Asia Pacific Science Press

Asia Pacific Economic and Management Review
ISSN(O):3005-9275
ISSN(P):3005-9267
Vol. 3 No.5 (2026)

Adaptive Multi-Scale Anomaly Detection for Cross-Sector Financial Time Series via End-to-End Wavelet-Attention Fusion

Mengting Zhang , Junxiang Lu*

Xi'an Polytechnic University, Xi'an, Shaanxi, 710600, China

Corresponding author: Junxiang Lu, junxianglu@xpu.edu.cn

Copyright: 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY-NC 4.0), permitting distribution and reproduction in any medium, provided the original author and source are credited, and explicitly prohibiting its use for commercial purposes.

Abstract: Traditional two-stage pipelines for financial anomaly detection suffer from a critical yet often overlooked flaw: the wavelet decomposition and the downstream detector are optimized independently. When a fixed Haar filter encounters the fat-tail shocks typical in A-share sectors, gradient flow breaks at the decomposition boundary, and the detector receives suboptimal frequency segments. We introduce EWAF-AE, an autoencoder that embeds discrete wavelet transform as a learnable layer, allowing the spectral filters to co-train with the encoder. Instead of a homogeneous backbone, we split the encoding: a shallow Transformer handles the low-frequency trend branch, while a masked locality-sensitive attention module processes high-frequency details. A gating network then calibrates the fusion ratio per sector—nonferrous metals receive heavier trend weighting (0.888), whereas banking shifts toward detail sensitivity (0.252)—based on rolling volatility and a learnable sector embedding. The anomaly score is not a single MSE, but a weighted composite of pixel-level, trend-consistency, detail-sparsity, and spectral losses, augmented by a triplet margin loss on pseudo-anomalous samples. On six A-share sector indices from 2015 to 2025, EWAF-AE hits 83.3% of benchmark crisis events, compared with 25.0% for Isolation Forest and 41.7% for LSTM-AE.

Keywords: Financial Time Series; Anomaly Detection; Differentiable Wavelet; Sector Heterogeneity; Adaptive Fusion

Published: Nov 4, 2026

DOI: https://doi.org/10.62177/apemr.v3i5.1855

1. Introduction

Financial anomaly detection rests on three implicit assumptions that collapse under real market conditions. The first is stationarity. ARIMA [1] and GARCH [2] calibrate their parameters on historical windows, yet when a liquidity crisis restructures the entire return distribution—as observed in our A-share dataset during February 2024—these models continue to forecast against a distribution that no longer exists. Distance-based detectors [3-5] uffer similarly: if the whole sector moves synchronously, the “local outlier” becomes a global constant, and the distance metric loses discriminative power.

The second assumption is that the temporal domain suffices. RNNs and CNNs [6,7] encode series as single-scale latent manifolds, which empirically conflates long-term commodity cycles with intraday flash crashes. Transformer variants [8-10] itigate this via self-attention, but their attention maps operate exclusively on time steps; in our preliminary tests on computer-sector log-returns, the attention weights spread uniformly across low-frequency trend periods, drowning the high-frequency spikes that preceded the October 2024 policy surge. Contrastive frameworks [11,12] arpen the decision boundary, yet they still

compress trend and detail into a shared embedding space, forcing the model to represent two statistically distinct regimes with a single vector geometry.

The third assumption is that decomposition and detection can be decoupled. Pre-trained wavelet filters or multi-scale period decomposition [13-16] do separate frequency components, but they treat this as preprocessing: the wavelet basis is fixed before the detector sees the data, and the fusion weights are hand-tuned or uniformly concatenated [17]. In our experiments, a fixed Haar decomposition followed by a standard LSTM-AE achieves only 33.3% event hit rate on nonferrous metals, because the filter cannot adapt its frequency boundary when the sector shifts from a low-volatility regime to a commodity super-cycle. The survey by [18] confirms that cross-sector volatility heterogeneity remains unresolved when fusion strategies ignore rolling market states and sector identity. These three failures—stationarity illusion, single-scale compression, and stage separation—motivate the end-to-end joint optimization proposed in this paper.

2. Related Theories and Techniques

Reconstruction-based anomaly detection presumes that the latent manifold of normal returns is compact and well-separated from anomalies [19-22] This presumption holds for stationary industrial sensor data, yet in our A-share setting, the "normal" manifold itself deforms across regimes: a bullish trend and a bearish volatility spike occupy fundamentally different regions of the latent space, and a single-scale bottleneck forces the encoder to average these geometries into a blurred compromise. We observed this empirically when a standard LSTM-AE reconstructed the October 2024 surge with low error, precisely because it learned to treat extreme policy-driven jumps as within-distribution noise.

