TF-ChebKAN: Time-Frequency Coupled Chebyshev Kolmogorov-Arnold Networks for Multivariate Time Series Anomaly Detection

Feng Wang[1,2,3], Xiaoheng Wang[3,4], Shuaipeng Wu[1,3], Kejiang Ye[3,2,*]
[1]Southern University of Science and Technology, Shenzhen, China
[2]Shenzhen University of Advanced Technology, Shenzhen, China
[3]Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China
[4]University of Chinese Academy of Sciences, Beijing, China
{f.wang3, xh.wang5, sp.wu1, kj.ye}@siat.ac.cn

***Abstract*—Multivariate time series anomaly detection requires a balance between model capability and computational efficiency. Many existing methods struggle to simultaneously capture complex nonlinear patterns and multivariate dependencies. In this paper, we propose a lightweight two-stream architecture, TF-ChebKAN, which integrates a Chebyshev II-enhanced Kolmogorov-Arnold network to address the boundary instability problem commonly encountered in multivariate time series anomaly detection. Furthermore, we use a multivariate frequency mixer to capture global cross-channel correlations by processing spectral components with linear complexity. Finally, an adaptive fusion mechanism dynamically adjusts the balance between the two modules. Experiments on MSL, PSM, SMD and SWaT show that TF-ChebKAN achieves strong performance, while using fewer parameters than baselines. Furthermore, TF-ChebKAN significantly reduces training and inference times, demonstrating efficient anomaly detection capabilities in resource-limited scenarios.**

***Index Terms*—Frequency Domain, Kolmogorov-Arnold Networks, Multivariate Time Series Anomaly Detection**

I. Introduction

Multivariate time series (MTS) anomaly detection is an essential component in modern industrial and cyber-physical systems, such as spacecraft monitoring, server maintenance, industrial detection, and network platforms [1]–[3]. However, effectively modeling complex relationships between variables under real-time processing and limited computational resources remains a major challenge.

There are two opposing paradigms in the field of MTS anomaly detection. On the one hand, Transformer-based architectures such as Autoformer [4] and iTransformer [5] are good at capturing global dependencies but suffer from high computational cost and prone to overfitting on small datasets. On the other hand, lightweight linear models such as DLinear [6] and LightTS [7] are efficient but suffer from poor performance in capturing complex nonlinear patterns. In recent years, Kolmogorov-Arnold Network (KAN) [8] has emerged as a promising approach, leveraging learnable activation functions to approximate nonlinear mappings. KAN-AD [9] was the first to use KAN in the field of temporal anomaly detection by using Fourier coefficients as learnable parameters.

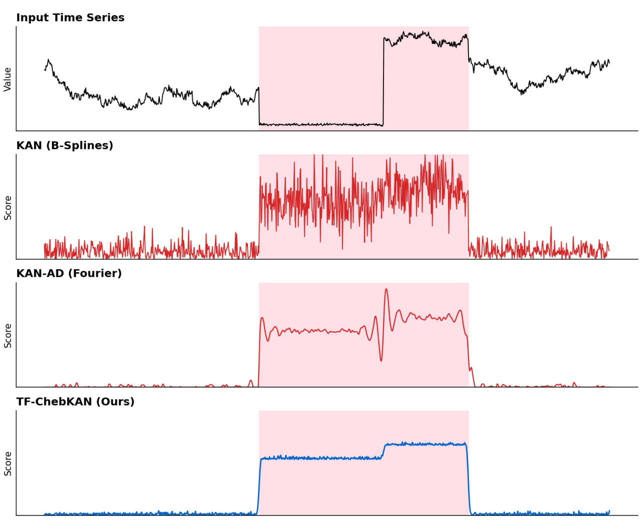


Fig. 1: Comparison of anomaly detection scores on a time series with sudden transitions. Standard KAN (B-spline) exhibits high-frequency noise due to overfitting. KAN-AD (Fourier basis) shows Gibbs phenomenon at the boundaries. TF-ChebKAN reaches stability at the boundaries.

However, as shown in Fig.1, while KAN-AD achieves efficient anomaly detection, two drawbacks hinder its performance in complex MTS scenarios. (1) Boundary instability: KAN-AD relies on Fourier basis functions, which are prone to the Gibbs phenomenon, exhibiting anomalous fluctuations near boundaries. In anomaly detection, the ability to fit these boundaries is crucial for immediate decision-making, and this instability can lead to severe false alarms. (2) Channel isolation: KAN-AD processes each variable individually, ignoring potential connections between sensors. This independent strategy weakens system-level modeling. Recent studies [10] demonstrate that modeling cross-channel dependencies is essential to distinguish anomalies from isolated noise.

* Corresponding author

To bridge these gaps, we propose **TF-ChebKAN** (Time-Frequency Coupled Chebyshev Kolmogorov-Arnold Network), a novel two-stream architecture combining the advantages of time-domain and frequency-domain processing. First, we use Chebyshev II polynomials as basis functions which provide better approximation accuracy at data boundaries to reduce false alarms in real-time detection. Second, we introduce a multivariate frequency mixer (MFM) to capture missing inter-variable dependencies. By directly interacting with real and imaginary frequency components, MFM can simulate complex relationships between sensors without incurring the high computational cost of attention mechanisms.

Our main contributions are summarized as follows:

- We propose a Chebyshev II-enhanced KAN encoder to mitigate boundary instability. By integrating Chebyshev polynomials with frequency features, it enables accurate modeling of complex temporal variations.
- We design a multivariate frequency mixer that overcomes the channel isolation problem in existing KAN-based methods. It captures global cross-channel correlations by processing spectral components with linear complexity.
- We design an adaptive fusion mechanism that can dynamically balance local time flow and global frequency flow, enabling our model to adapt to diverse anomaly patterns.
- Extensive experiments show that TF-ChebKAN achieves competitive performance compared to state-of-the-art baselines. Remarkably, it requires orders of magnitude fewer parameters than Transformer-based methods while significantly reducing training time, offering an efficient paradigm for multivariate time series anomaly detection.

