Reinforcement Learning for Time Series Data with Partially Labeled Anomalies

Kio Yun
Department of Industrial and Management Engineering
Korea University
Seoul, South Korea
ykio@korea.ac.kr

Jun-Geol Baek*
Department of Industrial and Management Engineering
Korea University
Seoul, South Korea
jungeol@korea.ac.kr

***Abstract*— In recent manufacturing processes, the escalation of sensors usage has led to the continuous collection of large volumes of data. This trends present the challenge of extracting significant patterns and rapidly identifying anomalies within this extensive time-series data. A major obstacle in this area is the scarcity of labeled data for anomaly detection, which limits the effective training of machine learning models. Existing research has mainly concentrated on utilizing the limited anomaly data with supervised learning or investigating unsupervised learning method. This study employs deep reinforcement learning to simultaneously utilize both the abundant unlabeled data and the minimal labeled data for anomaly detection. This paper aims to learn established anomaly patterns and actively explore potential anomalies in unlabeled data, thereby covering both known and undiscovered anomaly patterns. Experiments on multivariate time-series datasets have shown proposed method to outperform existing models in similar situations. The findings of this research are expected to significantly advance effective anomaly detection in manufacturing environments, particularly in contexts where labeled anomaly data is limited.**

Keywords—Anomaly Detection, Deep Learning, Reinforcement Learning, Time Series, Semi-supervised learning settings

I. INTRODUCTION

Anomaly detection aims to identify data objects or behaviors that significantly deviate from the majority of data patterns. It is an essential practice in various domains, including financial fraud detection, cybersecurity attack detection, medical diagnostics, and manufacturing processes. In manufacturing, particularly, large-scale time-series data that changes over time is collected from numerous sensors. Identifying subtle anomalies in this data is critical, as they can lead to product defects or reduced process efficiency, making the technology for quickly and accurately detecting anomaly patterns in time-series data important for enhancing manufacturing quality and reducing costs. However, detecting accurate anomaly patterns in time-series data is a significant challenge due to the complex anomaly patterns that can be spatial or temporal depending on the context. And this data often exhibits periodicity or seasonality, adding to the complexity of accurate anomaly [1].

Moreover, these anomaly patterns can originate from various causes, resulting in different types of anomaly patterns with distinct characteristics. For example, in manufacturing processes, anomalies can exhibit entirely different patterns depending on their cause. Additionally, since anomalies occur infrequently and unpredictably, it is challenging to obtain labeled training data for all types of anomaly patterns. For these reasons, anomaly detection research has primarily focused on unsupervised learning. However, in practice, there often exists a small set of known anomaly data for significant patterns. Despite their small size, this data can provide crucial prior knowledge, and effectively utilizing it can lead to substantial performance improvements over unsupervised learning [2]. Therefore, a method is needed that can leverage this limited anomaly data without assuming it covers all types of anomaly patterns [3]. In this study, we propose a methodology capable of effectively learning both known and unknown types of anomalies, based on a combination of large-scale unlabeled data, predominantly composed of normal data, and a small set of labeled anomaly data that includes only some types of anomalies.
Related works

A. Time-Series Representation Learning

Research on representation learning for time series is actively ongoing. Some approaches utilize pretext tasks to learn the representation of time series data. [8] learns the representation of time series by applying transformations to the original time series and then performing binary classification between the original and transformed data to recognize human behavior. [9] aimed to learn the representation of ECG data by applying six types of transformations to the original data and then classifying which transformation was used. Additionally, research leveraging contrastive learning for time series representation learning has been gaining attention, inspired by the success of contrastive learning. In [10], CPC achieved significant results in speech recognition by predicting the future in latent space. [11] expanded the application of the well-known contrastive learning methodology, SimCLR [12], to EEG data. The study most relevant to our research is TS2Vec [13], which proposed a methodology using contrastive learning to effectively capture the contextual information of time series.

