2025 International Conference on Intelligent Computing
July 26-29, Ningbo, China
https://www.ic-icc.cn/2025/index.php

Multivariate Time Series Anomaly Detection Model Selection based on Rank Aggregation of Performance Metrics

Mingyu Liu[1], Yijie Wang[1] (✉), Xiaohui Zhou[1], and Yongjun Wang[2]

[1] National Key Laboratory of Parallel and Distributed Computing, College of Computer Science and Technology, National University of Defense Technology, China
[2] College of Computer Science and Technology,
National University of Defense Technology, China
{liumingyu13, wangyijie, zxh17, wangyongjun}@nudt.edu.cn

Abstract. Multivariate time series anomaly detection (MTAD) has significant practical relevance in various applications. Despite the recent proposals of numerous MTAD models, none have demonstrated consistent optimal performance across various scenarios. Hence, there is an urgent need to investigate the accurate selection of the most appropriate MTAD model for a specific dataset. Most studies on model selection depend on extensive pre-trained models. Nevertheless, in real-world situations, labels for time series anomaly data are seldom accessible, and the training cost of pre-trained models is significant. This paper presents an unsupervised method for selecting multivariate time series anomaly detection model based on rank aggregation of performance metrics. We create a reliable performance ranking by aggregating rankings from various unsupervised evaluation metrics. Subsequently, an early-stopping mechanism is applied to minimize computational expenses by identifying the Top-K models that consistently maintain their ranking in robust performance throughout the epochs. Extensive experiments on six real-world datasets demonstrates that our proposed unsupervised model selection method is comparably effective to the supervised method in selecting the optimal MTAD model.

Keywords: Multivariate Time Series, Anomaly Detection, Model Selection, Rank Aggregation.

1 Introduction

The rapid advancement of automated systems in practical applications has resulted in the generation of extensive volumes of time-varying data from diverse sensors, leading to the creation of multivariate time series data. **M**ultivariate **t**ime series **a**nomaly **d**etection (**MTAD**) has garnered significant interest in both academic and industrial research endeavors. Numerous MTAD models have been proposed, ranging from traditional machine learning models to complex deep learning models [16, 21, 22].

With the progression of anomaly detection technology, the detection performance of these MTAD models is consistently improving, albeit with notable variations across diverse scenarios or datasets. There is no universally optimal model for anomaly detection, indicating that no single model can consistently achieve the highest performance across various scenarios or datasets.

We aim to address the prevalent issue named multivariate time series anomaly detection model selection, which is frequently encountered in practice but often disregarded in academic research. That is, **How to select an appropriate anomaly detection model for unlabeled time series datasets to achieve optimal anomaly detection performance?**

The time series model selection method typically encounters the following challenges: (1) Time series data is essentially unlabeled [8]. Acquiring data labels for extensive time series datasets is a laborious, costly, and error-prone task. (2) The absence of standardized and universally acknowledged evaluation metrics leads to inconsistencies in defining "good" performance, consequently causing variations in model selection results. Most anomalies in time series tend to appear consecutively, forming multiple anomaly segments [18]. Time series anomaly detection ought to be evaluated using range-based metrics rather than point-based metrics [12]. The absence of acknowledged evaluation metrics leads to inconsistencies in defining "good" performance. (3) The computational cost of the model selection training process is substantial. Existing model selection methods depend on the performance results of all candidate models that have been pre-trained on different datasets. Currently, the predominant MTAD models are based on deep learning, requiring each model to train over numerous epochs on each dataset.

In this paper, we introduce an unsupervised method for multivariate time series anomaly detection model selection to address the aforementioned challenges. For Challenge 1, our method is based on two unsupervised "surrogate" metrics which are related to detection performance and do not require anomaly labels. For Challenge 2, our method employs a range of evaluation metrics to obtain performance rankings for two unsupervised metrics. Subsequently, the performance ranking ordered by these diverse evaluation metrics is aggregated to achieve a more robust performance ranking. For Challenge 3, we design an early-stopping mechanism to terminate the epochs of deep learning training in order to mitigate the computational cost.