II. Related Work

A. Multivariate Time Series Anomaly Detection

Traditional multivariate time series anomaly detection methods primarily include statistical models (e.g., PCA [11], COPOD [12], HBOS [13]) and density-based algorithms (e.g., LOF [14]). Additionally, classical machine learning methods such as IForest [15] and OCSVM [16] are widely used. However, limited by the representation capability, these methods struggle to capture complex temporal dependencies and nonlinear features in high-dimensional long sequences [17], [18].

In recent years, deep learning-based approaches have gained widespread attention [19], [20], such as OmniAnomaly [21], USAD [22] and GDN [10]. With the rise of Transformers, Anomaly Transformer utilizes an association discrepancy mechanism to capture the differences in attention distribution between anomalies and normal patterns. Furthermore, models like Informer [23] and Autoformer [4] have significantly improved long-sequence modeling through enhanced attention mechanisms. Despite the dramatic advances in detection capabilities, attention mechanisms based on Transformers still consume a large amount of training and inference resources.

Recent research suggests that complex architectures are not always necessary for time series data. Lightweight linear models such as DLinear [6] and LightTS [7] have achieved excellent performance in multiple benchmark tests. However, to improve efficiency, these methods often adopt channel-independent strategies [24], ignoring dependencies between variables and sacrificing the nonlinear modeling capabilities which are necessary to capture anomalous patterns in complex systems. Therefore, designing an architecture that explicitly captures multivariate dependencies while maintaining computational efficiency remains a significant challenge.

B. Frequency Domain Analysis in Time-Series

While time-domain methods can capture local variations and trend anomalies, frequency-domain analysis has a significant advantage in detecting subsequence anomalies (such as periodic fluctuations or oscillation patterns) that are difficult to identify in time-domain [25]. In recent years, frequency-based methods have received increasing attention. SR-CNN [26] was the first to use frequency domain saliency maps for anomaly detection, but it was limited to univariate scenarios and failed to address the complexity of multivariate time series.

In terms of multi-scale modeling, TFAD [25] uses time-frequency transformation for multivariate anomaly detection; TimesNet [27] extracts the period through frequency analysis and processes it with two-dimensional convolution; TimeMixer [28] mixes information of different resolutions in the time domain. However, most of these methods treat frequency information merely as auxiliary features.

Existing frequency domain methods still face two limitations: (1) Most methods process the spectrum of each variable independently, ignoring the phase coupling and amplitude correlations inherent in multivariate time series; (2) Some methods introduce complex lifting operations, contradicting the goal of lightweight design.

C. Kolmogorov-Arnold Networks

The Kolmogorov–Arnold representation theorem [29] states that any multivariate continuous function can be represented as a superposition of univariate functions. Based on this, KANs [8] achieve higher parameter efficiency by replacing fixed activation functions in MLPs with learnable activation functions on edges (typically parameterized as B-splines).

TKAN [30] first applied KAN to time series forecasting, leveraging its adaptive function properties to capture nonlinear temporal patterns. However, directly using B-spline basis functions tends to overfit local features, failing to eliminate the impact of local peaks or drops. To introduce KAN into anomaly detection, KAN-AD [9] uses Fourier series as basis functions to capture global features. However, while Fourier bases excel at handling periodicity, they are prone to the Gibbs phenomenon when dealing with non-stationary sudden changes, producing spurious ringing artifacts near local mutations in normal data, which leads to false alarms.

Existing KAN-related methods suffer from two main limitations: (1) Boundary Approximation Instability: Existing B-spline or Fourier basis functions lack the Minimax property at interval boundaries, where instability easily leads to false positives; (2) Lack of Multivariate Spectral Interaction: Most

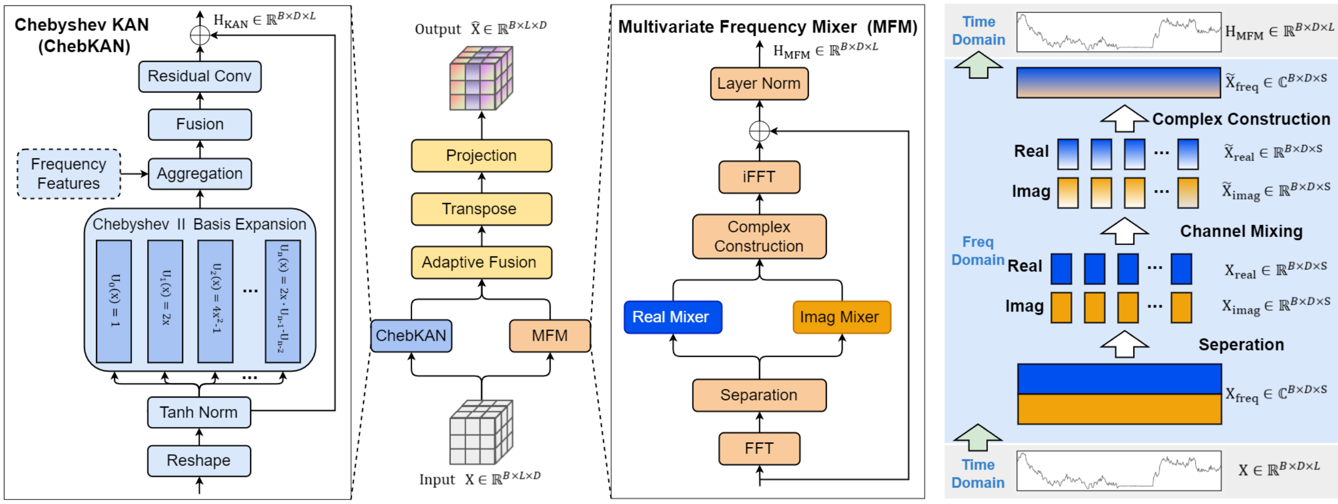


Fig. 2: TF-ChebKAN overall architecture. TF-ChebKAN consists of a temporal stream (ChebKAN) and a frequency stream (MFM). The temporal stream models local univariate dynamics, while the frequency stream captures global cross-channel dependencies in spectral space. Their outputs are adaptively fused to produce the final prediction.

