

# KAN-AD: Time Series Anomaly Detection with Kolmogorov–Arnold Networks

Quan Zhou<sup>1,2</sup> Changhua Pei<sup>1,3</sup> Fei Sun<sup>4</sup> Jing Han<sup>5</sup> Zhengwei Gao<sup>5</sup> Haiming Zhang<sup>1</sup> Gaogang Xie<sup>1</sup>  
Dan Pei<sup>6</sup> Jianhui Li<sup>7,1</sup>

## Abstract

Time series anomaly detection (TSAD) underpins real-time monitoring in cloud services and web systems, allowing rapid identification of anomalies to prevent costly failures. Most TSAD methods driven by forecasting models tend to overfit by emphasizing minor fluctuations. Our analysis reveals that effective TSAD should focus on modeling “normal” behavior through smooth local patterns. To achieve this, we reformulate time series modeling as approximating the series with smooth univariate functions. The local smoothness of each univariate function ensures that the fitted time series remains resilient against local disturbances. However, a direct KAN implementation proves susceptible to these disturbances due to the inherently localized characteristics of B-spline functions. We thus propose **KAN-AD**, replacing B-splines with truncated Fourier expansions and introducing a novel lightweight learning mechanism that emphasizes global patterns while staying robust to local disturbances. On four popular TSAD benchmarks, KAN-AD achieves an average 15% improvement in detection accuracy (with peaks exceeding 27%) over state-of-the-art baselines. Remarkably, it requires fewer than 1,000 trainable parameters, resulting in a 50% faster inference speed compared to the original KAN, demonstrating the approach’s efficiency and practical viability.

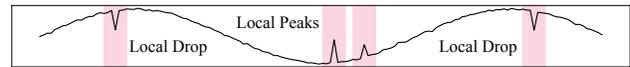


Figure 1. Illustration of local drops and peaks.

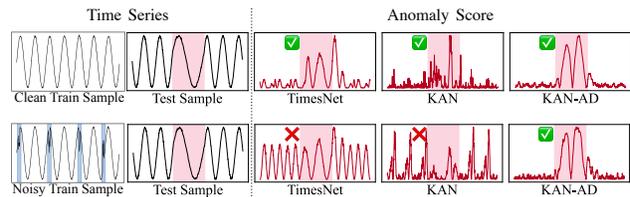


Figure 2. Comparison of anomaly detection performance. Top: All methods successfully detect anomalies when trained on clean data (black curve, anomalous segments in pink). Bottom: TimesNet and KAN fail to detect anomalies when trained on noisy data. Blue markers indicate local drops and peaks; red curve shows anomaly scores.

## 1. Introduction

Time Series Anomaly Detection (TSAD) serves as a critical component in modern IT infrastructure (Li et al., 2019; Qu et al., 2024) and manufacturing systems (Zhan et al., 2021; Wang et al., 2022), enabling rapid identification of potential anomalies and providing sufficient clues for fault localization (Sun et al., 2024; Kieu et al., 2022). The emergence of deep learning-based forecasting approaches (Xu et al., 2022; Wu et al., 2023; Zhou et al., 2023) have superseded traditional rule-based methods (Breunig et al., 2000; Siffer et al., 2017), establishing new state-of-the-art performance through their capacity to fit historical data and detect anomalies via prediction-observation comparisons.

However, the effectiveness of the forecasting-based approach declines when encountering time series with localized disturbances. As illustrated in Figure 1, time series data frequently exhibit local peaks and drops that can significantly impact model learning. Existing deep learning methods (Tuli et al., 2022; Wu et al., 2023) often overfit to these local disturbances, compromising their ability to detect anomalies effectively. From the third column of Figure 2, we can observe that compared to training with clean data, TimesNet (Wu et al., 2023) trained on noisy data fails to detect anomalies in the samples.

Our experimental analysis reveals that forecasting-based

<sup>1</sup>Computer Network Information Center, Chinese Academy of Sciences <sup>2</sup>University of the Chinese Academy of Sciences <sup>3</sup>Hangzhou Institute for Advanced Study, University of the Chinese Academy of Sciences <sup>4</sup>Institute of Computing Technology, Chinese Academy of Sciences <sup>5</sup>ZTE <sup>6</sup>Department of Computer Science and Technology, Tsinghua University <sup>7</sup>School of Frontier Sciences, Nanjing University. Correspondence to: Jianhui Li <lijh@nju.edu.cn>, Changhua Pei <chpei@cnic.cn>.

TSAD methods suffer performance degradation by attempting to model every detailed patterns in raw time series data. While these methods aim to identify anomalies through comparison with predicted behavior, such detailed modeling proves unnecessary and potentially detrimental, especially given that real-world time series typically contain various forms of anomalies and irrelevant disturbances, presenting two significant challenges: firstly, the difficulty in establishing a universal criterion for filtering these disturbances, and secondly, developing another model to ensure the forecasting model’s input is free of local disturbances is resource-intensive. Given these inherent limitations in both filtering-based and dual-modeling approaches, researchers have explored VAE-based approaches to address the challenge of local disturbance mitigation. VAE-based approaches (Xu et al., 2018; Wang et al., 2024) assume that normal patterns in time series cluster in a low-dimensional latent space and can be effectively reconstructed, thereby overcoming interference from data perturbations. Nevertheless, as demonstrated in FCVAE (Wang et al., 2024), VAE-based approaches struggle with underfitting, which impairs their ability to reconstruct the original time series and limits their effectiveness.

To mitigate local disturbances, we reformulate TSAD by approximating time series using smooth univariate functions, building on the theoretical foundation that normal sequences exhibit greater local smoothness than abnormal ones (Xu et al., 2022). To achieve this formulation, Kolmogorov–Arnold Networks (KANs) (Liu et al., 2025) offer a promising direction by decomposing complex objectives into combinations of learnable univariate functions based on the Kolmogorov–Arnold representation theorem (Kolmogorov, 1957). This decomposition approach has shown remarkable effectiveness in various domains (Yu et al., 2024; Bodner et al., 2024). However, direct application of KAN to TSAD presents significant challenges. From the fourth column in the upper part of Figure 2, it can be observed that models trained on clean training samples can detect anomalies in the test samples. But we find that KAN fails to detect anomalies when the training samples contain noisy samples. The main reason is that, although KAN can specify univariate functions, *i.e.*, B-spline function, these functions are not specifically designed for time series and can still overfit local features, failing to completely eliminate the impact of local peaks or drops.

To address these challenges, we propose KAN-AD, adopting KAN as our backbone. By considering the characteristics of time series, we redesign KAN in three aspects. First, we replace the B-spline function with Fourier series. Fourier series have local smoothness compared to spline functions, while their natural periodicity allows for better modeling of global patterns (Dym & HP, 1972; Stein & Shakarchi, 2011). Second, as the Fourier series contains unlimited

terms which is computation intensive, we only use the first  $N$  terms of Fourier series. To overcome the limitation that the first  $N$  terms of Fourier series can only model periodic no smaller than  $\frac{1}{N}$ , we designed an alternative index-based univariate function to capture the fine-scale periodic missing from the first  $N$  terms. Third, we incorporated differencing to isolate time series trend effects on coefficient estimation, leading to improved modeling accuracy through more precise coefficients.

Our comprehensive evaluation demonstrates that KAN-AD achieves 15% higher F1 accuracy while being 50% faster than the original KAN architecture. Our code is publicly available at <https://github.com/CSTCloudOps/KAN-AD>. Our contributions are as follows:

- We reformulate the problem to assist deep learning-based forecasting models for time series anomaly detection (TSAD) tasks by minimizing overfitting to local perturbations.
- We introduce KAN-AD, an innovative TSAD approach. KAN-AD, built meticulously on the KAN backbone, exhibits substantial improvements in both detection precision and inference efficiency.
- We performed comprehensive experiments on four publicly available datasets, verifying the effectiveness and efficiency against state-of-the-art TSAD benchmarks.