2 Related Work

Anomaly detection model selection aims to determine the most suitable anomaly detection model for a specific dataset. Utilizing pre-training in model selection has been a well-established practice in the field. Recently, Zhao et al. explored the application of meta-learning for outlier detection methods [27]. The method relies on a collection of historical outlier detection datasets that contain ground-truth (anomaly) labels and the historical performance of models on these meta-training datasets. The meta-features of datasets are utilized to construct a model that "recommends" an anomaly detection method for a given dataset. Concurrently, Kotlar et al. proposed a meta-learning method

with novel meta-features to select anomaly detection models using labeled data during the training phase [10]. The aforementioned studies do not specifically address time series datasets.

For time series datasets, Ying et al. concentrated on model selection towards time series datasets [24]. The time series are characterized based on a predetermined set of features. An anomaly-labeled time series dataset from a knowledge base is utilized to train a classifier for model selection and a regressor for hyper-parameter estimation. A comparable method is proposed by Zhang et al. who extracted time series features to select optimal models through exhaustive hyper-parameter tuning. And a classifier is then trained by using the extracted time series features and results on model performance as labels [26].

In contrast to methods based on pre-training models, Goswami et al. introduce a method to combine rankings from three categories of surrogate metrics to create a robust composite metric for unsupervised model selection [8]. Nevertheless, the method is computationally intensive, requiring over 5,000 training models and numerous metrics.

3 Problem Statement

Let $\{x_t, y_t\}_{t=1}^{T}$ denote a time series with samples $(x_1, \cdots, x_T)$, $x_t \in R^d$ and anomaly labels $(y_1, \cdots, y_T)$, $y_t \in \{0,1\}$, where $y_t = 1$ indicates that the observation x_t is an anomaly. Anomaly labels are only used to evaluate our model selection method and do not play a role in the actual model selection process.

Inspired by [8], let $\mathrm{M} = \{A_i\}_{i=1}^{N}$ denote a set of N candidate anomaly detection models. Each model A_i is a tuple including detector and hyperparameters, e.g. (GDN [4], out_layer_inter_dim=128, topk=5, $\cdots$).

The candidate anomaly detection models require unlabeled time series as training data. We consider a training / test set $\{x_t\}_{t=1}^{t_{test}-1}$ and $\{x_t\}_{t=t_{test}}^{T}$, where the training set $\{x_t\}_{t=1}^{t_{test}-1}$ is used without labels during the training process. It should be noted that A_i denotes a trained model after the training process.

We assume that a trained model A_i, when applied to the test set $\{x_t\}_{t=t_{test}}^{T}$, produces anomaly scores $\{s_t^i\}_{t=t_{test}}^{T}$, $s_t^i \in R \geq 0$. We argue that a higher anomaly score signifies a greater likelihood that the observation is more likely to be an anomaly. Subsequently, the detection performance of the model is determined by calculating the anomaly score using the corresponding evaluation metrics $Q(\{s_t^i\}_{t=1}^{T}, \{y_T\}_{t=1}^{T})$. In the next section, we discuss the choice of evaluation metrics.

Let X_{train}, X_{test}, and Y_{test} denote the training and testing data sets of observations and the label testing data set, respectively. So, the problem of **Unsupervised Time series Anomaly Detection Model Selection** can be described as follows: Given observations X_{test} and a set of models $\mathrm{M} = \{A_i\}_{i=1}^{N}$ trained using X_{train}, select a model

from M that maximizes the anomaly detection evaluation metric $Q(A_i(X_{test}), Y_{test})$. The whole model selection process cannot use anomaly labels.

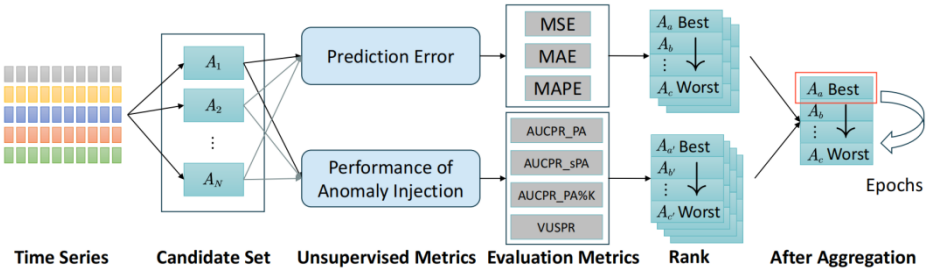

Fig. 1. The framework of our proposed model selection method.