B. Reinforcement Learning

Deep reinforcement learning has demonstrated human-level capabilities in various domains, including gaming [14]. Based on these successes, recent efforts have been made to apply deep reinforcement learning to solve real-world problems. The

*Corresponding author—Tel: +82-2-3290-3396; Fax: +82-2-3290-4550

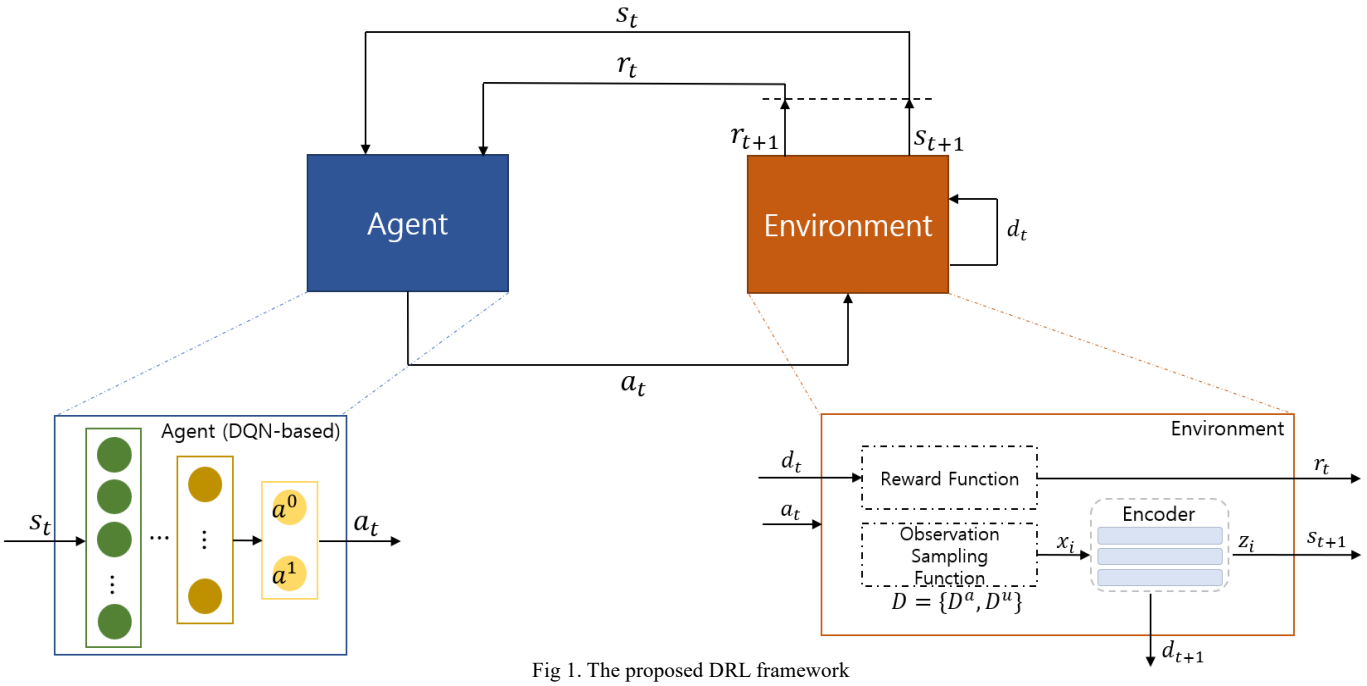

Fig 1. The proposed DRL framework

research most relevant to our study is [3], which proposed an approach named DPLAN (Deep Q-learning with Partially Labeled ANomalies). DPLAN utilizes reinforcement learning to learn anomalies in data that partially contains labels. However, DPLAN is limited to data in tabular form and cannot accommodate the temporal and spatial characteristics of sensor data collected in time-series format. Furthermore, DPLAN shows significant differences from our methodology in aspects such as the reward function for learning anomaly from unlabeled data and the observation sampling function.

II. Proposed Method

A. Time-Series Representation Learning

When given a time series $X = \{x_1, x_2, \dots, x_N\}$ in D, composed of N instances, each x_i is mapped to its most representative form z_i through a nonlinear embedding function. The inputted time series x_i is composed of dimensions $T \times F$, where T denotes the length of the time series and F represents the feature dimensions. The representation $z_i = \{z_{i,1}, z_{i,2}, \dots, z_{i,T}\}$ includes a representation vector $z_{i,t} \in \mathbb{R}^K$ at each time point t, where K is the dimension of the representation vector.

For this representation of the time series, the nonlinear embedding function utilizes the framework proposed in TS2Vec [13]. This allows for measuring the extent of anomalies in unlabeled time series using representations that accurately reflect the characteristics of the original time series data. These representations can then be used as inputs for a reinforcement learning agent.

