. 1, pp. 457–468, Mar. 2022.

[12] A. Deng and B. Hooi, "Graph neural network-based anomaly detection in multivariate time series," in *Proc. AAAI Conf. Artif. Intell.*, 2021, pp. 4027–4035.

[13] H. Liang, L. Song, J. Du, X. Li, and L. Guo, "Consistent anomaly detection and localization of multivariate time series via cross-correlation graph-based encoder–decoder GAN," *IEEE Trans. Instrum. Meas.*, vol. 71, 2021, Art. no. 3504210.

[14] M. Jiang et al., "Weakly supervised anomaly detection: A survey," 2023, *arXiv:2302.04549*.

[15] T. G. Dietterich, R. H. Lathrop, and T. Lozano-Pérez, "Solving the multiple instance problem with axis-parallel rectangles," *Artif. Intell.*, vol. 89, no. 1–2, pp. 31–71, 1997.

[16] A. Tibo, M. Jaeger, and P. Frasconi, "Learning and interpreting multi-multi-instance learning networks," *J. Mach. Learn. Res.*, vol. 21, no. 193, pp. 1–60, 2020.

[17] M.-A. Carbonneau, V. Cheplygina, E. Granger, and G. Gagnon, "Multiple instance learning: A survey of problem characteristics and applications," *Pattern Recognit.*, vol. 77, pp. 329–353, 2018.

[18] V. M. Janakiraman, "Explaining aviation safety incidents using deep temporal multiple instance learning," in *Proc. 24th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2018, pp. 406–415.

[19] M.-H. Bleu Laine, T. G. Puranik, D. N. Mavris, and B. Matthews, "Multi-class multiple-instance learning for predicting precursors to aviation safety events," *J. Aerosp. Inf. Syst.*, vol. 19, no. 1, pp. 22–36, 2022.

[20] H. Wang, H. Zhao, J. Zhang, D. Ma, J. Li, and J. Wei, "Survey on unmanned aerial vehicle networks: A cyber physical system perspective," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 2, pp. 1027–1070, Secondquarter 2020.
[21] A. Keipour, M. Mousaei, and S. Scherer, "Alfa: A dataset for uav fault and anomaly detection," *Int. J. Robot. Res.*, vol. 40, no. 2–3, pp. 515–520, 2021.
[22] Y. Zhou, H. Ren, Z. Li, and W. Pedrycz, "An anomaly detection framework for time series data: An interval-based approach," *Knowl.-Based Syst.*, vol. 228, 2021, Art. no. 107153.
[23] J. F. R. Rochac, L. Liang, N. Zhang, and T. Oladunni, "A Gaussian data augmentation technique on highly dimensional, limited labeled data for multiclass classification using deep learning," in *Proc. IEEE 10th Int. Conf. Intell. Control Inf. Process.*, 2019, pp. 145–151.
[24] D. Roman, S. Saxena, V. Robu, M. Pecht, and D. Flynn, "Machine learning pipeline for battery state-of-health estimation," *Nature Mach. Intell.*, vol. 3, no. 5, pp. 447–456, 2021.
[25] M. Arslan, M. Guzel, M. Demirci, and S. Ozdemir, "SMOTE and Gaussian noise based sensor data augmentation," in *Proc. IEEE 4th Int. Conf. Comput. Sci. Eng.*, 2019, pp. 1–5.
[26] G. R. Mode and K. A. Hoque, "Adversarial examples in deep learning for multivariate time series regression," in *Proc. IEEE Appl. Imagery Pattern Recognit. Workshop*, 2020, pp. 1–10.
[27] D. Xu et al., "Deep multi-instance contrastive learning with dual attention for anomaly precursor detection," in *Proc. SIAM Int. Conf. Data Mining*, 2021, pp. 91–99.
[28] M. Ilse, J. Tomczak, and M. Welling, "Attention-based deep multiple instance learning," in *Proc. 35th Int. Conf. Mach. Learn.*, 2018, pp. 2127–2136.
[29] Y. N. Dauphin, A. Fan, M. Auli, and D. Grangier, "Language modeling with gated convolutional networks," in *Proc. 34th Int. Conf. Mach. Learn.*, 2017, pp. 933–941.
[30] Y. Su, Y. Zhao, C. Niu, R. Liu, W. Sun, and D. Pei, "Robust anomaly detection for multivariate time series through stochastic recurrent neural network," in *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2019, pp. 2828–2837.
[31] R. Zhang, S. Zhang, S. Muthuraman, and J. Jiang, "One class support vector machine for anomaly detection in the communication network performance data," in *Proc. 5th Conf. Appl. Electromagn., Wireless Opt. Commun.*, 2007, pp. 31–37.

Jie Yang received the B.S. degree from Shandong University, Jinan, China, in 2017, and the master's degree in 2020 from Beihang University, Beijing, China, where she is currently working toward the Ph.D. degree with the School of Automation Science and Electrical Engineering. Her research interests include prognostic and health management, explainable anomaly detection, flight data analysis, and reliability.

Diyin Tang (Member, IEEE) received the B.S. and Ph.D. degrees from Beihang University, Beijing, China, in 2008 and 2015, respectively. From 2012 to 2013, she was a Visiting Ph.D. Student with the Department of Mechanical and Industrial Engineering, University of Toronto, Toronto, ON, Canada. She is currently an Associate Professor with the School of Automation Science and Electrical Engineering, Beihang University. Her research interests include fault prognostics, degradation-based modeling, and condition-based maintenance.

Jinsong Yu (Member, IEEE) received the Ph.D. degree from the School of Automation Science and Electrical Engineering, Beihang University, Beijing, China, in 2004. From 2013 to 2014, he was a Visiting Scholar with the Department of Mechanical Engineering, University of Canterbury, Christchurch, Canterbury, New Zealand. He is currently a Professor with the School of Automation Science and Electrical Engineering, Beihang University. His research interests include automatic testing and prognostic and health management.

Jian Zhang received the Ph.D. degree from China Agricultural University, Beijing, China, in 2007. He is currently with the School of Management Science and Engineering, Beijing Information Science and Technology University, Beijing, China. His main research interests include knowledge management and intelligent decision technology.

Haigang Liu is currently a Researcher with Shenyang Aircraft Design and Research Institute, Shenyang, China. His research interests include prognostic and health management and condition-based maintenance.