TAIL-MIL: Time-Aware and Instance-Learnable Multiple Instance Learning for Multivariate Time Series Anomaly Detection

Jaeseok Jang and Hyuk-Yoon Kwon *

Graduate School of Data Science, Seoul National University of Science and Technology
{jangjs1027, hyukyoon.kwon}@seoultech.ac.kr

Abstract

This study addresses the challenge of detecting anomalies in multivariate time series data. Considering a bag (e.g., multi-sensor data) consisting of two-dimensional spaces of time points and multivariate instances (e.g., individual sensors), we aim to detect anomalies at both the bag and instance level with a unified model. To circumvent the practical difficulties of labeling at the instance level in such spaces, we adopt a multiple instance learning (MIL)-based approach, which enables learning at both the bag- and instance- levels using only the bag-level labels. In this study, we introduce time-aware and instance-learnable MIL (simply, TAIL-MIL). We propose two specialized attention mechanisms designed to effectively capture the relationships between different types of instances. We innovatively integrate these attention mechanisms with conjunctive pooling applied to the two-dimensional structure at different levels (i.e., bag- and instance-level), enabling TAIL-MIL to effectively pinpoint both the timing and causative multivariate factors of anomalies. We provide theoretical evidence demonstrating TAIL-MIL's efficacy in detecting instances with two-dimensional structures. Furthermore, we empirically validate the superior performance of TAIL-MIL over the state-of-the-art MIL methods and multivariate time-series anomaly detection methods.

Introduction

Anomaly refers to data that deviates from normal patterns, and abnormal patterns in data typically arise when the relationships between instances in the data deviate from the norm (Chandola, Banerjee, and Kumar 2009). Anomalies can manifest in various forms across domains such as healthcare, finance, smart grids, manufacturing, and education, appearing as diseases, fraud, system malfunctions, operational errors, defects, violence, and more (Wong et al. 2002; Sailusha et al. 2020; Wang, Masoud, and Khojandi 2020; Liebert et al. 2022; Pang et al. 2021; Bermejo Nievas et al. 2011). Due to these diverse manifestations, active research efforts have been undertaken in each domain to detect and prevent anomalies.

This study deals with the problem of detecting anomalies in multivariate time series. In the problem of detecting which variate at which time point caused an anomaly in multivariate time series data, the space can be considered two-dimensional, as illustrated in Fig. 1. Here, the entire dataset is composed of a collection of time-aware bags (TABs), where each bag consists of time-point instances with multivariate (TPIMs). Each TPIM, in turn, is composed of multivariate instances in TPIM (VI-TPIMs). When an anomaly occurs in a TAB, it means that one or more TPIMs within the TAB have anomalies, as illustrated in Figure 1. An anomaly in a TPIM indicates that one or more VI-TPIMs within the TPIM have anomalies. This study aims to design a unified model that can effectively detect anomalies for 1) TABs, 2) TPIMs, and 3) VI-TPIMs. Specifically, instance-level anomaly detection for 2) and 3) has the effect of explaining the cause of the detected anomaly bag in a two-dimensional time-aware data structure. Labeling anomalies, known for their extremely rare occurrence in the overall data, usually requires substantial time and resources from

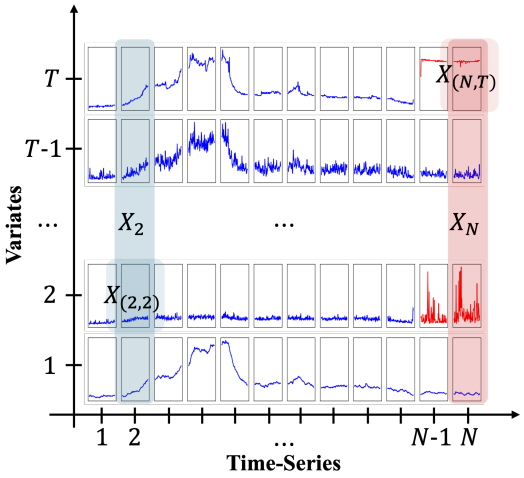

Figure 1: A two-dimensional time-aware bag (TAB) for multivariate time series data (blue: normal data, red: abnormal data). Since all VI-TPIMs, including $X_{(2,2)}$, in a TPIM X_2 are normal, X_2 is normal. On the other hand, because a TPIM X_N includes an abnormal VI-TPIM, $X_{(N,T)}$, X_N is abnormal. Consequently, the TAB including X_N is also considered abnormal.

*Corresponding author

experts. Multiple instance learning (MIL) is a feasible solution to these problems. MIL is a methodology that learns from bag-level labels to predict the labels of individual instances. It is effective when a single bag is composed of multiple instances, and individual labels for each instance are not provided (Javed et al. 2022; Carbonneau et al. 2018). In other words, it learns from the labels of TABs and predicts the labels of TPIMs. Due to these characteristics, MIL has recently been used in the field of time series (Early et al. 2024; Chen et al. 2024), where actual anomalies in a sliding window of data occur at specific time points.