## 2. Preliminaries and Problem Formulation

### 2.1. Problem Statement

This paper primarily addresses the issue of anomaly detection in single time series curves, also known as univariate time series (UTS). To elaborate on the problem more comprehensively, consider the following UTS observational data:  $x_{0:t} = \{x_0, x_1, x_2, \dots, x_t\}$  and anomaly labels  $C = \{c_0, c_1, c_2, \dots, c_t\}$ , where  $x_t \in \mathbb{R}$ ,  $c \in \{0, 1\}$ , and  $t \in \mathbb{N}$ . Here,  $x_{0:t}$  represents the entire observed time series, and  $C$  denotes the temporal anomaly labels.

*Given a UTS  $x = [x_0, x_1, x_2, \dots, x_t]$ , the objective of UTS anomaly detection is to utilize the data  $[x_0, x_1, \dots, x_i]$  preceding each point  $x_i$  to predict  $c_i$ .*

### 2.2. Kolmogorov–Arnold Networks

#### 2.2.1. THEORETICAL FOUNDATION

The Kolmogorov–Arnold representation theorem demonstrates that any multivariate continuous function can be decomposed into a finite sum of univariate functions, as shown in Equation (1), where  $\varphi_{q,p}$  are univariate functions that map each input variable  $x_p$ , and  $\Phi_q$  are continuous

functions.

$$f(x_1, x_2, \dots, x_n) = \sum_{q=1}^{2n+1} \Phi_q \left( \sum_{p=1}^n \varphi_{q,p}(x_p) \right) \quad (1)$$

$$\text{KAN}(x) = (\Phi_{L-1} \circ \Phi_{L-2} \circ \dots \circ \Phi_0)(x) \quad (2)$$

### 2.2.2. NETWORK ARCHITECTURE AND FUNCTION REPRESENTATION

KAN consists of a series of interconnected univariate sub-networks, each responsible for learning distinct features of the data. Unlike traditional multi-layer perceptrons (MLPs), which employ fixed activation functions at each node, KAN replaces each weight parameter with a univariate function. The resulting functional form for deeper KAN can be expressed as Equation (2), where each  $\Phi_l$  represents a layer of univariate functions applied to the input or intermediate outputs. The vanilla KAN (Liu et al., 2025) implements these univariate functions using B-splines (De Boor, 1978), which provide localized function approximation capabilities. However, this localization property presents a notable consideration in anomaly detection contexts. Since anomalous patterns typically manifest as localized features (Xu et al., 2022), B-splines may inadvertently fit these outliers, potentially compromising model accuracy.

## 3. Methodology

The core challenge in time series anomaly detection (TSAD) lies in establishing accurate normal patterns while maintaining robustness to local disturbances (Li et al., 2021). Traditional approaches that directly predict based on historical data inevitably incorporate local noise into their learned patterns. Building on the observation that normal sequences exhibit greater smoothness than abnormal ones, we propose KAN-AD, a novel anomaly detection framework that leverages this smoothing feature to identify anomalies in complex time series data.

### 3.1. Design of KAN-AD

The pipeline of KAN-AD consists of three main stages: **mapping**, **reducing**, and **projection**. In the mapping phase, we decompose the input time window into multiple univariate functions. The reducing phase then combines these functions through learned coefficients to reconstruct the “normal” pattern. Finally, the projection phase leverages this pattern to predict future behavior, enabling anomaly detection through comparison with real-time observations.

$$f(x_{0:i}) = A_0 + \underbrace{\sum_{n=1}^N (A_n \cos(nx_{0:i}) + B_n \sin(nx_{0:i}))}_{g(x_{0:i})} + \epsilon \quad (3)$$

$$\begin{aligned} \mathbf{H} &= \text{Stack}(\cos(x_{0:i}), \sin(x_{0:i}), \dots, \cos(nx_{0:i}), \sin(nx_{0:i})) \\ \Theta(x_{0:i}) &= [A_1, B_1, A_2, B_2, \dots, A_n, B_n] \\ x'_{0:i} &= A_0 + \Theta(x_{0:i}) \times \mathbf{H} \end{aligned} \quad (4)$$

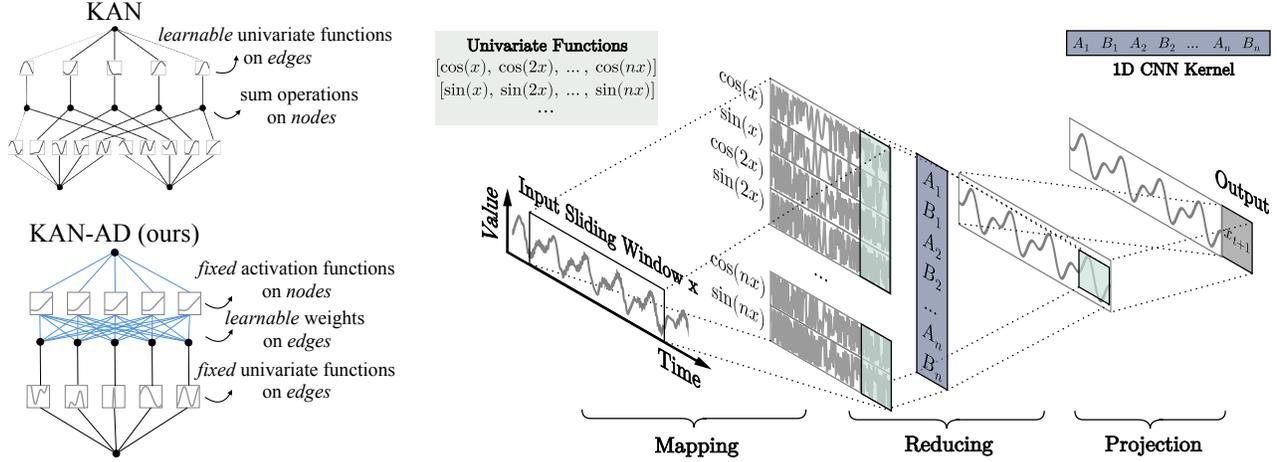
Formally, we employ Fourier series for normal pattern representation, motivated by two key advantages over alternative approaches such as B-spline functions. First, the constituent sine and cosine functions exhibit superior local smoothness, avoiding the potential overfitting to local noise. Second, Fourier series naturally capture global patterns, particularly excelling at modeling periodic behaviors in time series. Following this motivation, we introduce the function deconstruction (FD) mechanism, where  $f$ , the mapping between the historical window  $x_{0:i}$  and its next behavior  $x_{i+1}$ , can be expanded as shown in Equation (3). The normal pattern can be represented by the finite  $N$  terms of the series (Kolmogorov, 1957), denoted as  $g(x)$ , while the terms beyond  $N$  capture the stochastic observational noise  $\epsilon$ . The normal pattern  $x'_{0:i}$  can then be expressed as in Equation (4), where  $\mathbf{H}$  denotes the univariate function matrix. This decomposition combined with learnable coefficients filters out potential noise and significantly simplifies the construction of normal patterns.

### 3.2. Mapping Phase

As shown in Figure 3b, the primary purpose of the mapping phase is to transform the original time series signal  $x_{0:i} \in \mathbb{R}^T$  into multiple new sets of values  $x_{0:i} \in \mathbb{R}^{T \times (N+N)}$  through a series of univariate functions. Here,  $T$  is the size of the sliding window. The first  $N$  represents the number of sine series univariate functions, and the other  $N$  represents the number of cosine series univariate functions. The detailed calculation method is shown in Equation (3). Notably, besides the univariate function terms, an  $A_0$  term representing the average value within the sliding window is also present, which varies across different windows. To mitigate the impact of fluctuating  $A_0$  on coefficient fitting, a constant term elimination module is employed.