KAN variants are either univariate or perform simple mixing only in the time domain, lacking a mechanism to explicitly capture inter-variable dependencies within the spectral domain.

III. Methodology

A. Problem Formulation

We consider the unsupervised multivariate time series anomaly detection task formulated as a next-step forecasting problem. Let $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_T)$ denote a multivariate time series, where $\mathbf{x}_t \in \mathbb{R}^D$ is the observation vector at time step t. For each time step t, we define a sliding window $\mathcal{W}_t = (\mathbf{x}_{t-L+1}, \dots, \mathbf{x}_t) \in \mathbb{R}^{L \times D}$ of length L. Our objective is to learn a mapping function $f_\theta : \mathbb{R}^{L \times D} \to \mathbb{R}^D$ that predicts the subsequent observation $\hat{\mathbf{x}}_{t+1}$ based on the historical window $\mathcal{W}_t$. The model is trained to minimize the prediction error on normal data:

$$\mathcal{L} = \sum_t \|\mathbf{x}_{t+1} - f_\theta(\mathcal{W}_t)\|_2^2 \tag{1}$$

During inference, we identify anomalies by the deviation between predicted and actual values. The anomaly score at time t is defined as $s_t = \|\mathbf{x}_t - \hat{\mathbf{x}}_t\|_2^2$, where $\hat{\mathbf{x}}_t$ is the prediction generated from the preceding window $\mathcal{W}_{t-1}$.

B. Theoretical Foundation

When applying the KAN-based methods for time series anomaly detection, we discovered two theoretical bases:

1) **Boundary Instability:** The accuracy of time series anomaly detection is highly dependent on the latest timestamp (sliding window boundary). Traditional polynomial bases and Fourier series, due to the effects of the Runge or Gibbs phenomenon, exhibit severe oscillations at interval boundaries.
2) **Minimax Approximation:** We propose using a second-type Chebyshev polynomial (U_n). Theoretical analysis shows that Chebyshev nodes minimize the Lebesgue constant, providing an optimal uniform approximation (minimax property), particularly at the interval endpoints ± 1. This is crucial to our task, ensuring that our model's predictions free of spurious oscillations.

C. Overall Architecture

The overall architecture of TF-ChebKAN is shown in Fig.2. We first stabilize the multivariate time series window through input normalization to ensure the data distribution matches the Chebyshev definition domain. The normalized series then undergoes univariate unfolding and is fed into the Chebyshev II-enhanced KAN encoder, which performs independent temporal modeling on each variable using orthogonal polynomial bases augmented with frequency features. Simultaneously, the original series is input into the multivariate frequency mixer, which transforms data into the frequency domain and captures cross-variable correlations through spectral channel mixing. Subsequently, the outputs from these two complementary streams are integrated via adaptive fusion to balance local temporal dynamics and global multivariate dependencies. Finally, the fused representation is mapped to the output space via a projection layer to generate the next-step prediction. TF-ChebKAN effectively combines the efficiency of channel-independent processing in the temporal stream with the expressiveness of channel-mixing in the frequency stream.

D. Temporal Stream: Chebyshev KAN

This stream captures univariate temporal patterns.

1) Input Normalization: A key premise of orthogonal polynomial approximation is ensuring that the input distribution lies within the domain $[-1, 1]$. Industrial sensor data often exhibits large variations and non-stationary trends, which can lead to saturation of bounded activation functions. Therefore, we apply Tanh normalization before basis projection. For each instance $x \in \mathbb{R}^{L\times D}$, we apply the hyperbolic tangent function:

$$\tilde{x}_{t,d} = \tanh(x_{t,d}) \tag{2}$$

This activation maps the data to the open interval $(-1, 1)$, ensuring the input optimally activates the linear response region of the Chebyshev polynomials.

2) Chebyshev Encoder: For each variable d, we approximate its temporal function using a learnable combination of Chebyshev polynomials $U_n(\tilde{x})$:

$$\begin{aligned} U_0(\tilde{x}) &= 1, \\ U_1(\tilde{x}) &= 2\tilde{x}, \\ U_2(\tilde{x}) &= 4\tilde{x}^2 - 1, \\ U_n(\tilde{x}) &= 2\tilde{x}U_{n-1}(\tilde{x}) - U_{n-2}(\tilde{x}) \end{aligned} \tag{3}$$

The polynomial basis is concatenated with frequency features $F_{freq} \in \mathbb{R}^{2\times L}$ extracted via FFT. To capture complex nonlinear dynamics, the fused representation undergoes deep feature refinement. Specifically, we apply univariate unfolding to reshape the input from $\mathbb{R}^{B\times L\times D}$ to $\mathbb{R}^{(B\cdot D)\times L}$ and process it through a multi-layer one-dimensional convolutional network with residual connections and Batch Normalization. This parameter-sharing design ensures that the temporal modeling capacity is decoupled from the variable dimension D, achieving $O(C)$ complexity per variable.

E. Frequency Stream: Multivariate Mixer

While the temporal stream processes variables independently, this stream explicitly models inter-variable dependencies in the spectral domain.

1) Spectral Channel Mixing: The input $X \in \mathbb{R}^{B\times L\times D}$ (without tanh normalization) is first transposed to channel-first format $X \in \mathbb{R}^{B\times D\times L}$ and then transformed to the frequency domain via Real FFT, yielding $X_{freq} \in \mathbb{C}^{B\times D\times S}$, where $S = \lfloor L/2 \rfloor + 1$. We decompose this into real and imaginary components and apply separate non-linear mixing transformations:

$$\begin{aligned} \tilde{X}_{real} &= \sigma(\text{Conv1d}(X_{real})), \\ \tilde{X}_{imag} &= \sigma(\text{Conv1d}(X_{imag})) \end{aligned} \tag{4}$$

where Conv1d denotes a 1×1 convolution layer acting as a learnable mixing matrix $W \in \mathbb{R}^{D\times D}$ shared across all frequency bins, and σ represents the GELU activation function followed by Dropout. This design allows the model to capture amplitude and phase correlations among sensors efficiently.