Existing MIL algorithms (Early et al. 2024; Chen et al. 2024) for processing time series data focus on predicting TPIMs based on the labels of TABs. As a result, these algorithms cannot explain which VI-TPIMs are responsible for anomalies within the anomalous TPIMs. This limitation arises because most MIL algorithms, including MIL for time series data, focus on aggregating instances within a single dimension. However, most real-world data are not one-dimensional, i.e., multivariate time series data, as shown in Figure 1. Recently, MIL methodologies supporting multi-dimensional instances have been proposed (Tibo, Jaeger, and Frasconi 2020; Fuster, Eftestøl, and Engan 2022), called *MD-MIL*. MD-MIL can utilize instance-level information, resulting in better performance in instance-level anomaly detection compared to most existing MIL methods. However, existing MD-MIL approaches have a limitation in that they do not account for relationships between instances from different bags. That is, in Figure 1, while VI-TPIMs should ideally be predicted by considering their temporal relationships with VI-TPIMs in different TPIMs, the existing MD-MIL methods fail to leverage these relationships.

To address this issue, this study proposes time-aware and instance-learnable multiple instance learning (TAIL-MIL). We first present two kinds of attention mechanisms, which are specially designed for effectively capturing the relationships between different types of instances. Next, based on the theoretical foundations, we innovatively employ conjunctive-pooling with two kinds of attention mechanisms into two-dimensional instances at different levels (i.e., TPIM and VI-TPIM). Therefore, TAIL-MIL can effectively detect the anomaly occurrence time points and its casual multivariate. We theoretically prove that TAIL-MIL can effectively detect instances with two-dimensional structures. We also demonstrate the superior performance of TAIL-MIL in detecting anomalies in multivariate time series data through experiments compared to existing state-of-the-art MIL methods and multivariate time-series anomaly detection methods in terms of three targets: 1) TABs, 2) TPIMs, and 3) VI-TPIMs. Consequently, by revealing the instance-level aspects of the black-box problem at the bag level, TAIL-MIL is expected to improve the interpretability of the problem, ultimately improving the bag-level prediction performance.

Background

Multiple Instance Learning

The typical MIL is trained using only the labels Y of a given bag $X = \{X_1, X_2, ..., X_{N-1}, X_N\}$ consisting of N instances. It then predicts the labels Y and, concurrently, the labels corresponding to individual instances. Y is considered anomalous if there is at least one anomalous instance, as illustrated in Fig. 1. Conversely, if all instances are normal, it is considered normal. Based on this assumption, the MIL problem is defined as in Problem 1.

Problem 1. ***(Anomaly Detection with MIL)****: The MIL model $g(X)$ generates the output based on the instance features extracted by the encoder $f(X)$. The prediction follows two cases: 1) If any instance prediction $p(\cdot)$ based on $f(X)$ is abnormal, then must be predicted as abnormal. 2) If all instances are predicted as normal, then $g(X)$ must be predicted as normal.*

$$g(X) = \begin{cases} 1 & \textit{if } \exists x \in X \colon (p \circ f)(x) = 1 \\ 0 & \textit{otherwise} \end{cases} \tag{1}$$

Pooling Types in Multiple Instance Learning

To solve Problem 1, MIL models perform aggregation of $f(X)$, $m(\cdot)$, and $p(\cdot)$ to obtain $g(X)$. This process is represented by the aggregation function $Agg(\cdot)$, which combines these components to produce $g(X)$. Based on the pooling techniques used in $Agg(\cdot)$, existing studies can be classified into five categories:

1. **Instance-pooling** (Wang et al. 2018):

$$g(X) = Agg(p_{inst} \circ f(X)) \tag{2}$$

2. **Embedding-pooling** (Wang et al. 2018; Tibo, Jaeger, and Frasconi 2020):

$$g(X) = p_{bag} \circ Agg(f(X)) \tag{3}$$

3. **Attention-pooling** (Ilse, Tomczak, and Welling 2018; Shao et al. 2021; Fuster, Eftestøl, and Engan 2022):

$$g(X) = p_{bag} \circ Agg((m \circ f(X)) \cdot f(X)) \tag{4}$$

4. **Additive-pooling** (Javed et al. 2022):

$$g(X) = Agg(p_{inst}((m \circ f(X)) \cdot f(X))) \tag{5}$$

5. **Conjunctive-pooling** (Angelidis and Lapata 2018; Early et al. 2024):

$$g(X) = Agg((m \circ f(X)) \cdot (p_{inst} \cdot f(X))) \tag{6}$$

MIL Categorization by Instance Dimensions

MIL can be classified into 1-Dimensional MIL (1D-MIL) and Multi-Dimensional MIL (MD-MIL) based on the dimensionality of the instances that make up the data in MIL.