B. Reinforcement Learning For Anomaly Detection

Inspired by DPLAN [3], this study proposes a deep reinforcement learning approach for anomaly detection targeting time-series data. This approach aims to detect both known and unknown types of anomalies. The deep reinforcement learning approach comprises an anomaly detection agent, environment, action space, and reward system. The key is for the anomaly detection agent to explore information about known anomaly types from the labeled anomaly data D^a and simultaneously learn known and unknown anomalies in the unlabeled data D^u.

III. Reinforcement Learning for Anomaly Detection

A. Observation Sampling Function g

E refers to the collection of the agent's learning experiences, denoted as $e_t = (s_t, a_t, r_t, s_{t+1})$. The loss is computed using mini-batches that are uniformly and randomly sampled from these stored experiences. g_u samples s_{t+1} from D^u based on the action taken by the agent at the current observation. Specifically, g_u is defined as follows:

$$g_u(s_{t+1}|s_t, z_i, a_t; \theta^e) = \begin{cases} z_{i+1} & \text{if } a_t = a^1 \\ \underset{s \in S}{\operatorname{argmax}}\, d(s_t, s; \theta^e) & \text{if } a_t = a^0 \end{cases} \quad (1)$$

$S \in D^u$ is a random subsample, and θ^e represents the parameters of the TS2Vec encoder. The term d refers to the Euclidean distance between s_t and s_{t+1}, indicating the distance in the representation space of time series as processed by the TS2Vec encoder. Specifically, if agent A classifies the current observation s_t as an anomaly and takes action a^1, g_u returns the next time series window in the sequence. This enables A to actively search for data in the unlabeled set D^u that resembles suspected anomalies. If A deems the current data as normal and takes action a^0, g_u returns the data furthest away in the time series representation space. This approach leads A to explore

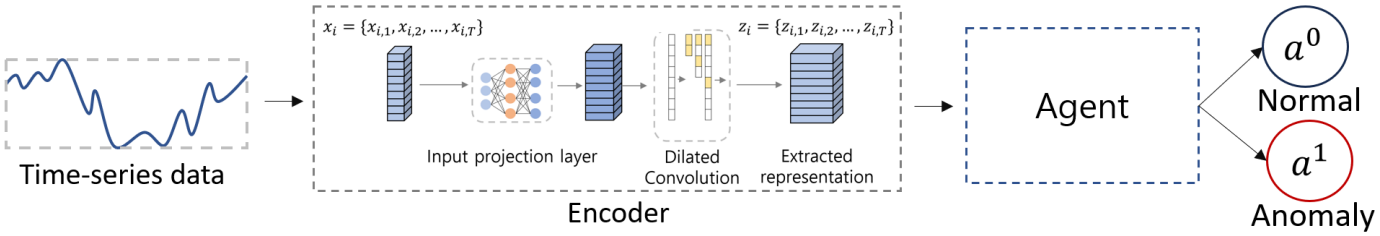

Fig 2. Inference phase of proposed method

potential anomalies that are most distant from the current normal observation. In both scenarios, A is encouraged to actively seek anomalies in the large set D^u. Otherwise, g_a samples randomly from the labeled anomaly dataset D^a.

During the interaction between the agent and the environment, both g_a and g_u are used. In this study, g_a and g_u are selected with equal probability. This ensures that the agent sufficiently explores the small labeled anomaly data while also exploring the unlabeled data.

B. Reward Function

The extrinsic reward function h is defined to provide a reward signal r_t^e based on the known anomaly data and is the same as defined in DPLAN.

$$r_t^e = h(s_t, a_t) = \begin{cases} 1 & \text{if } a_t = a^1 \text{ and } s_t \in D^a \\ 0 & \text{if } a_t = a^0 \text{ and } s_t \in D^u \\ -1 & \text{otherwise} \end{cases} \quad (2)$$

According to the extrinsic reward function h, the agent receives a positive reward only when it correctly identifies the data as an anomaly during its interaction with the known anomaly data. It incurs a penalty in cases of false positives or false negatives. Thus, r^e encourages the agent to actively explore the labeled data D^a.

To maximize rewards, the agent learns to recognize known anomalies, accurately detecting known anomaly data while simultaneously avoiding false negatives and false positives. This approach enables proposed method to utilize known information more effectively than traditional semi-supervised learning methodologies [7], thereby enhancing its capability in anomaly detection.