2) Frequency Reconstruction: The processed spectral features are recombined ($\tilde{X}_{freq} = \tilde{X}_{real} + j \cdot \tilde{X}_{imag}$) and transformed back to the time domain via Inverse Real FFT. Finally, we apply a learnable residual connection and Layer Normalization:

$$H_{MFM} = \text{LayerNorm}(X + \lambda \cdot \text{IRFFT}(\tilde{X}_{freq})) \tag{5}$$

where λ is a learnable scalar initialized to 0.5.

TABLE I: Benchmark details. AR represents the anomaly rate.

Dataset	Dimension	Training	Validation	Testing	AR(%)
MSL	55	46,653	11,663	73,729	10.53%
PSM	25	105,984	26,496	87,841	27.76%
SMD	38	566,724	141,681	708,420	4.16%
SWaT	51	378,719	94,679	449,919	12.14%

F. Adaptive Fusion

The outputs from the temporal stream (H_{KAN}) and frequency stream (H_{MFM}) are fused via a learnable weight α (initialized to 0.5):

$$H_{fused} = H_{KAN} + \alpha \cdot H_{MFM} \tag{6}$$

This mechanism allows the model to dynamically balance the contributions of univariate temporal patterns and multivariate dependencies. The fused representation is directly used as the prediction, transposed to match the input format $\mathbb{R}^{B\times L\times D}$. Since the channel dimension is preserved throughout both streams, no additional projection parameters are required, further contributing to the model's efficiency.

G. Optimization and Anomaly Criterion

Optimization: To train the model, we employ the Mean Squared Error (MSE) loss function. For a training batch of size B and window length L, the loss is computed as:

$$\mathcal{L} = \frac{1}{B \cdot L \cdot D} \sum_{b,t,d} (x_{b,t,d} - \hat{x}_{b,t,d})^2 \tag{7}$$

Anomaly Criterion: During inference, we utilize the point-wise prediction error averaged across dimensions as the anomaly score. For a test observation $\mathbf{x}_t$, the score is:

$$s_t = \frac{1}{D} \|\mathbf{x}_t - \hat{\mathbf{x}}_t\|_2^2 \tag{8}$$

A higher score indicates a larger deviation from normal patterns, signaling a potential anomaly.

IV. Experiments

A. Experimental Settings

Datasets. In the experiment, we use four different multivariate time series datasets to evaluate our model performance, as shown in Table I. Here is the description of these datasets:

- MSL (Mars Science Laboratory): Telemetry data from NASA's Mars rovers, including sensor and control signals [31].
- PSM (Pooled Server Metrics): Monitoring metrics from eBay's server cluster, reflecting the status of a large-scale distributed system [32].
- SMD (Server Machine Dataset): Recording operational data on the resource utilization (such as CPU and memory) of multiple servers in a large internet company [21].
- SWaT (Secure Water Treatment): Collected from a real water treatment plant testbed, including physical state data under various simulated cyberattacks [33].

TABLE II: Performance comparison on four real-world datasets

Dataset	MSL			PSM			SMD			SWaT		
Metric	**V-ROC**	**V-PR**	**Aff-F1**	**V-ROC**	**V-PR**	**Aff-F1**	**V-ROC**	**V-PR**	**Aff-F1**	**V-ROC**	**V-PR**	**Aff-F1**
OCSVM	0.6013	0.2027	0.7185	0.4814	0.2910	0.6999	0.7864	0.3007	0.8625	0.6622	0.4638	0.7184
LOF	0.5997	0.1666	0.6863	0.4669	0.3077	0.6943	0.7453	0.1690	0.8495	0.5306	0.1585	0.7321
PCA	0.5812	0.1652	0.7086	0.5601	0.3609	0.7126	0.7908	0.3393	0.7960	0.7057	0.2666	0.6932
IForest	0.5420	0.1418	0.7120	0.4616	0.3314	0.6943	0.7657	0.2331	0.7663	0.6750	0.3640	0.6931
HBOS	0.6546	0.2206	0.7225	0.5565	0.3520	0.7090	0.8001	0.2512	0.7965	0.7292	0.4097	0.7128
COPOD	0.5817	0.1639	0.7105	0.4952	0.3200	0.7037	0.7811	0.1988	0.7774	0.7913	0.4136	0.7160
OmniAnomaly	0.6254	0.2312	0.6823	0.7003	**0.5511**	0.7229	0.7987	0.3676	0.8392	**0.8662**	0.6391	0.6925
Autoformer	0.6422	0.2019	0.7140	0.6662	0.4988	0.7396	0.7868	0.3583	0.8307	0.6787	0.5165	0.7209
FEDformer	0.6478	0.2051	0.7061	0.6703	0.5011	0.7604	0.7907	0.3619	0.8301	0.6918	0.4963	0.7227
AnomalyTrans	0.5568	0.1592	0.6788	0.5708	0.3772	0.6947	0.5117	0.0917	0.7184	0.4599	0.1263	0.7000
LightTS	0.6841	0.2218	0.7276	0.6585	0.4633	0.8117	0.8021	0.3822	0.7955	0.8036	0.3633	**0.8219**
DLinear	0.7085	0.2305	0.7299	0.6403	0.4425	0.8107	0.7884	0.3713	0.7939	0.2455	0.0829	0.7028
TimesNet	0.7014	**0.2338**	0.7297	0.6920	0.5284	0.8378	0.8300	0.3380	0.8527	0.2808	0.1090	0.7345
iTransformer	0.6787	0.2316	0.7237	0.6781	0.5026	0.8389	0.8025	0.3825	0.8444	0.2574	0.0871	0.7201
TimeMixer	**0.7158**	0.2309	0.7274	0.6715	0.4950	0.8341	0.8327	0.3984	0.8425	0.2794	0.1104	0.7382
KAN-AD	0.5798	0.1974	**0.7428**	0.6609	0.4541	0.7998	0.8192	0.3610	0.8530	0.6588	0.5023	0.7241
TF-ChebKAN	0.6718	0.2279	0.7312	**0.7054**	0.5495	**0.8399**	**0.8482**	**0.4204**	**0.8804**	0.8413	**0.6685**	0.7735