1D-MIL 1D-MIL assumes a single dimension for the instances that make up the bag and performs predictions for the bag by aggregating the predictions or features for all instances at once. Considering a two-dimensional time-aware bag in Figure 1, 1D-MIL needs to predict both TAB and TPIM based on the predictions VI-TPIM. Predicting TPIM through 1D-MIL is feasible by treating a TPIM as a bag and VI-TPIMs in TPIM as instances. However, it is challenging to predict TAB through 1D-MIL because all VI-TPIMs belonging to different TPIMs are considered in the same space without the information of TPIMs to which the VI-TPIMs belong.

MD-MIL MD-MIL is a recent MIL type that performs predictions for the top-level bag by recursively aggregating predictions for instances across multiple dimensions. Considering a two-dimensional time-aware bag, MD-MIL performs MIL pooling for VI-TPMI for each TPMI, making predictions or extracting features for each TPMI. Recursively, MIL pooling is performed again on these results to make predictions for TAB. Unlike 1D-MIL, MD-MIL has the advantage of being able to model the structure of multi-dimensional data in Figure 1, and it is known to achieve better performance for such data (Tibo, Jaeger, and Frasconi 2020; Fuster, Eftestøl, and Engan 2022).

MIL for Multivariate Time Series

The existing MIL algorithms for multivariate time series predict TAB from the learned TPIMs. Specifically, MILLET (Early et al. 2024), which uses conjunctive-pooling, demonstrated anomaly detection performance comparable to supervised learning with excellent interpretability. TimeMIL (Chen et al. 2024), which uses attention-pooling, achieved the best predictive performance compared to state-of-the-art models. These approaches 1) applied 1D-MIL to TPIMs, not considering the level of VI-TPIM, and 2) utilized additional information for representing time series, such as recurrent neural networks (RNNs) or positional encoding.

Motivation

Limitations of Existing MIL-Pooling Types

Existing MIL pooling types have the following limitations in predictions for instances: 1) **Instance-pooling** (Wang et al. 2018) fails to capture the relationships between instances. 2) **Embedding-pooling and attention-pooling** perform anomaly detection by calculating features for the bag through pooling without using predictions for instances. This approach predicts bags based on the contribution of these features, making it prone to overfitting to the decision boundary of the bag and ignoring the decision boundary at the instance level (Raff and Holt 2024). 3) **Additive-pooling** is significantly influenced by attention weights rather than instance features, which can lose original and meaningful instance features (Javed et al. 2022).

These limitations become more pronounced when dealing with multivariate time series data with a two-dimensional structure as follows. 1) Instance-pooling fails to capture the relationships between TPIM and VI-TPIM, leading to significant performance degradation. 2) Embedding-pooling and attention-pooling ignore the two types of instances during the learning process for the bag, leading to performance degradation in instance-level detection. 3) In additive-pooling, the instance feature's contamination in VI-TPIMs affects the predictions on both TPIM and TAB levels in turn.

Limitations of Existing MD-MILs The temporal relationship between VI-TPIMs in different TPIMs should be considered in multivariate time-series data. However, the existing MD-MIL algorithms are limited in that they cannot learn the relationships with instances in other bags. Specifically, Multi-Multi-MIL (Tibo, Jaeger, and Frasconi 2020) assumes that the relationships between all instances in a bag are independent and perform embedding-pooling based on this assumption. In Nested MIL (Fuster, Eftestøl, and Engan 2022), which is an attention-pooling-based MD-MIL, VI-TPIM can only learn the relationships within the same TPIM, but it cannot learn the relationships with VI-TPIMs in other TPIMs.

Limitations of Existing MILs for Multivariate Time Series The MIL algorithms for multivariate time series are as follows: 1) inputting TPIMs into an RNN-based model to extract dependency between them (Angelidis and Lapata 2018), or 2) utilizing additional positional encoding to the instances (Early et al. 2024; Chen et al. 2024). Both approaches have the following limitations. First, they include features extracted from other time points rather than just their own features. This additional information potentially improves the prediction performance for the bag but hinders that for the instance because it contaminates the original instance's features. The theoretical evidence for these points is detailed in Appendix. Second, they have been studied as 1D-MIL types focusing on TPIM. As a result, to perform VI-TPIM prediction, these methods must either 1) acquire additional labels for TPIM to train on VI-TPIM or 2) implement a MIL algorithm that predicts TAB as a bag by mapping VI-TPIMs belonging to different TPIMs to one dimension. The first approach faces the challenge of difficulty in obtaining actual TPIM labels, sacrificing the advantage of MIL. The second approach encounters the limitation of multicollinearity problems when analyzing the relationships between VI-TPIMs belonging to different TPIMs in the same space.