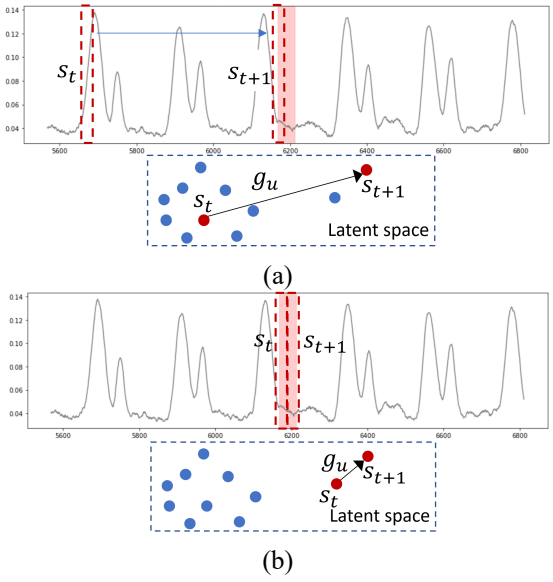

Fig 3. Operation of the observation sampling function g_u

Unlike r^e, which encourages the agent to explore the labeled data D^a, the intrinsic reward r^i motivates the agent to actively explore the unlabeled data D^u.

$$r_t^i = f(s_t; \theta^e) = d_{\text{cosine}}(z_i^u, z_i^m) \quad (3)$$

The function f assesses the anomaly of s_t by using the cosine distance between the original latent representation z_i^u of the current observation s_t and its masked version at the end point z_i^m as the anomaly score. TS2Vec learns the contextual consistency of instances at the same time point, allowing the detection of anomalies through the degree of discrepancy between representations extracted from original and masked data [13]. Cosine distance, commonly used in contrastive learning to measure similarity between two samples is employed for this purpose [12]. Additionally, since cosine distance ranges from 0 to 1, higher values of r_i indicating greater anomaly. Therefore, regardless of the agent's action, higher anomaly in the observation results in higher rewards for r_i, leading the agent to actively search for potential anomalies in D^u. To balance exploration and exploitation, the reward at time t, r_t, is defined as follows.

$$r_t = r_t^e + r_t^i \quad (4)$$

IV. Experiments

A. Datasets

In this study, three datasets were used to evaluate the anomaly detection performance of the methodology and its individual components. Each dataset is labeled, comprising training data in normal state and test data containing anomalies. Detailed descriptions of each dataset are as follows.

TABLE I. Summary of the Dataset Used in the Experiment

	SWaT	MSL	SMAP
# of datasets	1	27	55
Variables	52	55	25
% of anomalies	12.14	10.48	12.82
Training data points	495000	58317	138004
Testing data points	449919	73729	435826

- **Water Treatment (SWaT).** This dataset contains water treatment data collected over 11 days. The data from the first 7 days represent normal conditions, while the last 4 days include 36 different types of attacks, each with unique characteristics.
- **Mars Science Laboratory (MSL).** This dataset comprises data collected from NASA's Curiosity Rover on Mars. It consists of multivariate time series data in 55 dimensions, including 27 remote sensing signals.
- **Soil Moisture Active Passive dataset (SMAP).** Similar in format to the MSL dataset, SMAP includes data received from the Soil Moisture Active Passive satellite. It consists of 53 signal data and is structured similarly to the MSL data.

B. Evaluation Metrics

The evaluation metrics used were precision, recall, and the F1-score. Point Adjustment, as proposed by [15] was applied. Point Adjustment considers the entire anomaly interval as detected if only a part of it is identified. This concept is based on the practical viewpoint that detecting a single point within an anomaly interval is effectively equivalent to detecting the entire anomaly. However, there is a concern that this approach may lead to an overestimation of detection performance [16]. Therefore, this study presents results both with and without the application of Point Adjustment.

C. Comparisn Models

In this paper, the proposed methodology was compared with six widely known unsupervised learning models. iForest [17] and OC-SVM [18] were selected as baseline models. Deep learning models like USAD [19], MemAE [20], DAGMM [21] and TS2Vec [13] used in this research were also compared. Notably, TS2Vec, the time series representation extraction method used in this study, was compared with the anomaly detection technique used in the paper proposing TS2Vec. The comparison aimed to demonstrate the effectiveness of the deep reinforcement learning-based anomaly detection model proposed in this research, especially when using the same representation learning model as TS2Vec.