Wavelet analysis offers a natural escape from this single-scale trap. The Mallat algorithm [23] splits a series into trend (approximation) and detail coefficients, and Daubechies wavelets [24] have been the conventional choice in financial econometrics for their compact support. But this choice introduces a subtle epistemic problem: the analyst must decide the decomposition depth and wavelet family before the detector is trained. In our ablation experiments, a fixed db4 basis with two-level decomposition achieved strong trend separation on nonferrous metals yet completely missed the micro-fluctuations in computer-sector flash crashes—because the frequency boundary was carved in stone before the model saw the data. Zhang and Benveniste [25] bridged wavelets and neural networks, yet their parameterization remained frozen; Mallat [26] envisioned adaptive filter banks, but without end-to-end differentiability, the filters cannot follow the gradient of the detection loss. Differentiable wavelet layers have since emerged for image denoising and mechanical fault diagnosis, yet the temporal irregularity and fat-tail noise of financial returns demand a distinct architectural treatment.

Attention mechanisms [27,28] provide the complementary capability of dynamic focus, yet their marriage with wavelet decomposition remains shallow in existing work. The core issue is not that current methods lack either wavelets or attention, but that the two operate in isolation: decomposition hands over static coefficients, and attention processes them as if they were raw time steps. This disconnect creates three friction points that we address in this paper. First, the gradient of the reconstruction loss never reaches the wavelet filters, so the frequency split cannot adapt to the detector's needs. Second, the same encoder architecture processes both slow-moving trends and abrupt jumps, diluting the representational specificity of each. Third, once fusion weights are set—whether by concatenation or fixed averaging—they ignore the sector's current volatility regime. The following section describes how EWAF-AE dissolves these frictions by making the wavelet, the encoder, and the gate jointly trainable.

3. End-to-End Wavelet-Attention Fusion Autoencoder

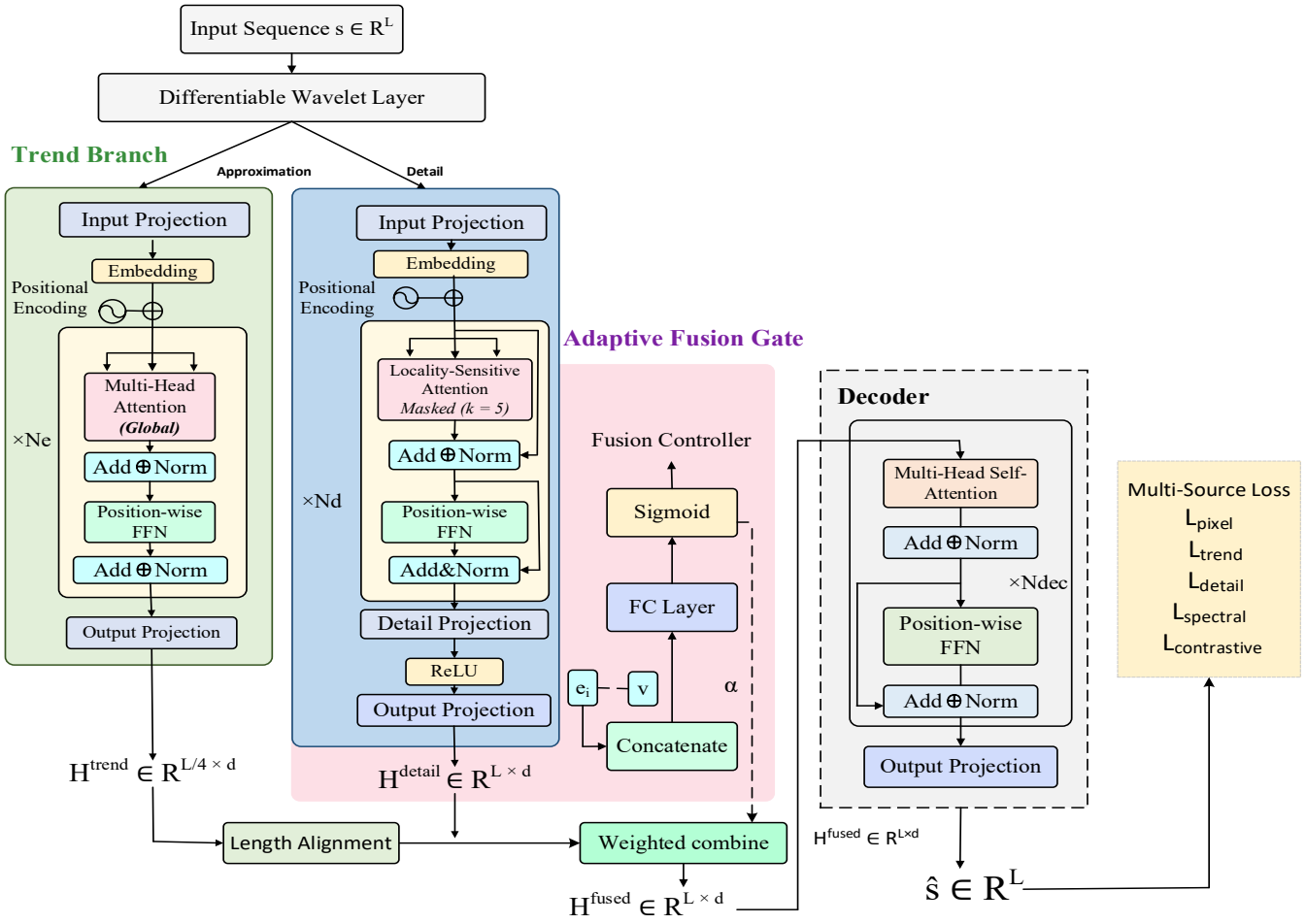
3.1 Problem Formulation and Differentiable Decomposition Architecture