Research Objectives

To overcome the limitations described from the three perspectives mentioned earlier, this study aims to achieve the following research objectives:

- **(R1)** The attention mechanism must be able to consider the relationships between VI-TPIMs in different TPIMs.
- **(R2)** Bag-level prediction must be made using instance-level prediction, enabling instance-level detection and enhancing bag-level detection.
- **(R3)** For predictions for instances, the original features extracted from the instances must be maintained without modification.

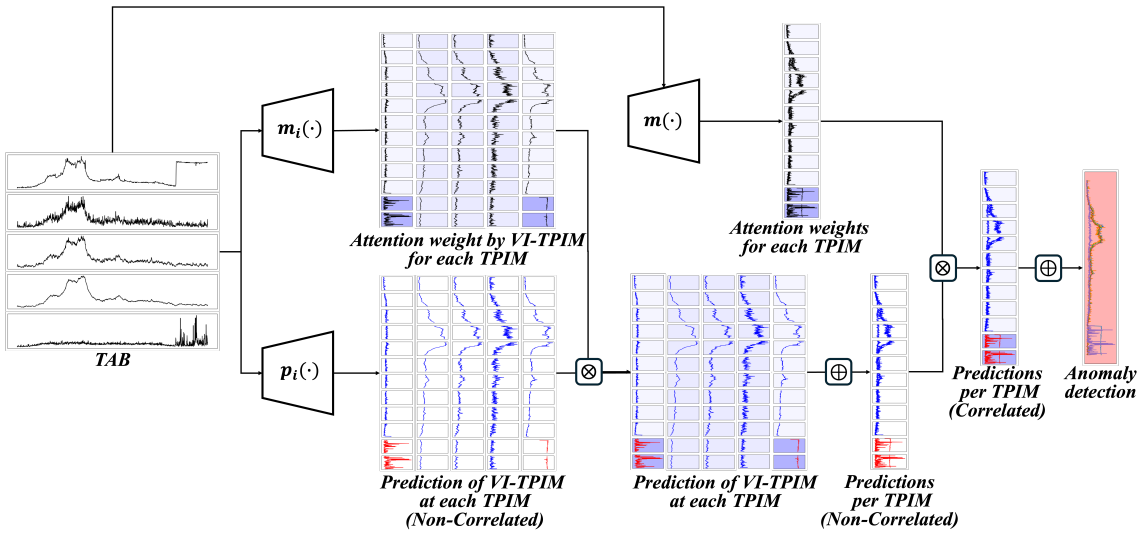

Figure 2: The proposed architecture of TAIL-MIL.

R1 aims to ensure that attention mechanisms reflect the temporal dependencies between VI-TPIMs located in different TPIMs. **R2** focuses on preventing overfitting to the decision boundary of the bag when predictions are made using aggregated features, such as in Embedding or Attention-pooling (Raff and Holt 2024). Lastly, **R3** aims to prevent the use of external information, like RNNs or Positional Encoding, from influencing the prediction of individual instances.

Proposed Method

In this study, we propose Time-Aware Instance Learning with Multiple Instance Learning (TAIL-MIL) to overcome the limitations of existing MIL and effectively learn both anomalous TPIM and VI-TPIM in TAB. TAIL-MIL consists of two main components: 1) two attention mechanisms, each of which adequately considers the relationship between TPIMs in a TAB or VI-TPIMs in a TPIM, and 2) a two-dimensional structure that supports anomaly detection for both TAB and two kinds of instances based on the two attention mechanisms.

Problem Formulation

This study targets the multivariate time-series anomaly detection in two-dimensional spaces (simply, MTAD), where each TAB consists of N TPIMs. Each i-th TPIM consists of T VI-TPIMs: $\{X_{(i,1)}, X_{(i,2)}, ..., X_{(i,T-1)}, X_{(i,T)}\}$. It aims to train on labels Y for the entire dataset X and predict the labels for TPIM, Y, $\{Y_1, Y_2, ..., Y_{N-1}, Y_N\}$, as well as the labels for VI-TPIM, $\{Y_{(i,1)}, Y_{(i,2)}, ..., Y_{(i,T-1)}, Y_{(i,T)}\}$. TAIL-MIL only requires labels Y for the bag level during the learning process (i.e., it does not require labels for individual TPIM and VI-TPIM). Only with the labels for the bag level can our model predict both TPIM and VI-TPIM by learning them while learning bags. MTAD with MIL can be defined as in Problem 2.