D. Experimental Results.

Firstly, to evaluate the proposed method, comparative experiments were conducted with other models. All experiments were repeated five times to measure average performance. Models other than the proposed methodology were trained on training data with all labels removed and then evaluated on the test data. The proposed methodology assumes a scenario with a large amount of unlabeled data and a small amount of labeled anomaly data. Therefore, the data was restructured to use a part of the anomaly data as D^a and the rest, along with normal data, as D^u with labels removed. The percentage alongside each model name indicates the proportion of anomaly data used for training. For example, proposed method (5%) means that 5% of all anomaly data were labeled and used as D^a, and the rest were used as D^u without labels.

Tables 2 and 3 show the performance of the comparison models and the proposed methodology on three benchmark datasets. Table 2 presents the performance with Point Adjustment applied, while Table 3 shows the performance without Point Adjustment. The results indicate that proposed method performed excellently across all datasets. Although some models had slightly higher recall than proposed method, their Precision was significantly lower, leading to proposed method having a much higher F1-Score. Additionally, the methodology used TS2Vec for extracting time series representations, and its performance was compared with TS2Vec without the application of reinforcement learning. The proposed methodology showed significantly higher performance compared to TS2Vec. This demonstrates the effectiveness of utilizing reinforcement learning to enhance

TABLE II. EXPERIMENTAL RESULTS WITH COMPARATIVE MODELS (WITHOUT POINT ADJUSTMENT)

Model	MSL			SMAP			SWaT		
	Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
ARIMA	0.28	**0.80**	0.41	0.17	0.82	0.28	0.13	**0.99**	0.23
IForest	0.18	0.16	0.17	0.10	0.04	0.08	0.23	0.83	0.36
MemAE	0.13	0.40	0.20	0.19	0.46	0.27	0.34	0.72	0.46
USAD	0.15	0.57	0.24	0.18	0.49	0.26	0.34	0.72	0.46
DAGMM	0.12	0.19	0.12	0.11	0.17	0.10	0.43	0.71	0.54
TS2Vec	0.23	0.62	0.34	0.47	0.76	0.58	0.37	0.88	0.52
Proposed method (5%)	**0.61**	0.78	0.68	0.56	0.85	0.68	0.84	0.89	0.86
Proposed method (10%)	0.60	0.81	**0.69**	**0.60**	**0.89**	**0.72**	**0.87**	0.91	**0.89**

TABLE III. EXPERIMENTAL RESULTS WITH COMPARATIVE MODELS (WITH POINT ADJUSTMENT)

Model	MSL			SMAP			SWaT		
	Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
ARIMA	0.31	**1**	0.47	0.18	0.96	0.30	0.13	**1**	0.23
IForest	0.47	0.66	0.55	0.34	0.40	0.36	0.26	0.97	0.40
MemAE	0.17	0.91	0.29	0.21	0.89	0.34	0.40	0.72	0.51
USAD	0.22	0.99	0.36	0.26	0.95	0.41	0.32	0.89	0.47
DAGMM	0.20	0.44	0.25	0.16	0.41	0.19	0.49	0.90	0.64
TS2Vec	0.27	0.90	0.42	0.51	0.94	0.66	0.42	0.90	0.57
Proposed method (5%)	0.60	0.96	0.74	0.61	0.94	0.74	0.86	0.93	0.89
Proposed method (10%)	**0.65**	0.99	**0.78**	**0.63**	**0.96**	**0.76**	**0.89**	0.97	**0.93**

performance by effectively using the information from known anomaly data. Moreover, the use of 10% of the anomaly data as labeled data showed better performance than using only 5%. This suggests that the more and diverse information available about anomalies, the better the detection performance of the model.

V. Conclusion

In this research, the focus was on detecting known and unknown anomaly patterns within large-scale time-series data in manufacturing processes. To address this, the study proposed a deep reinforcement learning framework which utilizes partially labeled anomaly data. This model demonstrated the ability to learn known anomaly patterns while actively searching for potential anomalies in unlabeled data. Additionally, it addressed the complexity of data by applying a representation learning methodology to reflect the spatial and temporal characteristics of time-series data.

The experimental results under various conditions showed that proposed method outperformed existing methodologies, especially in anomaly pattern detection in situations with limited labeled data. This finding is significant, as it demonstrates the considerable advantages of proposed method in practical applications within manufacturing processes. The results are expected to provide substantial value in real-world industrial applications, enhancing the efficiency and accuracy of anomaly detection in complex manufacturing environments.