4 Methodology

In prior research, the issue of unsupervised model selection has typically been approached as the task of identifying an unsupervised metric that exhibits strong correlation with supervised metrics [8]. As demonstrated by Fig. 1, we identified two classes of unsupervised metrics (**Prediction Error** and **Performance of Anomaly Injection**) due to their intuitive fit. After evaluating the performance ranking of candidate models using various metrics, we introduce a robust ranking aggregation method to obtain the composite ranking of model performance within each training epoch. We design an early-stopping mechanism to reduce the computational costs. This mechanism examines the Top-K models that consistently maintain their ranking in robust performance across epochs.

4.1 Prediction Error

One key attribute of an effective anomaly detection model is its ability to accurately predict or reconstruct time series data. This is the reason why numerous unsupervised MTAD models use prediction or reconstruction errors as anomaly scores without anomaly labels. Through the computation of prediction errors, various standard evaluation metrics are derived for time series forecasting tasks, including mean squared error (MSE), mean absolute error (MAE), and mean absolute percentage error (MAPE).

For multivariate time series, we calculate the average of each metric over all the variables. MSE assesses the square of the difference between the true value and the predicted value. MAE focuses on the absolute error between the true value and the predicted value. If a significant difference in magnitude exists between the true values of different samples, or if there is more concern about the percentage variance between the predicted value and the true value, MAPE is employed.

Using the evaluation metrics of prediction error as the metric for anomaly detection may not be perfect enough. Unsupervised training datasets for time series typically include anomaly data, which can have an impact on the accuracy of prediction results. Meanwhile, some anomalies are difficult to reconstruct or predict.

4.2 Performance of Anomaly Injection

Another key attribute of an effective anomaly detection model is its ability to effectively detect the injected anomalies. Given the challenge of acquiring labels for extensive time series datasets, one method is to transform the initial unlabeled dataset into a labeled dataset by injecting synthetical anomalies. So, the injected anomalies are assigned a pseudo label as 1, while the rest of the data points are labeled as 0. Previous study has investigated the utilization of synthetic anomaly injection for training anomaly detection models [3].

Thus, the anomaly detection performance of the candidate model set can be assessed using supervised evaluation metrics along this research line. Given an initial multivariate time series lacking anomaly labels, we inject six specific types of anomalies randomly. We have developed straightforward, productive programs to inject various types of anomalies as illustrated in Fig. 2.

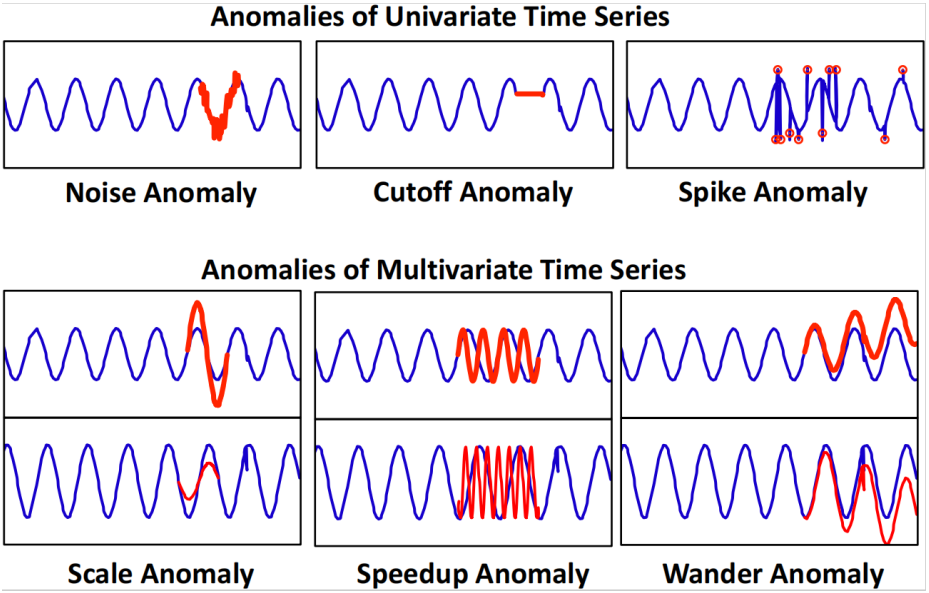

Fig. 2. Different types of injected anomalies. We inject different types of anomalies randomly across the initial time series. The blue line is initial series before anomalies are injected, the red line is the injected anomalies.