Problem 2. ***(MTAD with MIL)****: To address the MTAD problem, the instance-level aggregation module $g_i(X_i)$ and the bag-level aggregation module $g(X)$ in MIL are defined as follows.*
1) For the i^{th} TPIM X_i, if at least one of the predictions $p(X_{(i,j)})$ for the VI-TPIMs $X_{(i,j)}$ within X_i is anomalous, then the prediction $g_i(x_i)$ for the TPIM is considered anomalous; if all are normal, then it is predicted as normal:

$$g_i(X_i) = \begin{cases} 1 & \text{if } \exists x_{(i,j)} \in X_{(i)} : p_i(f_i(x_{(i,j)})) = 1 \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

2) For a TAB X, if at least one of the predictions $g_i(x_i)$ for the TPIMs x_i within X is anomalous, then the prediction $g(X)$ is considered anomalous; if all are normal, then it is predicted as normal:

$$g(X) = \begin{cases} 1 & \text{if } \exists x_i \in X : g_i(x_i) = 1 \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

Time-Aware and Instance-Learnable MIL

In this section, we propose TAIL-MIL to solve MTAD with MIL. To overcome the limitations of existing MIL-based approaches, we design TAIL-MIL to satisfy the requirements outlined in **R1, R2, R3**.

T-Attention: Attention Mechanism for TPIMs in a TAB. To consider the correlation between TPIMs in a TAB for **R1**, we propose an attention mechanism (*T-Attention*) as described in Eq. 9. When the dimension of the features for each instance is L, $W \in \mathbb{R}^{L\times 1}$ and $V \in \mathbb{R}^{L\times N}$ are parameters. Here, $sigm(\cdot)$ and $softm(\cdot)$ denote the sigmoid activation and the softmax activation functions, respectively. W and V generate features of size 1 representing each instance and a scalar feature of size N, respectively. These are then combined with features extracted from other instances, and the attention weights for each instance are computed through

Model	mi-NET	AMIL	MILNET	Additive	MILLET	TimeMIL	Nested	**TAIL-MIL**
Pooling operator	Instance	Attention	Conjunctive	Additive	Conjunctive	Attention	Attention	**Conjunctive**
Input Dimension	1D	1D	1D	1D	1D	1D	2D	**2D**
Position Dependency	None	None	Instance, Attention	None	Instance, Attention	Instance, Attention	None	**Attention**

Table 1: Comparison of MIL models.

the dot-product operation. T-Attention is designed to better reflect the relationships between TPIMs by crossing the features extracted from W and V.

$$m(X) = softm(\frac{sigm(W^\top f(X))^\top sigm(V^\top f(X))^\top}{\sqrt{N}}) \quad (9)$$

V-Attention: Attention Mechanism for VI-TPIMs in a TPIM. To consider the correlation between VI-TPIMs for **R1**, we propose an attention mechanism (*V-Attention*) as described in Eq. 11. Notably, VI-TPIMs at previous TPIMs influence them at subsequent time points. To reflect these characteristics, V-Attention incorporates the features of previous TPIMs into the attention mechanism by using the recurrent features (RF) calculated in Eq. 10, which are used as inputs for Eq. 11:

$$RF_{(i,j)} = \begin{cases} f(X_{(i,j)}) & \text{if } i < n \\ \frac{f(X_{(i,j)}) + \sum_{t=1}^{n-1} RF_{(i-t,j)}}{n} & \text{otherwise} \end{cases} \quad (10)$$

, where n is a configurable variable; in this study, it is set to 2 to calculate the average value.

The final calculation is expressed in Eq. 11. V-Attention m_i computed from VI-TPIMs of i^{th} TPIM uses the number of VI-TPIMs, T, as a scaling factor:

$$m_i(X_i) = softm(\frac{sigm(W^\top RF_{(i,j)})^\top sigm(V^\top f(X_{(i,j)}))^\top}{\sqrt{T}}) \quad (11)$$

TAIL-MIL with T-Attention and V-Attention To satisfy **R1**, we utilize two kinds of attention mechanisms, T-Attention and V-Attention, through $m(X)$ to reflect the relationships among instances at the different levels. To satisfy **R2** and **R3**, TAIL-MIL should directly input the features of an instance $f(X)$ into $p_{inst}(X)$ for prediction. In this study, we adopt conjunctive-pooling to aggregate instances at two different levels (i.e., bag level and instance level) to address these issues while satisfying **R1**, **R2**, and **R3**. The pooling operation process at each level of the MIL module $g(X)$, which aggregates instances from a specific dimension to predict the higher-dimensional bag, is as follows.

$$g(X) = \sum_{i=1}^{N} (m_i(f(X_i))p(f(X_i))) \quad (12)$$

The MIL module $g(X)$ at each level, if there is a higher level, is used again as the $p(X)$ for that higher level.

Figure 2 shows the overall framework for TAIL-MIL.