Baselines. We compare TF-ChebKAN with 16 baselines. These baselines include classic outlier detection algorithms: IForest [15], LOF [14], OCSVM [16]; statistical approaches: PCA [11], COPOD [12], HBOS [13]; representative Transformer variants: Autoformer [4], FEDformer [34], iTransformer [5]; advanced multi-scale architectures: TimesNet [27], TimeMixer [28]; Lightweight linear methods: DLinear [6], LightTS [7]; deep reconstruction-based methods: OmniAnomaly [21], Anomaly Transformer [35]. Finally, we specifically compare our model with KAN-AD [9] to demonstrate the efficacy of our architecture.

Metrics. We use score-based and label-based metrics to evaluate the anomaly detection performance of the models. For score-based metrics, recent research in TSB-AD [36] shows that VUS-PR [37] is the most robust, accurate and fair evaluation metric. Therefore, in our main results, we use Volume Under the Surface (VUS), including VUS-ROC and VUS-PR, to compare detection results. To ensure comprehensiveness, we also employ Area under the ROC Curve (AUC-ROC) [38] and Area under the Precision-Recall Curve (AUC-PR) [39] and Range-AUC metrics [40]. For label-based metrics, we report the Affiliation F1 score (Aff-F1) [41] to measure the alignment quality between predicted anomaly fragments and real anomaly fragments. Additionally, while the Point Adjustment F1 score (PA-F1) [42] is known to potentially overestimate detection performance, we still include it in our evaluation as it remains a widely used benchmark in the field.

Implementation Details. All experiments are conducted using PyTorch in Python 3.13, running on an NVIDIA Tesla-V100-32GB GPU. The input window length L is set to 96 for all datasets, consistent with standard benchmarks. For the Chebyshev II-enhanced KAN encoder, we use Second-kind Chebyshev polynomials with order $N = 2$. The frequency features extract via FFT have dimension 2 (real and imaginary components). The adaptive fusion weight α and the residual weight λ in the Multivariate Frequency Mixer are both initialized to 0.5. We employ the Adam optimizer with an initial learning rate of 10^{-2} and a step-wise learning rate scheduler (step size=5, decay factor=0.75). The batch size is set to 128, and models are trained for a maximum of 100 epochs, with early stopping based on validation loss (patience=3). The Mean Squared Error (MSE) loss is used for training.

B. Detection Results

Main results. We evaluate TF-ChebKAN on 16 competitive baselines across four real-world datasets, as shown in Table II. Across all datasets, our method achieves top-tier performance in VUS-PR, VUS-ROC, and Affiliation F1 metrics, demonstrating its ability to effectively distinguish between normal and abnormal samples while maintaining high detection accuracy and stability. These results validate the effectiveness of our model architecture.

Performance on Comprehensive Metrics. To ensure a comprehensive comparison, we evaluate the models using additional metrics: Point Adjustment F1 score (PA-F1), Affiliation F1 score (Aff-F1), AUC-ROC, AUC-PR, Range-AUC-ROC (R-AUC-ROC), Range-AUC-PR (R-AUC-PR), VUS-ROC and VUS-PR. Specifically, we compare TF-ChebKAN with three advanced models: OmniAnomaly, TimesNet, and TimeMixer. The study shows that TF-ChebKAN consistently achieves superior performance across multiple metrics, demonstrating its superior anomaly detection capabilities.

TABLE III: Performance on comprehensive metrics

Dataset	Method	PA-F1	Aff-F1	AUC-ROC	AUC-PR	R-AUC-ROC	R-AUC-PR	VUS-ROC	VUS-PR
MSL	OmniAnomaly	0.4257	0.6823	0.6030	**0.1734**	0.5216	0.1429	0.6254	0.2312
	TimesNet	0.8212	0.7297	0.6044	0.1385	**0.5956**	0.1390	0.7014	**0.2338**
	TimeMixer	0.8810	0.7274	0.2230	0.0785	0.2229	0.0783	**0.7158**	0.2309
	TF-ChebKAN	**0.8873**	**0.7312**	0.5744	0.1539	0.5740	**0.1488**	0.6718	0.2279
PSM	OmniAnomaly	0.9066	0.7229	**0.7492**	0.5250	0.6350	0.4635	0.7003	**0.5511**
	TimesNet	0.9779	0.8378	0.5993	0.3946	0.5877	0.3876	0.6920	0.5284
	TimeMixer	0.9800	0.8341	0.5886	0.3846	0.5740	0.3774	0.6715	0.4950
	TF-ChebKAN	**0.9838**	**0.8399**	0.6801	**0.5284**	**0.6480**	**0.4704**	**0.7054**	0.5495
SMD	OmniAnomaly	0.7737	0.8392	0.8212	**0.4577**	0.8169	0.3000	0.7987	0.3676
	TimesNet	0.9167	0.8527	**0.8701**	0.3962	**0.8648**	0.3725	0.8300	0.3380
	TimeMixer	0.9209	0.8425	0.8210	0.3942	0.8179	0.3689	0.8327	0.3984
	TF-ChebKAN	**0.9403**	**0.8804**	0.8139	0.4270	0.8037	**0.4029**	**0.8482**	**0.4204**
SWaT	OmniAnomaly	0.8325	0.6925	0.8039	0.6887	0.7662	0.6391	**0.8662**	0.6391
	TimesNet	0.9167	0.7345	0.2416	0.0939	0.2401	0.0914	0.2808	0.1090
	TimeMixer	0.9136	0.7382	0.2414	0.0954	0.2404	0.0931	0.2794	0.1104
	TF-ChebKAN	**0.9583**	**0.7735**	**0.8353**	**0.6654**	**0.8316**	**0.6615**	0.8413	**0.6685**

TABLE IV: Ablation study on different variations.