ACKNOWLEDGMENT

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (NRF-2022R1A2C2004457). This work was also supported by Samsung Electronics Co., Ltd. (IO201210-07929-01) and Brain Korea 21 FOUR.

References

[1] M. Yu and S. Sun, "Policy-based reinforcement learning for time series Anomaly detection," *Engineering Applications of Artificial Intelligence*, vol. 95, p. 103919, 2020.

[2] G. Pang, C. Shen, H. Jin, and A. van den Hengel, "Deep weakly-supervised anomaly detection," *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2023.

[3] G. Pang, Anton, C. Shen, and L. Cao, "Toward Deep Supervised Anomaly Detection," *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2021.

[4] T. Schlegl, P. Seeböck, S. M. Waldstein, U. Schmidt-Erfurth, and G. Langs, "Unsupervised Anomaly Detection with Generative Adversarial Networks to Guide Marker Discovery," *Lecture Notes in Computer Science*, pp. 146–157, 2017.

[5] C. Zhou and R. C. Paffenroth, "Anomaly Detection with Robust Deep Autoencoders," *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Aug. 2017.

[6] G. Pang, L. Cao, L. Chen, and H. Liu, "Learning Representations of Ultrahigh-dimensional Data for Random Distance-based Outlier Detection," *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining - KDD '18*, 2018.

[7] L. Ruff *et al.*, "Deep Semi-Supervised Anomaly Detection," *arXiv.org*, Feb. 14, 2020. https://arxiv.org/abs/1906.02694 (accessed Oct. 25, 2023).

[8] A. Saeed, T. Ozcelebi, and J. Lukkien, "Multi-task Self-Supervised Learning for Human Activity Detection," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 3, no. 2, pp. 1–30, Jun. 2019.

[9] P. Sarkar and A. Etemad, "Self-supervised ECG Representation Learning for Emotion Recognition," *IEEE Transactions on Affective Computing*, pp. 1–1, 2020.

[10] A. Oord, Y. Li, and O. Vinyals, "Representation Learning with Contrastive Predictive Coding," *arXiv (Cornell University)*, Jul. 2018.

[11] M. Mohsenvand, M. Izadi, and P. Maes, "Contrastive Representation Learning for Electroencephalogram Classification," *Machine Learning for Health*, pp. 238–253, Nov. 2020.

[12] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," *International conference on machine learning* , vol. 119, pp. 1597–1607, Jul. 2020.

[13] Z. Yue *et al.*, "TS2Vec: Towards Universal Representation of Time Series," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, no. 8, pp. 8980–8987, Jun. 2022.

[14] V. Mnih *et al.*, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, Feb. 2015.

[15] H. Xu *et al.*, "Unsupervised Anomaly Detection via Variational Auto-Encoder for Seasonal KPIs in Web Applications," *Proceedings of the 2018 World Wide Web Conference on World Wide Web - WWW '18*, pp. 187–196, 2018.

[16] S. Kim, K. Choi, H.-S. Choi, B. Lee, and S. Yoon, "Towards a Rigorous Evaluation of Time-Series Anomaly Detection," *Proceedings of the ... AAAI Conference on Artificial Intelligence*, vol. 36, no. 7, pp. 7194–7201, Jun. 2022.

[17] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation Forest," *2008 Eighth IEEE International Conference on Data Mining*, Dec. 2008.

[18] Bernhard Schölkopf, R. C. Williamson, A. Smola, J. Shawe-Taylor, and J. Platt, "Support Vector Method for Novelty Detection," *Advances in neural information processing systems*, vol. 12, pp. 582–588, Nov. 1999.

[19] A. A. Cook, G. Mısırlı, and Z. Fan, "Anomaly Detection for IoT Time-Series Data: A Survey," *IEEE Internet of Things Journal*, vol. 7, no. 7, pp. 6481–6494, Jul. 2020.

[20] D. Gong *et al.*, "Memorizing Normality to Detect Anomaly: Memory-Augmented Deep Autoencoder for Unsupervised Anomaly Detection," *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, Oct. 2019.

[21] B. Zong *et al.*, "Deep Autoencoding Gaussian Mixture Model for Unsupervised Anomaly Detection," *International conference on learning representations*, Feb. 2018.