1. Extract features from the VI-TPIMs in each TPIM constituting each TPIM using $f_i(X)$.
2. Compute the V-attention and prediction for each VI-TPIM individually based on its features via $m_i(\cdot)$ and $p_i(\cdot)$.
3. **(Instance-level prediction)** Perform conjunctive-pooling (Eq. 12) on the predictions and V-attention weights of VI-TPIMs that make up each time point to compute the prediction for the TPIM.
4. Utilize the features of each TPIM to calculate the attention for each TPIM through T-Attention using $m(\cdot)$ (Eq. 9).
5. **(Bag-level prediction)** Aggregating the predictions and attention weights of TPIM to predict for TAB (Eq. 12).

Through this process, TAIL-MIL can perform end-to-end predictions, i.e., TABs, TPIMs, and VI-TPIMs, solely from TAB's labels. TAIL-MIL is optimized for MTAD through its two-dimensional architecture and overcomes the limitations of traditional MIL methods by enabling predictions that reflect only the necessary relationships for each level of instances through T-Attention and V-Attention. Notably, unlike conventional conjunctive-pooling, which only considers the relationships between TPIMs in each TAB, TAIL-MIL applies conjunctive-pooling across two levels, leading to a well-suited model to MTAD.

Explanability of TAIL-MIL Theorem 1 demonstrates that the predictions for instances in a MIL algorithm using conjunctive-pooling contribute to the prediction for the bag.

Theorem 1. *The marginal instance contribution from a conjunctive-pooling-based MIL, $g'(x_i)$, is proportional to the Shapley value of that instance, ϕ_i:*

$$g'(x_i) \propto \phi_i(V, x) \quad (13)$$

***Proof**: The proof of Theorem 1 is provided in Appendix.*

Theorem 2 demonstrates that the predictions for VI-TPIMs by TAIL-MIL contribute not only to the predictions for TPIMs but also to the predictions for TABs.

Theorem 2. *The contribution $g_{sub}(X_{(i,j)})$ of the j-th sub-instance $X_{(i,j)}$ for the i-th instance calculated by TAIL-MIL is proportional to the Shapley value $\phi_{(i,j)}(V, X)$ of $X_{(i,j)}$ with respect to X_i.*

$$g(X_{(i,j)}) \propto \phi_{(i,j)}(V, X) \quad (14)$$

***Consequence**: The contributions of sub-instances in TAIL-MIL are proportional to the Shapley values of sub-instances for data predictions, making them interpretable values for instance-level predictions.*

***Proof**: The proof for Theorem 2 is conducted in Appendix.*

Performance Evaluation

Experimental Setup & Datasets

In this study, we evaluate the performance of TAIL-MIL in the multivariate time-series anomaly detection problem from three aspects: 1) anomaly detection in TABs, 2) anomaly detection in TPIMs, and 3) anomaly detection in VI-TPIMs. Experiments are performed using the following two types of datasets: 1) State of Health (SOH) dataset, 2) energy consumption anomaly detection dataset (5-ECK-2022 (Kim, Jang, and Kwon 2024))

Experimental Datasets Both TAIL-MIL and other MIL methodologies are consistently trained using labels at the TAB level for performance evaluation. 1D-MIL models cannot be directly applied to learning instances with a two-dimensional structure, and existing MIL algorithms for time series have primarily focused on predictions for TPIM. Therefore, to ensure a fair evaluation, this study first compares the performance with 1D-MIL models for time series, assuming TPMI as instances. Additionally, to measure prediction performance for VI-TPMI, 1D-MIL models are further adapted to consider multivariate time series data as single instances at each time point, using the maximum anomaly score of the instances at that time point for predictions. The used performance metrics are AUROC and F1-Score.

Experimental Methods To evaluate the performance of TAIL-MIL in comparison with other MIL models, experiments were conducted with 1D-MIL models such as mi-NET (Wang et al. 2018), Attention MIL (AMIL) (Ilse, Tomczak, and Welling 2018), MILNET (Angelidis and Lapata 2018), Additive MIL (Additive) (Javed et al. 2022), MILLET (Early et al. 2024), and an MD-MIL model, Nested-MIL (Nested) (Fuster, Eftestøl, and Engan 2022), under the same structure of encoders and hyperparameters. The features of each model are presented in Table 1. Additionally, for SOH and 5-ECK-2022 datasets, to measure the upper bound of prediction performance, supervised learning models that directly learn from the instance-level labels were used for comparison. The details regarding the model architectures, hyperparameter settings, and evaluation metrics for each scenario are described in Appendix.

Comparison with Existing MIL Models

Comparison with 1D-MIL for time series The experimental results for the MIL algorithms for time series, including MILNET, MILLET, and TimeMIL, are shown in Table 2. The results demonstrate that TAIL-MIL performed best in both TAB and TPIM predictions. This is because, unlike other MIL algorithms that did not adhere to **R1**, **R2**, and **R3**, TAIL-MIL followed these requirements through a model and computational process tailored to the structure of TAB.