Variation	**MSL**			**PSM**			**SMD**			**SWaT**		
	V-ROC	**V-PR**	**Aff-F1**	**V-ROC**	**V-PR**	**Aff-F1**	**V-ROC**	**V-PR**	**Aff-F1**	**V-ROC**	**V-PR**	**Aff-F1**
w/o FreqMixer	0.6696	0.2126	0.7143	0.6712	0.5362	0.8221	0.8276	0.4016	0.8746	0.7581	0.5990	0.7586
w/o Freq Domain	0.6221	0.1950	0.7149	0.6863	0.5295	0.8267	0.8377	0.4200	0.8754	0.7376	0.5798	0.7472
w/o Adaptive Fusion	0.6662	0.2094	0.7155	0.6806	0.5290	0.8135	0.8375	0.4164	0.8739	0.7909	0.6635	0.7659
w/o ChebKAN	0.6432	0.2039	0.7163	0.6664	0.4820	0.8253	0.8297	0.3974	0.8642	0.7070	0.5888	0.7512
TF-ChebKAN	**0.6718**	**0.2279**	**0.7312**	**0.7054**	**0.5495**	**0.8399**	**0.8482**	**0.4204**	**0.8804**	**0.8413**	**0.6685**	**0.7735**

C. Model Analysis

Ablation studies. To validate the effectiveness of TF-ChebKAN's components, we conduct an ablation study on the unique components and mechanisms of TF-ChebKAN (Table IV). Removing the Multivariate Frequency Mixer ("w/o FreqMixer") affects the model's capability to model inter-metric relationships, resulting in decreased detection performance on most datasets. Performing channel mixing in the time domain rather than the frequency domain ("w/o Freq Domain") results in significant performance degradation, especially on the SWaT dataset, indicating that the frequency domain is a more informative space for modeling global correlations. Removing the adaptive fusion mechanism ("w/o Adaptive Fusion") affects performance consistently, verifying the necessity of dynamically aligning temporal and frequency information instead of fixed summation. Lastly, replacing the Chebyshev KAN encoder with a standard MLP ("w/o ChebKAN") causes significant instability in the feature extraction step, degrading performance by constraining the model's capacity to fit complex temporal variations.

Furthermore, we investigate the impact of different basis functions in the temporal modeling stream, as shown in Fig.3. We compare our Chebyshev II basis against standard B-splines, Fourier series and Chebyshev I. The results consistently demonstrate that the Chebyshev II basis outperforms others across all metrics. While Fourier basis excels at capturing periodicity, it suffers from boundary instability due to Gibbs phenomenon. B-splines, although flexible, tend to overfit high-frequency noise. In contrast, Chebyshev polynomials provide superior approximation stability at boundaries and robustness to noise, validating their selection as the optimal basis for time series anomaly detection.

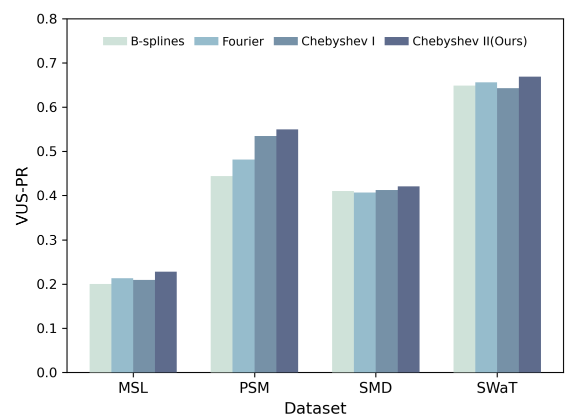


Fig. 3: Model performance under different basis functions.

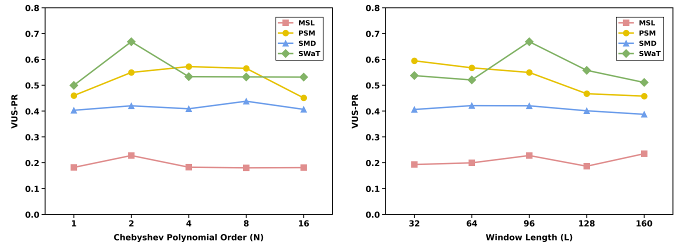
Model efficiency. We compare the efficiency of different time series anomaly detection methods on the SWaT dataset. As shown in Table V, while Linear-based methods like DLinear and LightTS generally exhibit higher efficiency than complex Transformer-based architectures, our proposed TF-ChebKAN achieves the most significant reduction in computational cost. Specifically, TF-ChebKAN possesses the lowest parameter count (only 5,721 parameters), which is orders of magnitude smaller than other models and notably more compact than the related KAN-AD. In terms of speed, TF-ChebKAN achieves the fastest training time and maintains a highly competitive inference speed, which is comparable to the most efficient method, DLinear. Unlike other methods where performance gains often necessitate higher model complexity, TF-ChebKAN demonstrates that superior detection performance can be achieved with a significantly lightweight architecture.

TABLE V: Model efficiency comparison on SWaT dataset.