**Constant Term Elimination:** In Fourier series,  $A_0$  represents the mean value of the function. Although normalization ensures that the entire time series has a mean of zero, individual time windows may still exhibit significant fluctuations in their means due to the presence of a trend. These variations in the constant term ultimately affect the model’s accurate estimation of Fourier coefficients, leading to biases in the construction of the normal pattern.

To mitigate the impact of mean fluctuations on the model’s approximation of normal time series patterns, we employ *first-order differencing* during data preprocessing to minimize the residual trend component in the data and subsequently renormalize the differenced data. This strategy



(a) Illustration of learning components in KAN and KAN-AD. KAN-AD learns the coefficients on edges with fixed univariate functions, and performs weighted sum operations on nodes. Blue lines indicate edges with weights.

(b) Illustration of the KAN-AD process using a sliding window approach. During the mapping phase, raw time windows are transformed into multiple univariate functions. In the reducing phase, a one-dimensional convolutional kernel learns coefficients for these univariate functions, aggregating them into a normal pattern for the current time window. In the projection phase, a single-layer MLP predicts future normal patterns.

Figure 3. Illustration of KAN-AD.

allows the model to focus on estimating Fourier coefficients  $A_{1:n}$  and  $B_{1:n}$ , thereby avoiding the need to learn frequently changing constant terms. After this differential strategy, the normal pattern  $x'_{0:i}$  can be expressed as  $x'_{0:i} \sim \Theta(x_{0:i}) \times \mathbf{H}$

**Periodic-Enhanced KAN-AD:** Fourier series of finite  $N$  terms cannot model a period smaller than  $\frac{1}{N}$ , which limits KAN-AD’s ability to express time series containing more subtle periods.

To address this limitation and enhance the model’s ability to capture periodic patterns in time series, we introduce additional univariate functions with different periods. Specifically, we incorporate trigonometric components  $\cos(\frac{2\pi ni}{T})$  and  $\sin(\frac{2\pi ni}{T})$  where  $i$  denotes the window index, with coefficients learned through one-dimensional convolutional networks. Our implementation utilizes three complementary univariate functions shown in Equation (5): the raw time variable  $X$ , the Fourier series  $S_n$ , and the sine-cosine wave  $P_n$ . This integration of multi-periodic univariate functions enhances KAN-AD’s capacity to model temporal patterns.

$$\begin{aligned} X &= x_{0:i} \\ S_n &= \{\sin(nx_{0:i}), \cos(nx_{0:i})\} \\ P_n &= \{\sin(\frac{2\pi ni}{T}), \cos(\frac{2\pi ni}{T})\} \end{aligned} \quad (5)$$

### 3.3. Reducing Phase

Another challenge in real-world time series anomaly detection is the high computational cost. Existing methods often sacrifice efficiency for accuracy, making them impractical in resource-constrained or large-scale settings.

The function deconstruction (FD) mechanism addresses this challenge by transforming the modeling of normal patterns into a weighted combination of univariate functions. This transformation substantially reduces the model’s parameter quantity - instead of requiring numerous parameters for fine-grained feature modeling, FD mechanism achieves efficient representation through estimating coefficients of a small number of univariate functions.

$$\mathbf{H}^{(0)} = \text{Stack}(X, S_1, P_1, \dots, S_n, P_n), \forall n \in [1, \dots, N] \quad (6)$$

$$\mathbf{H}^{(l)} = \text{CNN}(\text{CNN}(\mathbf{H}^{(l-1)})) \quad \forall l \in [1, 2, \dots, L] \quad (7)$$

$$\text{Conv}(\mathbf{H}) = \sum_{c=1}^{2N} \sum_{m=0}^2 W_c[m] \cdot \mathbf{H}_c[i+m-1] \quad (8)$$

$$\text{CNN}(\mathbf{H}) = \text{GELU}(\text{BN}(\text{Conv}(\mathbf{H}))) \quad (9)$$

To effectively estimate these univariate function coefficients, we employ a stacked one-dimensional convolutional neural network (1D CNN). This architecture choice is motivated by two key factors: 1D CNNs excel at sequence modeling through temporal dimension traversal, while their convolutional kernels naturally capture the diverse features introduced by the FD mechanism. As shown in Equation (6), KAN-AD first constructs a univariate function matrix  $\mathbf{H}^{(0)}$  by combining the required functions for a given time window. This matrix is then processed through multiple stacked 1D convolutional layers with a kernel size of 3, progressively approximating the normal pattern through coefficient learning, as expressed in Equation (7). Here,  $L$  denotes the number of CNN blocks, with the network  $\text{CNN}(\mathbf{H})$  and

convolution operation  $\text{Conv}(\mathbf{H})$  defined in Equations (8) and (9). The convolution operation in Equation (8) applies a kernel  $W_c$  to each channel  $\mathbf{H}_c$ , where indices  $m$  and  $t$  represent positions within the convolutional kernel and time window, respectively.

To ensure training stability and reduce internal covariate shift, we apply batch normalization (Ioffe & Szegedy, 2015) after each convolutional layer (Equation (9)), followed by Gaussian Error Linear Units (GELUs) (Hendrycks & Gimpel, 2016) for activation. The final stage of the reducing phase employs a residual connection (He et al., 2016) between the hidden state  $\mathbf{H}^{(L)}$  and the original input  $\mathbf{H}^{(0)}$  to maintain numerical stability, as shown in Equation (10). Finally, a 1-width convolutional kernel reduces the dimensionality of  $\mathbf{H}^{(L)'}$  to generate the normal pattern approximation  $x'_{0:i}$  within the current time window:

$$\mathbf{H}^{(L)'} = \mathbf{H}^{(L)} + \mathbf{H}^{(0)} \quad (10)$$

$$x'_{0:i} = \text{GELU}(\text{BN}(\text{DownConv}(\mathbf{H}^{(L)' }))) \quad (11)$$

Here,  $\text{DownConv}(\mathbf{H}) = \sum_{c=1}^{2N} W_c \cdot \mathbf{H}_c[i]$  denotes the convolution operation for reducing dimensions.

### 3.4. Projection Phase

After obtaining the current window’s normal mode approximation  $x'_{0:i}$ , we predict the future behavior  $x_{i+1}$  through a single-layer MLP, leveraging KAN-AD’s accurate approximation capability:

$$x_{i+1} = W \cdot x'_{0:i} + b \quad (12)$$

where  $W$  and  $b$  denote the weight matrix and bias term of the linear layer.

## 4. Evaluation

In this section, we conduct comprehensive experiments primarily aimed at answering the following research questions.

- RQ1:** How does KAN-AD compare to state-of-the-art anomaly detection methods in performance and efficiency?
- RQ2:** How sensitive is KAN-AD to hyperparameters?
- RQ3:** How effective is each design choice in KAN-AD?
- RQ4:** How sensitive is KAN-AD to anomalies in the training data?

In addition, we also evaluate our method on a multivariate time series anomaly detection dataset to demonstrate the application potential of KAN-AD in more scenarios.

### 4.1. Experimental settings

#### 4.1.1. DATASET

We evaluate KAN-AD on four publicly available UTS datasets: KPI (Competition, 2018), TODS (Lai et al., 2021),

Table 1. Dataset Statistics.

Dataset	Curves	Train	Train Ano%	Test	Test Ano%
KPI	29	3,073,567	2.70%	3,073,556	1.85%
TODS	15	75,000	5.32%	75,000	6.38%
WSD	210	3,829,373	2.43%	3,829,537	0.76%
UCR	203	3,572,316	0.00%	7,782,539	0.47%

WSD (Zhang et al., 2022), and UCR (Wu & Keogh, 2021). Dataset characteristics are summarized in Table 1, including curve counts, sizes, and anomaly rates. The anomaly interval length distributions, shown in Figure 6, reveal that while most anomalies span less than 10 points, WSD and UCR contain extended anomaly segments exceeding 300 points, enabling comprehensive evaluation. Detailed dataset descriptions are provided in Appendix A.1.