	TABs		TPIMs	
	AUROC	F1-Score	AUROC	F1-Score
MILNET	0.9533	0.9721	0.9029	0.9120
MILLET	0.9439	0.9698	0.9225	0.9321
TimeMIL	0.9563	0.9832	0.7623	0.7312
TAIL-MIL	**0.9574**	**0.9844**	**0.9539**	**0.9812**

Table 2: Performance comparison with 1D-MIL for time series (SOH dataset)

		SOH		5-ECK-2022	
		AUROC	F1-Score	AUROC	F1-Score
1D	mi-Net	0.5000	0.5282	0.5000	0.9260
	AMIL	0.9155	0.9783	0.8811	0.8651
	MILNET	0.9463	0.9824	0.9038	0.8935
	Additive	0.8982	0.9678	0.7805	0.7849
	MILLET	0.8982	0.9653	0.8713	0.8636
	TimeMIL	0.9321	0.9782	0.8926	0.8951
2D	Nested	0.9513	0.9811	0.8651	0.8965
	TAIL-MIL	**0.9574**	**0.9844**	**0.9709**	**0.9529**
	Superivsed	0.9570	0.9825	0.9374	0.9374

Table 3: Abnormal TABs Detection Performance Measurement Results

Experimental Results of Anomaly Detection for TABs The prediction results for TAB by MIL algorithms designed to predict both TPIM and VI-TPIM are shown in Table 3. The experimental results indicate that TAIL-MIL achieved the best anomaly TAB prediction performance across both datasets. Additionally, MIL algorithms that use attention mechanisms to analyze relationships between instances, such as MILNET, TimeMIL, and Nested MIL, performed well. In contrast, models like mi-Net, which do not analyze these relationships, failed to make accurate predictions.

Experimental Results of Anomaly Detection for TPIMs & VI-TPIMs The prediction performance of MIL algorithms designed to predict both TPIM and VI-TPIM is presented in Table 4 and Table 5, respectively. The main findings are as follows: 1) TAIL-MIL consistently demonstrated superior performance across all instances. 2) MD-MIL models outperformed 1D-MIL models in all aspects of performance. Specifically, TAIL-MIL showed performance comparable to supervised learning models trained with labels for all instances. 3) In this dataset, 1D-MIL performed poorly in instance prediction due to increased multicollinearity

		SOH		5-ECK-2022	
		AUROC	F1-Score	AUROC	F1-Score
1D	mi-Net	0.5022	0.8794	0.5000	0.3012
	AMIL	0.5156	0.8814	0.6179	0.3617
	MILNET	0.5585	0.8895	0.5187	0.2204
	Additive	0.6650	0.9031	0.5007	0.3015
	MILLET	0.5912	0.8872	0.5007	0.3015
	TimeMIL	0.5820	0.9137	0.6249	0.3472
2D	Nested	0.9513	0.9808	0.7891	0.7078
	TAIL-MIL	**0.9539**	**0.9812**	**0.9075**	**0.8978**
	Superivsed	0.9620	0.9849	0.9199	0.9041

Table 4: Abnormal TPIMs Detection Performance Measurement Results

		SOH		5-ECK-2022	
		AUROC	F1-Score	AUROC	F1-Score
1D	mi-Net	0.6496	0.6872	0.5000	0.1299
	AMIL	0.5307	0.5705	0.6215	0.3015
	MILNET	0.5643	0.5905	0.5164	0.2051
	Additive	0.5876	0.5947	0.5010	0.2446
	MILLET	0.5648	0.4626	0.5071	0.2327
	TimeMIL	0.5831	0.5319	0.5739	0.2759
2D	Nested	0.7272	0.5708	0.7386	0.7078
	TAIL-MIL	**0.9290**	**0.9454**	**0.9458**	**0.9204**
	Superivsed	0.9310	0.9472	0.9283	0.9171

Table 5: Abnormal VI-TPIMs Detection Performance Measurement Results

while reflecting relationships among all instances in different TPIMs within a single dimension.

To demonstrate the effectiveness of TAIL-MIL's two-dimensional structure and validate its efficacy, we conducted an ablation study. The results showed that all modules proposed in this study contributed to performance improvement, as detailed in Appendix.

Conclusions

In this study, we proposed TAIL-MIL as the first MIL-based model that supports time-aware two-dimensional data instances, overcoming the limitations of existing models. TAIL-MIL aims to improve the performance of anomaly detection for the respective TAB, TPIM, and VI-TPIM levels, providing explainability in the anomaly detection results. The experiments showed that our TAIL-MIL outperformed the existing MIL-based models and the SOTA models for multivariate time-series anomaly detection in all the anomaly detection levels. These findings indicate reducing the need for human labeling at the instance level while learning both instance and bag levels from only the bag-level labels.