Method	Training Time(s)	Inference Time(s)	Params	VUS-PR
Autoformer	864.34	157.04	323,379	0.5165
FEDformer	21723.73	997.92	853,734	0.4963
AnomalyTrans	5528.32	215.86	4,854,859	0.1263
LightTS	530.70	33.70	24,428	0.3633
DLinear	332.86	**23.17**	20,200	0.0829
TimesNet	1154.17	207.84	443,283	0.1090
iTransformer	1603.70	56.70	324,836	0.0871
TimeMixer	3414.54	132.55	316,049	0.1104
KAN-AD	652.89	23.53	6,551	0.5023
TF-ChebKAN	**248.73**	28.02	**5,721**	**0.6685**

D. Parameter Sensitivity Analysis

As shown in Fig.4, we comprehensively analyze the sensitivity of TF-ChebKAN to two key hyperparameters across four datasets. First, regarding the Chebyshev polynomial order (N), performance (VUS-PR) generally improves as N increases from 1 to 2, indicating that higher-order polynomials better capture intricate and complex non-linear dynamics. However, further increasing N beyond 3 yields diminishing returns or slight degradation due to potential overfitting of high-frequency noise; thus, $N = 2$ is selected as the optimal setting to successfully balance expressiveness and computational efficiency. Second, the model demonstrates significant robustness across varying input window lengths (L), with performance remaining highly stable for both short and long-range contexts. A window length of 96 consistently achieves optimal or near-optimal results across most datasets, effectively capturing long-term dependencies without introducing excessive computational cost, and is therefore formally adopted for all experiments. Overall, TF-ChebKAN is robust within the appropriate hyperparameter settings. This indicates that the model can maintain stable anomaly detection capabilities in different scenarios without careful parameter adjustment



(a) Chebyshev polynomial order (b) Window length

Fig. 4: Sensitivity analysis of key hyperparameters on detection performance (VUS-PR) across four datasets.

V. Conclusion

In this paper, we propose TF-ChebKAN, a lightweight dual-stream architecture that redefines the efficiency-performance trade-off in multivariate time series anomaly detection. Through the synergy of a Chebyshev II-enhanced KAN encoder and a Multivariate Frequency Mixer, our model successfully mitigates the boundary instability of Fourier-based methods and the channel isolation of univariate approaches. Our empirical assessment has shown that TF-ChebKAN achieves competitive performance across a variety of advanced evaluation metrics. Most importantly, it uses only 5,721 parameters, which is an order of magnitude improvement over state-of-the-art Transformer baselines, while simultaneously achieving a training speedup of orders of magnitude. Such extreme parameter efficiency makes TF-ChebKAN a practical solution for the industrial IoT edge, where computational overhead is a significant bottleneck.

Acknowledgement

This work is supported by the National Key R&D Program of China (No. 2025YFE0204100), Science and Technology Development Fund of Macao S.A.R (FDCT) under number 0074/2025/AMJ, Guangdong Basic and Applied Basic Research Foundation (No. 2023B1515130002), Shenzhen Basic Research Program (No. KJZD20230923113800001).

References

[1] A. Xiang, Z. Huang, X. Gao, K. Ye, and C.-z. Xu, "Bridgenet: A unified multimodal framework for bridging 2d and 3d industrial anomaly detection," in *Proceedings of the 33rd ACM International Conference on Multimedia*, 2025, pp. 1579–1587.

[2] C. Lu, K. Ye, W. Chen, and C.-Z. Xu, "Adgs: Anomaly detection and localization based on graph similarity in container-based clouds," in *2019 IEEE 25th International Conference on Parallel and Distributed Systems (ICPADS)*. IEEE, 2019, pp. 53–60.

[3] P. Lin, K. Ye, and C.-Z. Xu, "Netdetector: an anomaly detection platform for networked systems," in *2019 IEEE International Conference on Real-time Computing and Robotics (RCAR)*. IEEE, 2019, pp. 69–74.

[4] H. Wu, J. Xu, J. Wang, and M. Long, "Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting," *Advances in neural information processing systems*, vol. 34, pp. 22 419–22 430, 2021.

[5] Y. Liu, T. Hu, H. Zhang, H. Wu, S. Wang, L. Ma, and M. Long, "itransformer: Inverted transformers are effective for time series forecasting," in *The Twelfth International Conference on Learning Representations*, 2024.

[6] A. Zeng, M. Chen, L. Zhang, and Q. Xu, "Are transformers effective for time series forecasting?" in *Proceedings of the AAAI conference on artificial intelligence*, vol. 37, no. 9, 2023, pp. 11 121–11 128.

[7] D. Campos, M. Zhang, B. Yang, T. Kieu, C. Guo, and C. S. Jensen, "Lightts: Lightweight time series classification with adaptive ensemble distillation," *Proceedings of the ACM on Management of Data*, vol. 1, no. 2, pp. 1–27, 2023.

[8] Z. Liu, Y. Wang, S. Vaidya, F. Ruehle, J. Halverson, M. Soljacic, T. Y. Hou, and M. Tegmark, "Kan: Kolmogorov–arnold networks," in *The Thirteenth International Conference on Learning Representations*, 2025.

[9] Q. Zhou, C. Pei, F. Sun, J. Han, Z. Gao, H. Zhang, G. Xie, D. Pei, and J. Li, "Kan-ad: Time series anomaly detection with kolmogorov-arnold networks," in *Forty-second International Conference on Machine Learning (ICML)*, ser. Proceedings of Machine Learning Research, 2025.

[10] A. Deng and B. Hooi, "Graph neural network-based anomaly detection in multivariate time series," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 35, no. 5, 2021, pp. 4027–4035.

[11] S. Mei-Ling, "A novel anomaly detection scheme based on principal component classifier," in *Proc. of ICDM Foundation and New Direction of Data Mining workshop, 2003*, 2003.

[12] Z. Li, Y. Zhao, N. Botta, C. Ionescu, and X. Hu, "Copod: copula-based outlier detection," in *2020 IEEE international conference on data mining (ICDM)*. IEEE, 2020, pp. 1118–1123.

[13] M. Goldstein and A. Dengel, "Histogram-based outlier score (hbos): A fast unsupervised anomaly detection algorithm," *KI-2012: poster and demo track*, vol. 1, pp. 59–63, 2012.

[14] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, "Lof: identifying density-based local outliers," in *Proceedings of the 2000 ACM SIGMOD international conference on Management of data*, 2000, pp. 93–104.

[15] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation forest," in *2008 eighth ieee international conference on data mining*. IEEE, 2008, pp. 413–422.

[16] B. 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, 1999.