#### 4.1.2. MODEL TRAINING AND INFERENCE

We implement a systematic experimental protocol for both our method and baseline approaches. For each time series, we train dedicated KAN-AD models using consistent hyperparameters: batch size 1024, learning rate 0.01, and maximum 100 epochs. The validation strategy varies by dataset, with UCR reserving 20% of training data and other datasets employing a 4:1:5 ratio for training, validation, and testing splits. To ensure fair comparison, we faithfully replicate all baseline methods following their original implementations and hyperparameter settings as specified in their respective papers. During inference, we standardize the batch size to 1 across all methods for comparable efficiency assessment. Results presented in Table 2 report means and standard deviations from five independent trials with different random seeds.

#### 4.1.3. BASELINES

We conducted comparative experiments with ten state-of-the-art time series anomaly detection methods: LSTMAD (Malhotra et al., 2015), FCVAE (Wang et al., 2024), SRCNN (Ren et al., 2019), FITS (Xu et al., 2024), TimesNet (Wu et al., 2023), OFA (Zhou et al., 2023), TranAD (Tuli et al., 2022), SubLOF (Breunig et al., 2000), Anomaly Transformer (Xu et al., 2022) (abbreviated as AnoTrans in the tables), KAN (Liu et al., 2025) and SAND (Boniol et al., 2021). Detailed descriptions of these methods can be found in Appendix A.2. For datasets not featured in the baseline literature, we meticulously tuned hyperparameters via grid search to optimize the performance of the baseline method on the respective evaluation metrics.

#### 4.1.4. EVALUATION METRICS

In practical applications, operations teams are less concerned with point-wise anomalies (i.e., whether individual data points are classified as anomalous) and more focused on detecting sustained anomalous segments within time se-

Table 2. Performance comparison. Best scores are highlighted in bold, and second best scores are highlighted in bold and underlined. Metrics include F1 (Best F1),  $F1_e$  (Event F1),  $F1_d$  (Delay F1), AUPRC (area under the precision-recall curve) and Avg  $F1_e$  (average  $F1_e$  score across four datasets).

Method	KPI				TODS				WSD				UCR				Avg $F1_e$
	F1	$F1_e$	$F1_d$	AUPRC													
SRCNN	0.4137	0.0994	0.2266	0.3355	0.6239	0.1918	0.4399	0.6076	0.4092	0.1185	0.1951	0.3080	0.5964	0.1369	0.1656	0.5109	0.1367
SAND	0.2710	0.0397	0.1097	0.2022	0.5372	0.1879	0.5103	0.5145	0.1761	0.0839	0.1267	0.1238	0.7044	<u>0.5108</u>	<u>0.5116</u>	0.6550	0.2056
AnoTrans	0.6103	0.3020	0.3623	0.5676	0.4875	0.1915	0.2918	0.4148	0.4348	0.2311	0.1517	0.3527	0.6135	0.1696	0.1084	0.5458	0.2236
TranAD	0.7553	0.5611	0.6399	0.7399	0.5035	0.2460	0.3619	0.4501	0.7570	0.6338	0.4158	0.7106	0.5278	0.1840	0.1554	0.4599	0.4062
SubLOF	0.7273	0.2805	0.4994	0.7015	0.7997	0.4795	0.7169	0.7809	0.8683	0.6585	0.4917	0.8353	<u>0.8468</u>	0.4772	0.4151	<u>0.8001</u>	0.4739
TimesNet	0.8022	0.6363	0.6995	0.8166	0.6232	0.3327	0.4495	0.6031	0.9406	0.8444	0.6170	0.9376	0.5273	0.1805	0.1439	0.4536	0.4985
FITS	0.9083	0.6353	0.8175	0.9359	0.7773	0.5416	0.6312	0.7725	0.9732	0.8391	0.6535	0.9771	0.6664	0.2926	0.2912	0.5969	0.5772
OFA	0.8810	0.6150	0.7952	0.9009	0.6928	0.5811	0.5588	0.7206	0.9564	0.8344	0.6250	0.9615	0.6294	0.3176	0.1503	0.5699	0.5870
FCVAE	0.9398	0.7556	0.8624	0.9572	<u>0.8652</u>	0.6995	0.7482	<u>0.8798</u>	0.9650	0.8610	0.6583	0.9653	0.7651	0.3812	0.2857	0.7145	0.6743
LSTMAD	0.9376	0.7742	<b>0.8782</b>	0.9624	0.8633	0.6981	0.7655	0.8740	0.9866	<b>0.9028</b>	<b>0.6743</b>	0.9849	0.7040	0.3482	0.3121	0.6432	0.6808
KAN	0.9411	<u>0.7816</u>	0.8666	0.9664	0.8109	0.6466	0.7518	0.8286	0.9879	0.8939	0.6650	<b>0.9881</b>	0.8016	0.4120	0.3971	0.7489	0.6835
KAN-AD	<b>0.9442</b>	<b>0.7989</b>	<u>0.8755</u>	<b>0.9693</b>	<b>0.9425</b>	<b>0.8940</b>	<b>0.8391</b>	<b>0.9716</b>	<b>0.9888</b>	0.8927	0.6623	0.9868	<b>0.8554</b>	<b>0.5335</b>	<b>0.5177</b>	<b>0.8188</b>	<b>0.7798</b>
	$\pm 0.0007$	$\pm 0.0054$	$\pm 0.0024$	$\pm 0.0008$	$\pm 0.0040$	$\pm 0.0022$	$\pm 0.0055$	$\pm 0.0035$	$\pm 0.0005$	$\pm 0.0025$	$\pm 0.0022$	$\pm 0.0009$	$\pm 0.0040$	$\pm 0.0046$	$\pm 0.0042$	$\pm 0.0041$	

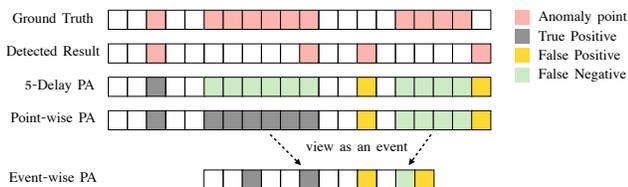


Figure 4. Illustration of the adjustment strategy. Point-wise PA gives an inflated score when some anomaly segments persist for a long duration. Event-wise PA treats each anomaly segment as an event, completely disregarding the length of the anomaly segment.  $k$ -delay PA considers only anomalies detected within the first  $k$  points after the anomaly onset, treating any detected later as undetected.

ries data. Furthermore, due to the potential impact of such segments, early identification is crucial. Previous work (Xu et al., 2018) proposed the **Best F1** metric, which iterates over all thresholds and applies a point adjustment strategy to calculate the F1 score. However, it has been criticized for performance inflation (Lai et al., 2021; Wu & Keogh, 2021).

To address this, we also adopt **Delay F1** (Ren et al., 2019) and **Event F1**. Delay F1 is similar to Best F1 but uses a delay point adjustment strategy. As shown in Figure 4, the second anomaly was missed because the detection delay exceeded the threshold of five time intervals. In all experiments, a delay threshold of five was used across all datasets. Event F1, on the other hand, treats anomalies of varying lengths as anomalies with a length of 1, minimizing performance inflation caused by excessively long anomalous segments. For the sake of convenience, unless otherwise stated, we use Event F1 as the primary metric, as it is more alignment with the need for real-time anomaly detection in real-world situations.

#### 4.2. RQ1. Performance and Efficiency Comparison

We present a comprehensive evaluation of KAN-AD across multiple time series anomaly detection (TSAD) experiments,

Table 3. Efficiency comparison on UCR dataset.