Let $\chi = \{x_t\}_{t=1}^{T}$ denote a financial time series， where $x_t \in R$ represents the logarithmic return on the t -th trading day.. The sample set $S = \{s_i\}_{i=1}^{N}$ is constructed via a sliding-window strategy, where each sample is defined as $s_i = [x_i, x_{i+1}, ..., x_{i+L-1}]^{\mathrm{T}} \in R^{L}$ The window length is set to $L = 64$, the stride is 1 day, and the total number of samples is given by $N = T - L + 1$.We frame anomaly detection as an unsupervised reconstruction task on sliding windows of length $L = 64$ (roughly three months of A-share trading days). The intuition is straightforward: a well-trained model should compress normal return patterns into a tight latent manifold, whereas anomalous windows—such as the February 2024

liquidity crisis or the October policy surge—should explode the reconstruction error. In practice, this intuition collapses under two-stage pipelines. When the wavelet decomposition is frozen before training, the gradient of the detection loss never reaches the filter bank; the filters remain stuck at their initialization while the downstream encoder optimizes against frequency segments that may be convenient for signal processing yet misleading for anomaly discrimination. Worse, once the trend-detail fusion weights are hand-tuned for one sector, they become blind to another: a 0.7:0.3 split that captures nonferrous metals' long cycles is disastrous for banking's abrupt policy shocks.

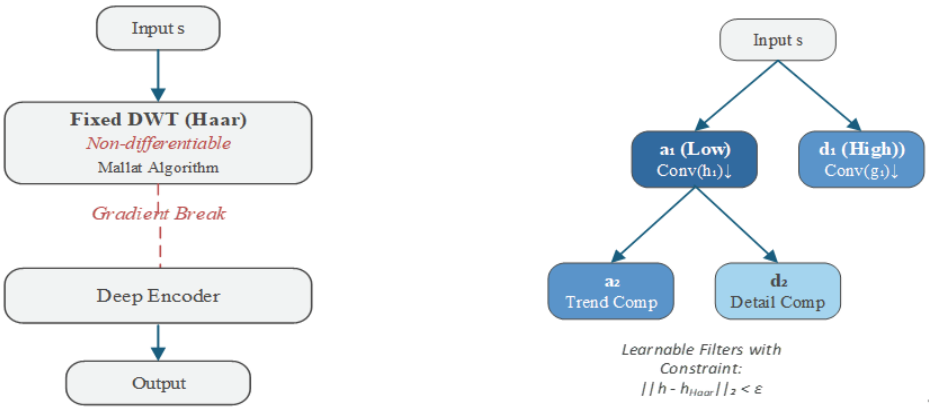
EWAF-AE removes these barriers by unrolling the entire pipeline into a single differentiable graph. The Discrete Wavelet Transform is no longer a preprocessing script but a neural layer whose low-pass and high-pass kernels co-train with the encoder. A lightweight Transformer digests the trend branch, while a masked locality-sensitive attention module isolates abrupt spikes in the detail branch. A gating network then calibrates the fusion ratio per sector based on rolling volatility and a learnable industry embedding, rather than relying on a fixed global average. The anomaly score is not a single pixel-wise MSE; it aggregates four distinct reconstruction losses and a triplet margin loss that pushes pseudo-anomalous samples away from the normal centroid. Because all components share one end-to-end objective, the wavelet filters, the branch encoders, and the fusion gate reshape themselves jointly to minimize reconstruction on normal windows and maximize it on extremes.