We design six types of injected anomalies, which belong to univariate and multivariate time series anomalies. Univariate time series anomalies change the data of a single

variable over time, including Noise anomaly, Cutoff anomaly, and Spike anomaly. And multivariate time series anomalies change the data of multiple variables simultaneously, including Scale anomalies, Speedup anomalies, and Wander anomaly. Only one type of anomalies is injected into each sliding window of the multivariate time series which has been standardized via Min-max normalization.

Noise Anomaly. We inject noise drawn from a standard normal distribution $\mathcal{N}(0,1)$ into one variable of time series.

Cutoff Anomaly. We set a random fixed value from $\mathcal{N}(0,1)$ into a period of one variable.

Spike Anomaly. We randomly set the extreme values (±1) within the sliding window.

Scale Anomaly. We scale the values of different variables using distinct scale factors. Differences within data magnitudes across different variables lead to scale anomaly.

Speedup Anomaly. We change the frequencies of different variables using distinct frequency factors. Differences within frequencies across different variables lead to speedup anomaly.

Wander Anomaly. We add a trend (up or down) to different variables. Differences within data trends across different variables lead to wander anomaly.

After injecting anomalies, we can obtain a pseudo-labeled dataset. Therefore, the performance of the model A_i can be measured using the supervised evaluation metrics calculated by confusion matrix. However, these point-based metrics ignore the sequential nature of time series; thus, time series anomaly detection is usually evaluated using the point-adjusted (PA) versions of precision and recall [15]. In this case, detecting any of these points is treated as if all points inside the anomaly segment were detected.

And considerable studies have shown the overestimation of anomaly detection performance caused by PA [6, 9, 12]. PA%K is proposed which applies PA only if the ratio of correctly detected anomalies to the segment length exceeds the threshold K [9]. Recently, sPA is proposed which considers both the position and frequency of alarms in the raw results via a ranging function and a rewarding function [12]. Here, we use AUCPR applied different point-adjusted versions as the supervised metrics, named AUCPR_PA, AUCPR_PA%K and AUCPR_sPA, respectively. AUCPR is more practical in real-world applications because it directly relates to benefits and costs of detection results without setting thresholds, and achieving high AUC-PR is more challenging. Recently, VUSPR is proposed as a robust evaluation metric [15]. Thus, we aim to identify models with the highest anomaly detection evaluation metrics (Q includes AUCPR_PA, AUCPR_PA%K, AUCPR_sPA and VUSPR).

Each metric possesses inherent limitations, and the comparability of values across different metrics lacks significance. The ranking of candidate models under various metrics can indicate the performance of anomaly detection to varying extents. Hence, our method involves selecting the optimal model by aggregating the model performance rankings of different imperfect metrics.

4.3 Performance of Anomaly Injection

Our candidate models have various noisy detection rankings for each evaluation metrics. Inspired by [8], we consider N models and a collection of Z evaluation metrics $\{\sigma_1, \sigma_2, \cdots, \sigma_{\mathrm{Z}}\}$. Here, we use σ_{z} to refer to the evaluation metric and the ranking it induces. For any given metric σ_{z}, the i^{th} model receives $\mathrm{N} - \sigma_{\mathrm{z}}(i)$ points. Thus, the i^{th} model accrues a total of $\sum_{\mathrm{z}=1}^{\mathrm{Z}}(N - \sigma_{\mathrm{z}}(\mathrm{i}))$ points. Finally, all the models are ranked in decreasing order of their total points.

In model selection, we only care about top-ranked models. Thus, to improve model selection performance at the top-ranks, we only consider models at the top-k rankings, setting the positions of the rest of the models to N [5]. Specifically, under the top-k aggregation scheme, the i^{th} model accrues a total of $\sum_{\mathrm{z}=1}^{\mathrm{Z}}(\Phi[\sigma_{\mathrm{z}}(\mathrm{i}) \leq \mathrm{k}] \cdot (\mathrm{N} - \sigma_{\mathrm{z}}(\mathrm{i})))$ points. This increases the probability of models which have top-k ranks consistently across evaluation metrics, to have top ranks upon aggregation.

4.4 Early-stopping Mechanism of Epochs

In the context of model selection, it is impractical to train every candidate model across numerous iterations of training using various datasets. We argue that multiple iterations of training epochs in deep learning lead to a gradual decrease in the loss function by employing continuous forward and back propagation processes, ultimately resulting in a consistent anomaly detection performance. Nevertheless, the relative performance rank among the models remains constant after low training iterations. Hence, we propose an early-stopping mechanism of training to minimize redundant iterative epochs. This mechanism operates by assessing the stability of the Top-α models within the robust performance rank.