Acknowledgements

This work was supported in part by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. 2022R1F1A1067008), and in part by the Ministry of Trade, Industry and Energy(MOTIE) and Korea Institute for Advancement of Technology(KIAT) through the International Cooperative R&D program (P0028216).

References

Angelidis, S.; and Lapata, M. 2018. Multiple instance learning networks for fine-grained sentiment analysis. *Transactions of the Association for Computational Linguistics*, 6: 17–31.

Bermejo Nievas, E.; Deniz Suarez, O.; Bueno García, G.; and Sukthankar, R. 2011. Violence detection in video using computer vision techniques. In *Computer Analysis of Images and Patterns: 14th International Conference, CAIP 2011, Seville, Spain, August 29-31, 2011, Proceedings, Part II 14*, 332–339. Springer.

Carbonneau, M.-A.; Cheplygina, V.; Granger, E.; and Gagnon, G. 2018. Multiple instance learning: A survey of problem characteristics and applications. *Pattern Recognition*, 77: 329–353.

Chandola, V.; Banerjee, A.; and Kumar, V. 2009. Anomaly detection: A survey. *ACM computing surveys (CSUR)*, 41(3): 1–58.

Chen, X.; Qiu, P.; Zhu, W.; Li, H.; Wang, H.; Sotiras, A.; Wang, Y.; and Razi, A. 2024. TimeMIL: Advancing Multivariate Time Series Classification via a Time-aware Multiple Instance Learning. In *Forty-first International Conference on Machine Learning*.

Early, J.; Cheung, G.; Cutajar, K.; Xie, H.; Kandola, J.; and Twomey, N. 2024. Inherently Interpretable Time Series Classification via Multiple Instance Learning. In *The Twelfth International Conference on Learning Representations*.

Fuster, S.; Eftestøl, T.; and Engan, K. 2022. Nested multiple instance learning with attention mechanisms. In *2022 21st IEEE International Conference on Machine Learning and Applications (ICMLA)*, 220–225. IEEE.

Ilse, M.; Tomczak, J.; and Welling, M. 2018. Attention-based deep multiple instance learning. In *International conference on machine learning*, 2127–2136. PMLR.

Javed, S. A.; Juyal, D.; Padigela, H.; Taylor-Weiner, A.; Yu, L.; and Prakash, A. 2022. Additive MIL: intrinsically interpretable multiple instance learning for pathology. *Advances in Neural Information Processing Systems*, 35: 20689–20702.

Kim, T.; Jang, J.-S.; and Kwon, H.-Y. 2024. Correlation-driven multi-level learning for anomaly detection on multiple energy sources. *Applied Soft Computing*, 111636.

Liebert, A.; Weber, W.; Reif, S.; Zimmering, B.; and Niggemann, O. 2022. Anomaly Detection with Autoencoders as a Tool for Detecting Sensor Malfunctions. In *2022 IEEE 5th International Conference on Industrial Cyber-Physical Systems (ICPS)*, 01–08. IEEE.

Pang, G.; Ding, C.; Shen, C.; and Hengel, A. v. d. 2021. Explainable deep few-shot anomaly detection with deviation networks. *arXiv preprint arXiv:2108.00462*.

Raff, E.; and Holt, J. 2024. Reproducibility in Multiple Instance Learning: A Case For Algorithmic Unit Tests. *Advances in Neural Information Processing Systems*, 36.

Sailusha, R.; Gnaneswar, V.; Ramesh, R.; and Rao, G. R. 2020. Credit card fraud detection using machine learning. In *2020 4th international conference on intelligent computing and control systems (ICICCS)*, 1264–1270. IEEE.

Shao, Z.; Bian, H.; Chen, Y.; Wang, Y.; Zhang, J.; Ji, X.; et al. 2021. Transmil: Transformer based correlated multiple instance learning for whole slide image classification. *Advances in neural information processing systems*, 34: 2136–2147.

Tibo, A.; Jaeger, M.; and Frasconi, P. 2020. Learning and interpreting multi-multi-instance learning networks. *Journal of Machine Learning Research*, 21(193): 1–60.

Wang, X.; Yan, Y.; Tang, P.; Bai, X.; and Liu, W. 2018. Revisiting multiple instance neural networks. *Pattern Recognition*, 74: 15–24.

Wang, Y.; Masoud, N.; and Khojandi, A. 2020. Real-time sensor anomaly detection and recovery in connected automated vehicle sensors. *IEEE transactions on intelligent transportation systems*, 22(3): 1411–1421.

Wong, W.-K.; Moore, A.; Cooper, G.; and Wagner, M. 2002. Rule-based anomaly pattern detection for detecting disease outbreaks. In *AAAI/IAAI*, 217–223.