Figure 1: Architecture of the proposed EWAF-AE



To address the decoupling limitation between wavelet analysis and deep learning mechanisms in existing research, the discrete wavelet transform (DWT) is reconstructed as a differentiable network layer, through which joint optimization of decomposition parameters and encoding parameters is realized. Specifically, two-dimensional convolution operations are employed to simulate the Mallat pyramidal decomposition algorithm; low-pass filters and high-pass filters are initialized to Haar wavelet coefficients, and fine-tuning via backpropagation is enabled during the training process, subject to $\left\| h - h_{Haar} \right\|_2 < \varepsilon (\varepsilon = 0.1)$.

Figure 2: Comparison between traditional two-stage non-differentiable wavelet decomposition (a) and differentiable wavelet (b).



(a) (b)

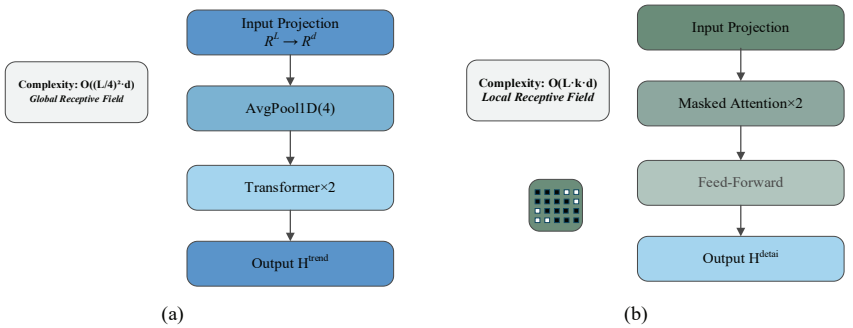
As illustrated in Figure 2(a), traditional two-stage non-differentiable wavelet decomposition relies on a fixed Haar discrete wavelet transform to decompose the input signal. The core design lies in sharing gradients between wavelet filters and the subsequent encoder, enabling the decomposition strategy to adaptively adjust frequency boundaries according to the downstream detection task.

To address the aforementioned limitation, this paper proposes a differentiable wavelet decomposition architecture, as illustrated in Figure 2(b). For an input sequence $s \in R^L$, the j-level decomposition is implemented via a convolution-downsampling operation: $a_{j+1} = Conv1D(a_j, h_j) \downarrow 2$, $d_{j+1} = Conv1D(a_j, g_j) \downarrow 2$, where $a_0 = s$, $\downarrow 2$ and $\downarrow 2$ denotes downsampling with a stride of 2. Following $J = 2$ layers of decomposition, approximation coefficients $a_J \in R^{L/4}$ and detail coefficients $\{d_j\}_{j=1}^{J}$ are obtained. The trend component s^{trend} and the detail component s^{detail} are subsequently reconstructed via an inverse transformation, both maintaining the original sequence length L. The core of this design lies in the fact that the wavelet filter parameters $\{h_j, g_j\}$ share the gradient of the loss function with the subsequent encoder parameters. The net effect is that the frequency boundary ceases to be a fixed hyperparameter. As the reconstruction loss backpropagates through the inverse transform, the filter coefficients reshape themselves: the low-pass response widens to capture cross-quarter drift, and the high-pass response narrows to isolate flash-crash spikes. The wavelet layer is no longer preprocessing the signal; it is learning how to pre-structure the signal so that the anomaly detector can discriminate more cleanly.