The mechanism is intuitive: if the sequence of the Top-α model within the robust performance ranking remains consecutively consistent for β epochs (where α and β are set as 3 by default), the model selection process will cease computation and suggest the Top-1 model to the user for anomaly detection.

5 Experimental Studies

5.1 Experimental Setup

Six public real-world anomaly detection datasets are used including **SWaT** [13], **SMAP** [14], **MSL** [14], **PSM** [1], **SMD** [17], and **ASD** [11]. In our proposed model selection method, we pre-set the candidate model set of **GDN** [4], **OmniAnomaly** [17], **MSCRED** [25], **LSTM-VAE** [16], **TimesNet** [21], **USAD** [2], **TranAD** [19], **DAGMM** [28], **FCSTGNN** [20] and **FourierGNN** [23] models. The experimental section of this paper primarily aims to validate the validity of our proposed model selection method.

We evaluate our model selection method by considering the adjusted best F1 score and AUCPR of the top-1 model in the robust performance rank. We conduct a comparative experiment of our method against four baseline methods: (1) "**Supervised**" involves utilizing labeled training datasets for the purpose of model selection. (2) "**Random**" involves selecting models from the candidate model set randomly, followed by determining detection performance on the testing set. (3) "**Ideal**" identifies the optimal model from the testing set based on the supervised metrics and provides its corresponding performance. (4) "**TimeGPT**" [7] is proposed as the first foundation model for time series, capable of detecting anomalies for diverse datasets not seen during pre-training.

5.2 Results and Discussion

Table 1 presents the anomaly detection performance of the selected methods from 5 model selection methods on six datasets in terms of F1 score and AUCPR. Each baseline provides a specific threshold selection method, and the best F1 score is reported correspondingly. The best results are highlighted in bold except for "Ideal" known as the best standard answer. Our results show that our proposed unsupervised model selection method is comparably effective to the "Supervised" method in selecting the optimal MTAD model.

Our results in Table 2 indicate that there is no single evaluation metric that consistently selects the best model across all datasets. The ablation study presented in Table 2 demonstrates that the impact on detection performance varies across datasets when different evaluation metrics are excluded from the comprehensive method. Understanding the potential anomalies present in datasets beforehand can aid in the process of model selection. After the removal of Top-K aggregation, there is a significant decrease in performance, suggesting that concentrating solely on the "top" candidate models can

Table 1. Best F1 scores and AUCPR of our method and its baselines on datasets. The best score under "Ideal" is boldfaced.

Methods	SWaT		SMAP		MSL		PSM		SMD		ASD	
	F1	AUCPR	F1	AUCPR	F1	AUCPR	F1	AUCPR	F1	AUCPR	F1	AUCPR
Supervised	0.699	0.662	**0.847**	0.769	**0.678**	0.693	0.772	**0.847**	0.847	0.495	**0.789**	0.582
Random	0.485	0.518	0.715	0.734	0.701	0.542	0.665	0.479	0.713	0.405	0.558	0.403
TimeGPT (2023)	0.587	0.631	0.743	0.759	0.685	0.584	0.736	0.754	0.724	**0.561**	0.718	0.613
Ideal (Best)	0.741	0.744	0.881	0.857	0.785	0.754	0.826	0.876	0.881	0.624	0.801	0.664
Ours	**0.713**	**0.693**	0.815	**0.799**	0.658	**0.711**	**0.788**	0.829	**0.858**	0.551	0.759	**0.644**

Table 2. Ablation study in terms of F1 score and AUCPR of our method and its variants.