Method	GPU Time	CPU Time	Parameters	$F1_e$
SAND	-	5637s	-	0.5108
SubLOF	-	299s	-	0.4772
OFA	220s	3087s	81.920 M	0.3176
AnoTrans	201s	1152s	4.752 M	0.1696
FCVAE	2327s	1743s	1.414 M	0.3812
TimesNet	182s	259s	73,449	0.1805
LSTMAD	73s	267s	10,421	0.3482
KAN	66s	<u>34s</u>	1,360	0.4120
FITS	<b>32s</b>	<b>17s</b>	624	0.2926
TranAD	113s	62s	369	0.1840
KAN-AD	<u>42s</u>	36s	<b>274</b>	<b>0.5335</b>

with results summarized in Table 2. Our analysis focuses on three key dimensions: detection accuracy, model efficiency, and computational requirements. Across diverse experimental settings, KAN-AD demonstrates consistent and robust performance advantages. In the TODS dataset, where training data contains a substantial proportion of anomalies, KAN-AD significantly outperforms SOTA by 27% on Event F1, highlighting its robust learning capabilities in handling noisy training data. For datasets exhibiting strong periodic characteristics (WSD and KPI), KAN-AD achieves comparable or superior performance relative to state-of-the-art approaches. Even in the challenging UCR dataset scenario, where the training set lacks anomaly samples and contains significant periodic variations, KAN-AD effectively captures normal patterns, whereas baseline methods show reduced effectiveness in pattern recognition. Quantitatively, KAN-AD achieves more than a 15% improvement in average Event F1 score compared to existing state-of-the-art methods.

The computational efficiency analysis, presented in Table 3, reveals another distinctive advantage of KAN-AD. We note that several baseline methods are excluded from this comparison due to implementation constraints: SAND’s CPU-only execution requirement and SubLOF’s limited multi-core utilization capabilities preclude fair comparison in modern hardware acceleration contexts. Among the other mod-

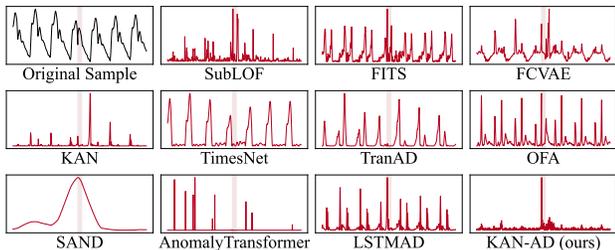


Figure 5. Case study on UCR InternalBleeding10. The black curve represents the original sample, the red curve represents the anomaly scores provided by the method, and the true anomaly segments are highlighted in pink.

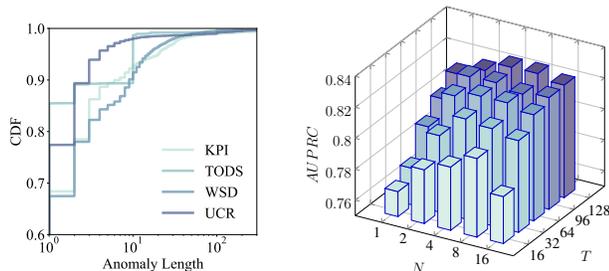


Figure 6. Anomalous lengths distribution.

Figure 7. Model performance under different hyperparameters.

els, we observe a wide spectrum of model complexities, with parameter counts ranging from millions to hundreds. Large-scale models like OFA utilize 81.92M parameters, while established approaches such as Anomaly Transformer, FCVAE, and TimesNet employ between 73k and 4.75M parameters. In contrast, KAN-AD achieves competitive performance with remarkable efficiency, requiring only 274 parameters, a 25% reduction compared to TranAD, the next most compact model in our comparison.

These empirical findings underscore KAN-AD’s exceptional efficiency-performance in TSAD tasks. By achieving state-of-the-art or near state-of-the-art performance while significantly reducing the parameter footprint, KAN-AD demonstrates the effectiveness of our design principles in creating efficient, practical solutions. This combination of high detection accuracy and minimal computational requirements positions KAN-AD as an ideal choice for resource-constrained or cost-sensitive deployment scenarios, offering a compelling balance between model complexity and detection capabilities.

#### 4.2.1. CASE STUDY

We analyzed anomaly detection performance on UCR dataset samples to illustrate how various methods respond to identical anomalies, as shown in Figure 5. The selected sample displayed pattern anomalies, marked by significant deviations from typical behavior. Both TranAD and TimesNet exhibit difficulty establishing normal patterns. Minor varia-

tions among normal samples across cycles lead to periodic false alarms during normal segments, consistent with our observations in Figure 2. Among the methods listed, while OFA, LSTMAD, SubLOF, and FITS can detect anomalies, their high anomaly scores during normal segments indicate excessive sensitivity to minor fluctuations in normal data. In contrast, KAN-AD excels in identifying anomalies while maintaining minimal anomaly scores during normal segments.

### 4.3. RQ2. Hyperparameter sensitivity

The KAN-AD model incorporates two key hyperparameters: the number of terms in univariate functions  $N$  and the window size  $T$ . To investigate the ultimate impact of these parameters on model performance, we conducted experiments on the UCR dataset while holding all other parameters constant. As findings summarized in Figure 7, a larger window size facilitates more accurate learning of normal patterns when  $N$  is fixed, leading to improved performance. When  $T$  is fixed, insufficient univariate functions limit KAN-AD’s expressive power, while excessive  $N$  can lead to overfitting. Overall, KAN-AD achieved its best performance with  $T = 96$  and  $N = 2$ . Notably, even with suboptimal hyperparameter settings like  $T = 16$  and  $N = 1$ , we surpassed SOTA methods on the UCR dataset.

### 4.4. RQ3. Ablation Studies

In this section, we investigated the impact of constant term elimination modules, different univariate function selections on algorithm performance and the influence of the function deconstruction mechanism.

#### 4.4.1. CONSTANT TERM ELIMINATION MODULE

We employed a constant term elimination (CTE) module during data preprocessing to mitigate the influence of the offset term  $A_0$  in Equation (3). Further experiments were conducted across all datasets to evaluate the impact of incorporating CTE within the preprocessing pipeline. As presented in Figure 8, the impact of CTE varies across datasets, reflecting inherent data characteristics. For datasets with pronounced periodicity or strong temporal stability (e.g., WSD), the benefits of CTE are less apparent. Conversely, for datasets exhibiting larger value fluctuations or trends (e.g., KPI, TODS and UCR), CTE yields significant improvements.

#### 4.4.2. SELECTION OF UNIVARIATE FUNCTIONS

To assess the impact of different univariate functions on model performance, we conducted experiments using common univariate functions listed in Table 4. In our implementations, due to varying input range requirements across

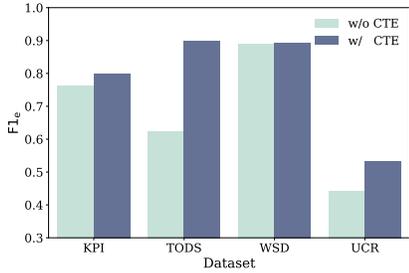


Figure 8. Model performance under different preprocessing.

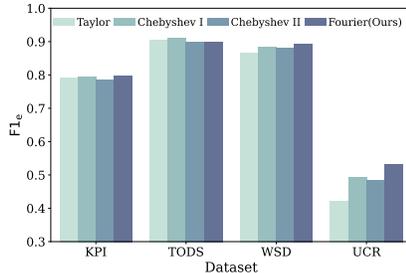


Figure 9. Model performance under different univariate function.