[17] M. Zhu, K. Ye, Y. Wang, and C.-Z. Xu, "A deep learning approach for network anomaly detection based on amf-lstm," in *IFIP international conference on network and parallel computing*. Springer, 2018, pp. 137–141.

[18] P. Lin, K. Ye, and C.-Z. Xu, "Dynamic network anomaly detection system by using deep learning techniques," in *International conference on cloud computing*. Springer, 2019, pp. 161–176.

[19] M. Zhu, K. Ye, and C.-Z. Xu, "Network anomaly detection and identification based on deep learning methods," in *International Conference on Cloud Computing*. Springer, 2018, pp. 219–234.

[20] S. Ji, T. Sun, K. Ye, W. Wang, and C.-Z. Xu, "Dafl: deep adaptive feature learning for network anomaly detection," in *IFIP International Conference on Network and Parallel Computing*. Springer, 2019, pp. 350–354.

[21] Y. Su, Y. Zhao, C. Niu, R. Liu, W. Sun, and D. Pei, "Robust anomaly detection for multivariate time series through stochastic recurrent neural network," in *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, 2019, pp. 2828–2837.

[22] J. Audibert, P. Michiardi, F. Guyard, S. Marti, and M. A. Zuluaga, "Usad: Unsupervised anomaly detection on multivariate time series," in *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*, 2020, pp. 3395–3404.

[23] H. Zhou, S. Zhang, J. Peng, S. Zhang, J. Li, H. Xiong, and W. Zhang, "Informer: Beyond efficient transformer for long sequence time-series forecasting," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 35, no. 12, 2021, pp. 11 106–11 115.

[24] Y. Nie, N. H. Nguyen, P. Sinthong, and J. Kalagnanam, "A time series is worth 64 words: Long-term forecasting with transformers," in *The Eleventh International Conference on Learning Representations*, 2023.

[25] C. Zhang, T. Zhou, Q. Wen, and L. Sun, "Tfad: A decomposition time series anomaly detection architecture with time-frequency analysis," in *Proceedings of the 31st ACM international conference on information & knowledge management*, 2022, pp. 2497–2507.

[26] H. Ren, B. Xu, Y. Wang, C. Yi, C. Huang, X. Kou, T. Xing, M. Yang, J. Tong, and Q. Zhang, "Time-series anomaly detection service at microsoft," in *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, 2019, pp. 3009–3017.

[27] H. Wu, T. Hu, Y. Liu, H. Zhou, J. Wang, and M. Long, "Timesnet: Temporal 2d-variation modeling for general time series analysis," in *The Eleventh International Conference on Learning Representations*, 2023.

[28] S. Wang, H. Wu, X. Shi, T. Hu, H. Luo, L. Ma, J. Y. Zhang, and J. Zhou, "Timemixer: Decomposable multiscale mixing for time series forecasting," in *The Twelfth International Conference on Learning Representations*, 2024.

[29] A. N. Kolmogorov, "On the representations of continuous functions of many variables by superposition of continuous functions of one variable and addition," in *Dokl. Akad. Nauk USSR*, vol. 114, 1957, pp. 953–956.

[30] R. Genet and H. Inzirillo, "Tkan: Temporal kolmogorov-arnold networks," arXiv preprint arXiv:2405.07344, 2024.

[31] K. Hundman, V. Constantinou, C. Laporte, I. Colwell, and T. Soderstrom, "Detecting spacecraft anomalies using lstms and nonparametric dynamic thresholding," in *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*, 2018, pp. 387–395.

[32] A. Abdulaal, Z. Liu, and T. Lancewicki, "Practical approach to asynchronous multivariate time series anomaly detection and localization," in *Proceedings of the 27th ACM SIGKDD conference on knowledge discovery & data mining*, 2021, pp. 2485–2494.

[33] A. P. Mathur and N. O. Tippenhauer, "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)*. IEEE, 2016, pp. 31–36.

[34] T. Zhou, Z. Ma, Q. Wen, X. Wang, L. Sun, and R. Jin, "Fedformer: Frequency enhanced decomposed transformer for long-term series forecasting," in *International conference on machine learning*. PMLR, 2022, pp. 27 268–27 286.

[35] J. Xu, H. Wu, J. Wang, and M. Long, "Anomaly transformer: Time series anomaly detection with association discrepancy," in *The Tenth International Conference on Learning Representations*, 2022.

[36] Q. Liu and J. Paparrizos, "The elephant in the room: Towards a reliable time-series anomaly detection benchmark," *Advances in Neural Information Processing Systems*, vol. 37, pp. 108 231–108 261, 2024.

[37] J. Paparrizos, P. Boniol, T. Palpanas, R. S. Tsay, A. Elmore, and M. J. Franklin, "Volume under the surface: a new accuracy evaluation measure for time-series anomaly detection," *Proceedings of the VLDB Endowment*, vol. 15, no. 11, pp. 2774–2787, 2022.

[38] T. Fawcett, "An introduction to roc analysis," *Pattern recognition letters*, vol. 27, no. 8, pp. 861–874, 2006.

[39] K. Boyd, K. H. Eng, and C. D. Page, "Area under the precision-recall curve: point estimates and confidence intervals," in *Joint European conference on machine learning and knowledge discovery in databases*. Springer, 2013, pp. 451–466.

[40] N. Tatbul, T. J. Lee, S. Zdonik, M. Alam, and J. Gottschlich, "Precision and recall for time series," *Advances in neural information processing systems*, vol. 31, 2018.

[41] A. Huet, J. M. Navarro, and D. Rossi, "Local evaluation of time series anomaly detection algorithms," in *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2022, pp. 635–645.

[42] H. Xu, W. Chen, N. Zhao, Z. Li, J. Bu, Z. Li, Y. Liu, Y. Zhao, D. Pei, Y. Feng *et al.*, "Unsupervised anomaly detection via variational auto-encoder for seasonal kpis in web applications," in *Proceedings of the 2018 world wide web conference*, 2018, pp. 187–196.