	SWaT		SMAP		MSL		PSM		SMD		ASD	
	F1	AUCPR	F1	AUCPR	F1	AUCPR	F1	AUCPR	F1	AUCPR	F1	AUCPR
Ours	**0.713**	**0.693**	**0.815**	**0.799**	**0.658**	**0.711**	**0.788**	**0.829**	**0.858**	**0.551**	**0.759**	**0.644**
w/o Pred.MSE	0.665	0.628	0.753	0.645	0.635	0.642	0.765	0.772	0.713	0.465	0.658	0.603
w/o Pred.MAE	0.627	0.612	0.759	0.630	0.613	0.628	0.718	0.687	0.717	0.530	0.686	0.624
w/o Pred.MAPE	0.601	0.653	0.717	0.709	0.558	0.649	0.737	0.738	0.551	0.521	0.727	0.642
w/o Inject.PA	0.647	0.653	0.758	0.660	0.582	0.601	0.643	0.611	0.628	0.450	0.608	0.635
w/o Inject.sPA	0.672	0.557	0.613	0.709	0.636	0.685	0.579	0.789	0.585	0.501	0.576	0.595
w/o Inject.PA%K	0.667	0.575	0.692	0.610	0.628	0.612	0.553	0.692	0.735	0.451	0.720	0.587
w/o Inject.VUS	0.647	0.596	0.708	0.664	0.601	0.584	0.736	0.759	0.724	0.491	0.618	0.573
w/o Top-K Aggre.	0.621	0.644	0.771	0.700	0.585	0.654	0.726	0.776	0.681	0.514	0.601	0.565
w/o Early-stopping	0.698	0.674	0.783	0.767	0.625	0.683	0.751	0.787	0.791	0.524	0.701	0.611

enhance performance after the ranking aggregation (w/o Top-K Aggre.). The removal of the early-stopping mechanism has minimal effect on performance, suggesting that the mechanism does not decrease computational cost at the expense of performance (w/o Early-stopping).

6 Conclusion

In this paper, we propose a model selection method to address the prevalent issue multivariate time series anomaly detection model selection. By aggregating the performance ranks obtained by individual evaluation metrics, we obtain a more robust performance rank of candidate models. We introduce an early-stopping mechanism to reduce the computational cost by terminating the epochs of training. The experimental results illustrate that our proposed unsupervised model selection method is as effective as the supervised selection of the optimal model.

Acknowledgments. This work was supported by the National Key R&D Program of China (Grant No.2022ZD0115302), the National Natural Science Foundation of China (Grant No.61379052), the Science Foundation of Ministry of Education of China (Grant No.2018A02002), the Natural Science Foundation for Distinguished Young Scholars of Hunan Province (Grant No.14JJ1026).

References

1. Abdulaal, A., Liu, Z., Lancewicki, T.: Practical approach to asynchronous multivariate time series anomaly detection and localization. In: Proceedings of the 27th ACM SIGKDD conference on knowledge discovery & data mining. pp. 2485–2494 (2021)
2. Audibert, J., Michiardi, P., Guyard, F., Marti, S., Zuluaga, M.A.: Usad: Unsupervised anomaly detection on multivariate time series. In: Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining. pp. 3395–3404 (2020)
3. Carmona, C.U., Aubet, F.X., Flunkert, V., Gasthaus, J.: Neural contextual anomaly detection for time series. arXiv preprint arXiv:2107.07702 (2021)