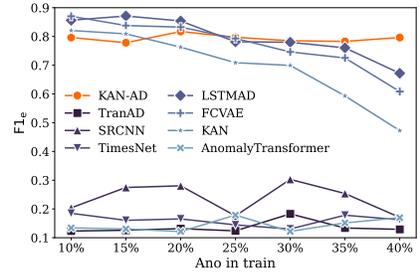


Figure 10. Model performance under different anomaly ratios in training.

Table 4. Commonly used univariate functions for time series approximation.

Name	$\Phi_n(x)$
Taylor Series	$x^n$
Fourier Series	$\cos(nx) + \sin(nx)$
Chebyshev Polynomial I	$\cos(n \arccos(x))$
Chebyshev Polynomial II	$\frac{\sin((n+1) \arccos(x))}{\sin(\arccos(x))}$

univariate functions, appropriate normalization techniques are employed. Specifically, min-max scaling to the range  $x \in [-1, 1]$  was utilized for both types of Chebyshev polynomials, while z-score was employed for Taylor series and Fourier series. The performance of all four univariate functions was compared using the same configuration. As results presented in Figure 9, Fourier series consistently achieved the top two performance across all datasets. In contrast, Taylor series exhibited persistent bias due to non-zero function values in most cases, hindering optimal model performance. The objective of both types of Chebyshev polynomials is to minimize the maximum error, which potentially conflicts with anomaly detection methods that minimize mean squared prediction error, thus leading to suboptimal performance.

#### 4.5. RQ4. Robustness to Anomalous Data

To evaluate KAN-AD’s robustness to anomalies in the training set, we conducted additional experiments using synthetic datasets constructed in accordance with the TODS dataset generation methodology. We synthesized test datasets containing local peaks and drops anomalies, and progressively increased the proportion of these anomalies in the initially anomaly-free training set. As illustrated in Figure 10, KAN-AD demonstrates stable performance across all anomaly ratios. Popular methods such as LSTMAD, perform well at lower anomaly ratios but experience a significant decline as the ratio increases. Other approaches, like TranAD, fail to achieve optimal performance due to overfitting to fine-grained structures within the time series.

Table 5. Model performance on UCR dataset under different function deconstruction strategies.

Variation	F1 <sub>e</sub>	F1 <sub>d</sub>	AUPRC
KAN-AD	0.5335	0.5177	0.8188
w/o X	0.5153	0.4974	0.8066
w/o P	0.5081	0.4810	0.8007
w/o S	0.5056	0.5113	0.7998
w/o X&P	0.4737	0.4583	0.7872
w/o X&S	0.4698	0.4610	0.7767
w/o S&P	0.4561	0.4637	0.7595

#### 4.6. Ablation on function deconstruction mechanism

To investigate the impact of the function deconstruction mechanism, we compared the model’s detection capabilities under different univariate function combination strategies. For clarity, the specific definitions are provided in Equation (5). As the results presented in Table 5, the model’s detection performance exhibited a notable improvement with an increasing number of univariate functions. Both Fourier series and cosine waves outperformed the raw input data, likely due to their smoother representations compared to the original signal, enabling higher detection accuracy. The combination of different features, particularly those involving Fourier series and cosine waves, resulted in significant performance gains as the feature count increased. Ultimately, KAN-AD achieved optimal detection performance by integrating all features. It is worth noting that even the variant of KAN-AD utilizing only the raw time series X outperforms KAN, clearly demonstrating the advantage of Fourier series over the use of spline functions for optimizing univariate functions.

#### 4.7. Performance on Multivariate Time Series

To extend KAN-AD’s application to the multivariate time series (MTS) scenario, we adopt a channel-independent approach. Specifically, an MTS input with the shape  $(batch\_size, window\_length, n\_features)$  is reshaped into  $(batch\_size * n\_features, window\_length)$ . Each of the  $n\_features$  channels

Table 6. Best F1 and parameter counts for multivariate time series anomaly detection. Best and second best results are in **bold** and underline.

Methods	SMD	MSL	SMAP	SWaT	PSM	Avg F1	Parameters@MSL
Informer (Zhou et al., 2021)	0.8165	0.8406	0.6992	0.8143	0.7710	0.7883	504,174
Anomaly Transformer (Xu et al., 2022)	0.8549	0.8331	0.7118	0.8310	0.7940	0.8050	4,863,055
DLinear (Zeng et al., 2023)	0.7710	0.8488	0.6926	0.8752	0.9355	0.8246	20,200
Autoformer (Wu et al., 2021)	0.8511	0.7905	0.7112	0.9274	0.9329	0.8426	325,431
FEDformer (Zhou et al., 2022)	0.8508	0.7857	0.7076	0.9319	0.9723	0.8497	1,119,982
TimesNet (Wu et al., 2023)	0.8462	0.8180	0.6950	0.9300	0.9738	0.8526	75,223
UniTS (Gao et al., 2024)	<b>0.8809</b>	0.8346	0.8380	0.9326	<b>0.9743</b>	0.8921	8,066,376
<b>KAN-AD (ours)</b>	0.8429	<b>0.8501</b>	<b>0.9450</b>	<b>0.9350</b>	0.9650	<b>0.9076</b>	<b>4,491</b>

is thus treated as an independent univariate time series instance. KAN-AD is then applied to these individual series. This channel-independent strategy has proven effective (Nie et al., 2023). By adopting a similar principle, KAN-AD can leverage its robust univariate modeling capabilities across all channels of an MTS dataset. The model is trained on the collection of these reshaped univariate instances, allowing it to learn generalized normal patterns.

We implemented MTS versions of KAN-AD in popular time series library (THUML) and evaluated them on the common SMD (Su et al., 2019), MSL (Hundman et al., 2018a), SMAP (Hundman et al., 2018b), SWaT (Mathur & Tippenhauer, 2016), and PSM (Abdulaal et al., 2021) datasets. Our evaluation metric uses the Best F1 score which is consistent with the baseline methods. We introduce these datasets and baseline methods in detail in the Appendix B. As detailed in Table 6, KAN-AD achieves the highest average Best F1 score of 0.9076, across all five benchmark datasets, outperforming all listed SOTA methods. A significant advantage of KAN-AD is its exceptional parameter efficiency. With only 4,491 trainable parameters (measured on MSL), KAN-AD utilizes substantially fewer parameters than all other compared methods.

## 5. Related Work

**Time Series Forecasting Methods:** These methods can be categorized into prediction-based and reconstruction-based methods, both aiming to identify deviations from normal patterns through temporal analysis. Prediction-based methods, like FITS (Xu et al., 2024) achieves efficient detection through frequency domain analysis with minimal parameters, while LSTMAD (Malhotra et al., 2015) leverages LSTM networks (Hochreiter & Schmidhuber, 1997) to capture complex temporal dependencies. Reconstruction-based approaches, like Donut (Xu et al., 2018) focus on time series denoising, while FCVAE (Wang et al., 2024) enhances the VAE (Kingma & Welling, 2022) framework by incorporating frequency domain information. Recent advances in Transformer architectures have further strengthened re-

construction capabilities: TranAD (Tuli et al., 2022) employs adversarial learning for robust pattern capture, while OFA (Zhou et al., 2023) leverages GPT-2 (Radford et al., 2019) for modeling complex temporal dependencies.

**Pattern Change Detection Methods:** These approaches identify anomalies through comparative analysis of current and historical patterns. Early methods, like SubLOF (Breunig et al., 2000) quantify pattern variations using window-based distance metrics. SAND employs temporal shape-based clustering to distinguish anomalous patterns. Recent advances, exemplified by TriAD (Sun et al., 2024), leverage multi-domain contrastive learning frameworks, demonstrating superior performance on UCR datasets.