4. Deng, A., Hooi, B.: Graph neural network-based anomaly detection in multivariate time series. In: Proceedings of the AAAI conference on artificial intelligence. vol. 35, pp. 4027–4035 (2021)
5. Fagin, R., Kumar, R., Sivakumar, D.: Comparing top k lists. SIAM Journal on discrete mathematics 17(1), 134–160 (2003)
6. Garg, A., Zhang, W., Samaran, J., Savitha, R., Foo, C.S.: An evaluation of anomaly detection and diagnosis in multivariate time series. IEEE Transactions on Neural Networks and Learning Systems 33(6), 2508–2517 (2021)
7. Garza, A., Mergenthaler-Canseco, M.: Timegpt-1. arXiv preprint arXiv:2310.03589 (2023)
8. Goswami, M., Challu, C., Callot, L., Minorics, L., Kan, A.: Unsupervised model selection for time-series anomaly detection. arXiv preprint arXiv:2210.01078 (2022)
9. Kim, S., Choi, K., Choi, H.S., Lee, B., Yoon, S.: Towards a rigorous evaluation of time-series anomaly detection. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 36, pp. 7194–7201 (2022)
10. Kotlar, M., Punt, M., Radivojević, Z., Cvetanović, M., Milutinović, V.: Novel meta-features for automated machine learning model selection in anomaly detection. IEEE Access 9, 89675–89687 (2021)
11. Li, Z., Zhao, Y., Han, J., Su, Y., Jiao, R., Wen, X., Pei, D.: Multivariate time series anomaly detection and interpretation using hierarchical inter-metric and temporal embedding. In: Proceedings of the 27th ACM SIGKDD conference on knowledge discovery & data mining. pp. 3220–3230 (2021)
12. Liu, M., Wang, Y., Xu, H., Zhou, X., Li, B., Wang, Y.: Smoothing point adjustment-based evaluation of time series anomaly detection. In: ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 1–5. IEEE (2023)
13. Mathur, A.P., Tippenhauer, N.O.: Swat: A water treatment testbed for research and training on ics security. In: 2016 international workshop on cyber-physical systems for smart water networks (CySWater). pp. 31–36. IEEE (2016)
14. O'Neill, P., Entekhabi, D., Njoku, E., Kellogg, K.: The nasa soil moisture active passive (smap) mission: Overview. In: 2010 IEEE International Geoscience and Remote Sensing Symposium. pp. 3236–3239. IEEE (2010)
15. Paparrizos, J., Boniol, P., Palpanas, T., Tsay, R.S., Elmore, A., Franklin, M.J.: Volume under the surface: a new accuracy evaluation measure for time-series anomaly detection. Proceedings of the VLDB Endowment 15(11), 2774–2787 (2022)
16. Park, D., Hoshi, Y., Kemp, C.C.: A multimodal anomaly detector for robot-assisted feeding using an lstm-based variational autoencoder. IEEE Robotics and Automation Letters 3(3), 1544–1551 (2018)
17. Su, Y., Zhao, Y., Niu, C., Liu, R., Sun, W., Pei, D.: Robust anomaly detection for multivariate time series through stochastic recurrent neural network. In: Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining. pp. 2828–2837 (2019)
18. Tatbul, N., Lee, T.J., Zdonik, S., Alam, M., Gottschlich, J.: Precision and recall for time series. Advances in neural information processing systems 31 (2018)
19. Tuli, S., Casale, G., Jennings, N.R.: Tranad: Deep transformer networks for anomaly detection in multivariate time series data. arXiv preprint arXiv:2201.07284 (2022)
20. Wang, Y., Xu, Y., Yang, J., Wu, M., Li, X., Xie, L., Chen, Z.: Fully-connected spatial-temporal graph for multivariate time-series data. In: Proceedings of the AAAI conference on artificial intelligence. vol. 38, pp. 15715–15724 (2024)

21. Wu, H., Hu, T., Liu, Y., Zhou, H., Wang, J., Long, M.: Timesnet: Temporal 2d-variation modeling for general time series analysis. In: The eleventh international conference on learning representations (2022)
22. Xu, H., Chen, W., Zhao, N., Li, Z., Bu, J., Li, Z., Liu, Y., Zhao, Y., Pei, D., Feng, Y., et al.: Unsupervised anomaly detection via variational auto-encoder for seasonal kpis in web applications. In: Proceedings of the 2018 world wide web conference. pp. 187–196 (2018)
23. Yi, K., Zhang, Q., Fan, W., He, H., Hu, L., Wang, P., An, N., Cao, L., Niu, Z.: Fouriergnn: Rethinking multivariate time series forecasting from a pure graph perspective. Advances in neural information processing systems 36, 69638–69660 (2023)
24. Ying, Y., Duan, J., Wang, C., Wang, Y., Huang, C., Xu, B.: Automated model selection for time-series anomaly detection. arXiv preprint arXiv:2009.04395 (2020)
25. Zhang, C., Song, D., Chen, Y., Feng, X., Lumezanu, C., Cheng, W., Ni, J., Zong, B., Chen, H., Chawla, N.V.: A deep neural network for unsupervised anomaly detection and diagnosis in multivariate time series data. In: Proceedings of the AAAI conference on artificial intelligence. vol. 33, pp. 1409–1416 (2019)
26. Zhang, P., Jiang, X., Holt, G.M., Laptev, N.P., Komurlu, C., Gao, P., Yu, Y.: Self-supervised learning for fast and scalable time series hyper-parameter tuning. arXiv preprint arXiv:2102.05740 (2021)
27. Zhao, Y., Rossi, R., Akoglu, L.: Automatic unsupervised outlier model selection. Advances in Neural Information Processing Systems 34, 4489–4502 (2021)
28. Zong, B., Song, Q., Min, M.R., Cheng, W., Lumezanu, C., Cho, D., Chen, H.: Deep autoencoding gaussian mixture model for unsupervised anomaly detection. In: International conference on learning representations (2018)