## 6. Conclusion

Training time series anomaly detection models with datasets containing anomalies is essential for deployment in production environments. Existing algorithms often rely on carefully selected features and complex architectures to achieve minor accuracy gains, neglecting robustness during training. This paper introduces KAN-AD, a robust anomaly detection model rooted in the Kolmogorov–Arnold representation theorem. KAN-AD transforms the prediction of time points into the estimation of coefficients of Fourier series, achieving strong performance with few parameters, significantly reducing costs while enhancing robustness to outliers. KAN-AD includes a constant term elimination module to address temporal trends and leverages frequency domain information for better performance. KAN-AD surpasses the SOTA model across four public datasets with a 15% improvement in average Event F1 score, simultaneously achieving an 80% reduction in parameter count and 50% faster inference speed compared to vanilla KAN. With KAN-AD, a promising research direction is to explore whether normal patterns in time series can be represented more efficiently by leveraging additional data.

## Acknowledgments

This work was partially funded by the National Key Research and Development Program of China (No.2022YFB2901800), the National Natural Science Foundation of China (62202445), the National Natural Science Foundation of China-Research Grants Council (RGC) Joint Research Scheme (62321166652), and the National Natural Science Foundation of China (Grant No. W2412136).

## Impact Statement

This paper presents work whose goal is to advance the field of Machine Learning. There are many potential societal consequences of our work, none which we feel must be specifically highlighted here.

## References

- Abdulaal, A., Liu, Z., and 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.
- Bodner, A. D., Tepsich, A. S., Spolski, J. N., and Pourteau, S. Convolutional kolmogorov-arnold networks, 2024. URL <https://arxiv.org/abs/2406.13155>.
- Boniol, P., Paparrizos, J., Palpanas, T., and Franklin, M. J. Sand: streaming subsequence anomaly detection. *Proceedings of the VLDB Endowment*, 14(10): 1717–1729, 2021.
- Breunig, M. M., Kriegel, H.-P., Ng, R. T., and Sander, J. Lof: identifying density-based local outliers. In *Proceedings of the 2000 ACM SIGMOD international conference on Management of data*, pp. 93–104, 2000.
- Competition, A. Kpi dataset. <https://github.com/iopsai/iops>, 2018.
- De Boor, C. A practical guide to splines. *Springer-Verlag google schola*, 2:4135–4195, 1978.
- Dym, H. and HP, M. Fourier series and integrals. 1972.
- Gao, S., Koker, T., Queen, O., Hartvigsen, T., Tsiligkaridis, T., and Zitnik, M. Units: A unified multi-task time series model. *Advances in Neural Information Processing Systems*, 37:140589–140631, 2024.
- He, K., Zhang, X., Ren, S., and Sun, J. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.
- Hendrycks, D. and Gimpel, K. Gaussian error linear units (gelus). *arXiv preprint arXiv:1606.08415*, 2016.
- Hochreiter, S. and Schmidhuber, J. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.
- Hundman, K., Constantinou, V., Laporte, C., Colwell, I., and Soderstrom, T. Detecting spacecraft anomalies using lstms and nonparametric dynamic thresholding. In *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*, pp. 387–395, 2018a.
- Hundman, K., Constantinou, V., Laporte, C., Colwell, I., and Soderstrom, T. Detecting spacecraft anomalies using lstms and nonparametric dynamic thresholding. In *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*, pp. 387–395, 2018b.
- Ioffe, S. and Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning*, pp. 448–456. pmlr, 2015.
- Kieu, T., Yang, B., Guo, C., Cirstea, R.-G., Zhao, Y., Song, Y., and Jensen, C. S. Anomaly detection in time series with robust variational quasi-recurrent autoencoders. In *2022 IEEE 38th International Conference on Data Engineering (ICDE)*, pp. 1342–1354. IEEE, 2022.
- Kingma, D. P. and Welling, M. Auto-encoding variational bayes, 2022.
- Kolmogorov, A. N. On the representations of continuous functions of many variables by superposition of continuous functions of one variable and addition. In *Dokl. Akad. Nauk USSR*, volume 114, pp. 953–956, 1957.
- Lai, K.-H., Zha, D., Xu, J., Zhao, Y., Wang, G., and Hu, X. Revisiting time series outlier detection: Definitions and benchmarks. In *Thirty-fifth conference on neural information processing systems datasets and benchmarks track (round 1)*, 2021.
- Li, D., Chen, D., Jin, B., Shi, L., Goh, J., and Ng, S.-K. Madgan: Multivariate anomaly detection for time series data with generative adversarial networks. In *International conference on artificial neural networks*, pp. 703–716. Springer, 2019.
- Li, Z., Zhao, Y., Han, J., Su, Y., Jiao, R., Wen, X., and 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.

- Liu, Z., Wang, Y., Vaidya, S., Ruehle, F., Halverson, J., Soljačić, M., Hou, T. Y., and Tegmark, M. Kan: Kolmogorov-arnold networks. In The Thirteenth International Conference on Learning Representations, 2025.
- Malhotra, P., Vig, L., Shroff, G., Agarwal, P., et al. Long short term memory networks for anomaly detection in time series. In Esann, volume 2015, pp. 89, 2015.
- Mathur, A. P. and 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.
- Nie, Y., H. Nguyen, N., Sinthong, P., and Kalagnanam, J. A time series is worth 64 words: Long-term forecasting with transformers. In International Conference on Learning Representations, 2023.
- Qu, X., Liu, Z., Wu, C. Q., Hou, A., Yin, X., and Chen, Z. Mfgan: Multimodal fusion for industrial anomaly detection using attention-based autoencoder and generative adversarial network. Sensors, 24(2):637, 2024.
- Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., Sutskever, I., et al. Language models are unsupervised multitask learners. OpenAI blog, 1(8):9, 2019.
- Ren, H., Xu, B., Wang, Y., Yi, C., Huang, C., Kou, X., Xing, T., Yang, M., Tong, J., and Zhang, Q. Time-series anomaly detection service at microsoft. In Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining, pp. 3009–3017, 2019.
- Siffer, A., Fouque, P.-A., Termier, A., and Largouet, C. Anomaly detection in streams with extreme value theory. In Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining, pp. 1067–1075, 2017.
- Stein, E. M. and Shakarchi, R. Fourier analysis: an introduction, volume 1. Princeton University Press, 2011.
- Su, Y., Zhao, Y., Niu, C., Liu, R., Sun, W., and 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.
- Sun, Y., Pang, G., Ye, G., Chen, T., Hu, X., and Yin, H. Unraveling the anomaly in time series anomaly detection: A self-supervised tri-domain solution. In 2024 IEEE 40th International Conference on Data Engineering (ICDE). IEEE, 2024.
- Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A. Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1–9, 2015.
- THUML. thuml/time-series-library: A library for advanced deep time series models. URL <https://github.com/thuml/Time-Series-Library>.
- Tuli, S., Casale, G., and Jennings, N. R. Tranad: deep transformer networks for anomaly detection in multivariate time series data. Proceedings of the VLDB Endowment, 15(6):1201–1214, 2022.
- Wang, Y., Perry, M., Whitlock, D., and Sutherland, J. W. Detecting anomalies in time series data from a manufacturing system using recurrent neural networks. Journal of Manufacturing Systems, 62:823–834, 2022.
- Wang, Z., Pei, C., Ma, M., Wang, X., Li, Z., Pei, D., Rajmohan, S., Zhang, D., Lin, Q., Zhang, H., Li, J., and Xie, G. Revisiting vae for unsupervised time series anomaly detection: A frequency perspective. In Proceedings of the ACM on Web Conference 2024, WWW '24, pp. 3096–3105, New York, NY, USA, 2024. Association for Computing Machinery. ISBN 9798400701719. doi: 10.1145/3589334.3645710. URL <https://doi.org/10.1145/3589334.3645710>.
- Wu, H., Xu, J., Wang, J., and Long, M. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. Advances in Neural Information Processing Systems, 34:22419–22430, 2021.
- Wu, H., Hu, T., Liu, Y., Zhou, H., Wang, J., and Long, M. Timesnet: Temporal 2d-variation modeling for general time series analysis. In International Conference on Learning Representations, 2023.
- Wu, R. and Keogh, E. J. Current time series anomaly detection benchmarks are flawed and are creating the illusion of progress. IEEE transactions on knowledge and data engineering, 35(3):2421–2429, 2021.
- 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.
- Xu, J., Wu, H., Wang, J., and Long, M. Anomaly transformer: Time series anomaly detection with association discrepancy. In International Conference on Learning Representations, 2022. URL [https://openreview.net/forum?id=LzQQ89U1qm\\_](https://openreview.net/forum?id=LzQQ89U1qm_).

- Xu, Z., Zeng, A., and Xu, Q. Fits: Modeling time series with  $10k$  parameters. In International Conference on Learning Representations (ICLR), 2024.
- Yu, R., Yu, W., and Wang, X. Kan or mlp: A fairer comparison. arXiv preprint arXiv:2407.16674, 2024.
- Zeng, A., Chen, M., Zhang, L., and Xu, Q. Are transformers effective for time series forecasting? In Proceedings of the AAAI conference on artificial intelligence, volume 37, pp. 11121–11128, 2023.
- Zhan, P., Wang, S., Wang, J., Qu, L., Wang, K., Hu, Y., and Li, X. Temporal anomaly detection on iiot-enabled manufacturing. Journal of Intelligent Manufacturing, 32: 1669–1678, 2021.
- Zhang, S., Zhong, Z., Li, D., Fan, Q., Sun, Y., Zhu, M., Zhang, Y., Pei, D., Sun, J., Liu, Y., et al. Efficient kpi anomaly detection through transfer learning for large-scale web services. IEEE Journal on Selected Areas in Communications, 40(8):2440–2455, 2022.
- Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., and Zhang, W. Informer: Beyond efficient transformer for long sequence time-series forecasting. In Proceedings of the AAAI conference on artificial intelligence, volume 35, pp. 11106–11115, 2021.
- Zhou, T., Ma, Z., Wen, Q., Wang, X., Sun, L., and Jin, R. Fedformer: Frequency enhanced decomposed transformer for long-term series forecasting. In International conference on machine learning, pp. 27268–27286. PMLR, 2022.
- Zhou, T., Niu, P., Sun, L., Jin, R., et al. One fits all: Power general time series analysis by pretrained lm. Advances in neural information processing systems, 36: 43322–43355, 2023.

## A. Datasets and Baselines on Univariate Time Series

### A.1. Datasets

We selected four datasets from diverse domains, with samples originating from:

- **KPI** (Competition, 2018): This dataset comprises service metrics collected from five major Internet companies: Sogou, eBay, Baidu, Tencent, and Alibaba. The data points are primarily recorded every 1-2 minutes, with some sections exhibiting a 5-minute interval.
- **TODS** (Lai et al., 2021): TODS comprises artificially created time series, each designed to present specific types of anomalies. Its excellent interpretability and carefully constructed data distributions make it suitable for in-depth case studies.
- **WSD** (Zhang et al., 2022): This dataset consists of web server metrics collected from three companies providing large-scale web services: Baidu, Sogou, and eBay.
- **UCR** (Wu & Keogh, 2021): This archive contains data from multiple domains with a single anomalous segment on each time series. In addition to real anomalies, UCR also includes synthetic but highly plausible anomalies.

### A.2. Baselines

We selected the following baseline approaches to further elaborate on the performance differences between KAN-AD and SOTA methods:

- **SubLOF** (Breunig et al., 2000) represents traditional outlier detection techniques based on distance metrics.
- **SRCNN** (Ren et al., 2019) is a supervised approach reliant on high-quality labeled data.
- **LSTMAD** (Malhotra et al., 2015) leverages Long Short-Term Memory (LSTM) networks (Hochreiter & Schmidhuber, 1997) for deep learning-based anomaly detection.
- **FITS** (Xu et al., 2024) achieves parameter-efficient anomaly detection by upsampling frequency domain information using a low-pass filter and simple linear layers.
- **FCVAE** (Wang et al., 2024) is unsupervised reconstruction method based on Variational Autoencoder (VAE) (Kingma & Welling, 2022), designed to reconstruct normal patterns.
- **Anomaly Transformer** (Xu et al., 2022) employs attention mechanism to compute the association discrepancy.
- **TranAD** (Tuli et al., 2022) incorporates the principles of adversarial learning to develop a training framework with two stages while integrating the strengths of self-attention encoders to capture the temporal dependency embedded in the time series.
- **SAND** (Boniol et al., 2021) utilizes a novel statistical approach based on curve shape clustering for anomaly detection in a streaming fashion.
- **TimesNet** (Wu et al., 2023) leverages an Inception (Szegedy et al., 2015)-based computer vision backbone to enhance learning capabilities.
- **OFA** (Zhou et al., 2023), with GPT-2 (Radford et al., 2019) as its backbone, improves its ability to capture point-to-point dependencies.
- **KAN** (Liu et al., 2025) leverages Kolmogorov-Arnold representation theory to decompose complex learning objectives into linear combinations of univariate functions.

These baseline methods encompass a variety of anomaly detection paradigms: shape-based SAND, subsequence distance-based SubLOF, Transformer-based approaches like OFA, TranAD, and Anomaly Transformer for modeling sequence relationships, and frequency domain information enhanced methods FCVAE and FITS.

## B. Datasets and Baselines on Multivariate Time Series

### B.1. Datasets

We evaluated KAN-AD on five widely-used public benchmark datasets for multivariate time series anomaly detection:

- **SMD** (Su et al., 2019): A dataset collected from a large internet company, containing data from many server machines over several weeks.
- **MSL** (Hundman et al., 2018a): A dataset from NASA containing telemetry data from the Mars Science Laboratory rover.
- **SMAP** (Hundman et al., 2018b): Another NASA dataset, containing telemetry data from the SMAP satellite.
- **SWaT** (Mathur & Tippenhauer, 2016): A dataset generated from a scaled-down real-world water treatment testbed, including normal and attack scenarios.
- **PSM** (Abdulaal et al., 2021): A dataset from eBay, consisting of aggregated metrics from multiple application servers.

### B.2. Baselines

We selected the following baseline approaches to further evaluate KAN-AD and SOTA methods on multivariate time series datasets:

- **Informer** (Zhou et al., 2021): A Transformer-based model designed for long sequence time-series forecasting, featuring a ProbSparse self-attention mechanism to improve efficiency. For anomaly detection, it typically relies on reconstruction error or forecast error.
- **Anomaly Transformer** (Xu et al., 2022): A Transformer architecture specifically tailored for time series anomaly detection, which aims to learn prior-associations and series-associations to better distinguish anomalies.
- **DLinear** (Zeng et al., 2023): A simple yet effective linear model that decomposes the time series into trend and remainder components, challenging the necessity of complex Transformer architectures for some forecasting tasks and adaptable for anomaly detection via reconstruction.
- **Autoformer** (Wu et al., 2021): A Transformer model with a novel decomposition architecture and an Auto-Correlation mechanism, designed to discover series-wise connections and improve long-term forecasting accuracy.
- **FEDformer** (Zhou et al., 2022): A Transformer variant that enhances performance for long sequence forecasting by employing frequency-enhanced decomposition and a mixture of expert design in the frequency domain.
- **TimesNet** (Wu et al., 2023): A model that transforms 1D time series into a set of 2D tensors based on identified periods and applies a 2D kernel (e.g., Inception block) to capture both intra-period and inter-period variations for general time series analysis.
- **UniTS** (Gao et al., 2024): Aims to provide a unified framework for time series analysis, often leveraging large-scale pre-training on diverse datasets to build a universal representation for both univariate and multivariate time series tasks.