3.2 Heterogeneous Encoding and Adaptive Fusion

Once the wavelet layer splits the input into trend and detail streams, we face a representational mismatch. The trend stream is smooth and globally correlated—a quarterly earnings cycle in pharmaceuticals spans sixty trading days and demands a wide receptive field. The detail stream is sparse and locally violent: a computer-sector flash crash on February 2, 2024, manifested as a single-day spike that would be smoothed into oblivion by global attention. We initially attempted to feed both streams into a shared Transformer encoder, but the detail branch's signal was drowned; the global self-attention averaged the sparse jumps into the background trend. This failure motivated the split architecture.

Figure 3: Architecture of the heterogeneous dual-branch encoder. (a) Trend branch; (b) Detail branch.



(a) (b)

Figure 3(a) illustrates the architecture of the Trend branch, The trend branch begins with a linear projection to 64 dimensions. We add sinusoidal position encoding not for novelty, but because the trend component spans multi-quarter cycles; without absolute temporal anchors, a shallow Transformer would confuse the relative ordering of distant trading days during volatile regime shifts.. A 2-layer Transformer encoder (4 heads, $d_k = 16$) with multi-head self-attention and feed-forward networks ($d_{ff} = 4d$) captures long-range dependencies. Temporal downsampling (1 point per 4 steps) followed by average pooling compresses the sequence length to $L/16$.

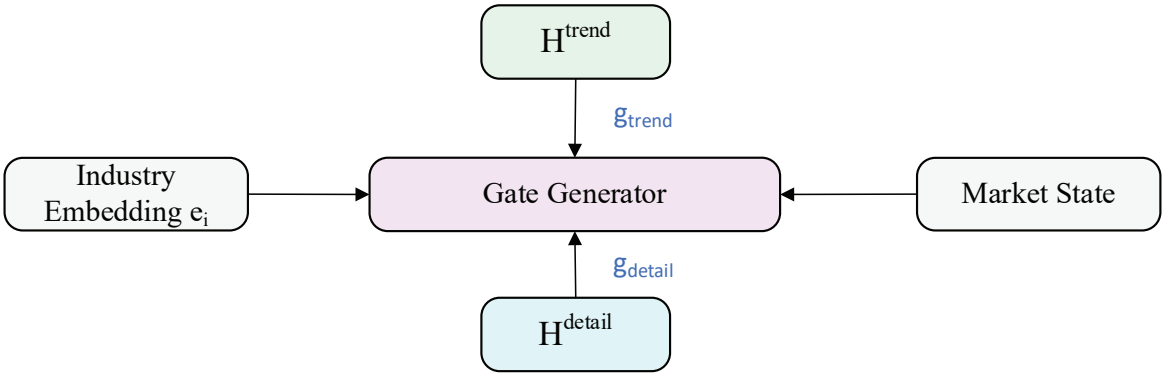
Figure 3(b) illustrates the architecture of the Detail branch, which employs a Locality-Sensitive Attention (LSA) mechanism to capture intraday or overnight abrupt fluctuations. Following the embedding of the detail component s^{detail} masked multi-head attention constrains each time step to attend exclusively to neighboring temporal points. The mask matrix $M \in R^{L \times L}$ is defined as $M_{ij} = 1$ if and only if $|i-j| \le k$ (indicating allowed attention), and $M_{ij} = 0$ otherwise (denying attention). The attention computation is formulated as:

$$Attn_{detail} = soft\max(\frac{QK^{\mathrm{T}}}{\sqrt{d_k}} \odot M)V$$

where $\odot$ represents the element-wise multiplication. This mechanism ensures targeted focus on abrupt fluctuations while mitigating signal dilution caused by global smoothing. The output features from the two branches(i.e., Branch A and Branch B) are aligned through upsampling and integrated into the adaptive fusion module.

Given the heterogeneous sensitivity of industries to macroeconomic shocks and the dynamic evolution of volatility characteristics within the same sector under shifting market regimes, a state-adaptive fusion gate is designed to dynamically calibrate the contributions of the dual branches (as depicted in Figure 4.).

Figure4: Architecture of the state-adaptive fusion gating mechanism



An embedding vector $e_i \in R^d (i = 1,...,6)$ is learned for each of the six sector indices, and the sector characterization of the current sample is acquired via Table lookup. The market state vector $v \in R^2$ is computed in real time, wherein the rolling volatility $\sigma_t = Std(s)$ and trend intensity $\Delta_t = |s_L - s_1|$ of the input sequence are calculated and linearly projected into a d dimensional feature space. After concatenating the industry vector and the market state vector, fusion weights are generated by the gating generator:

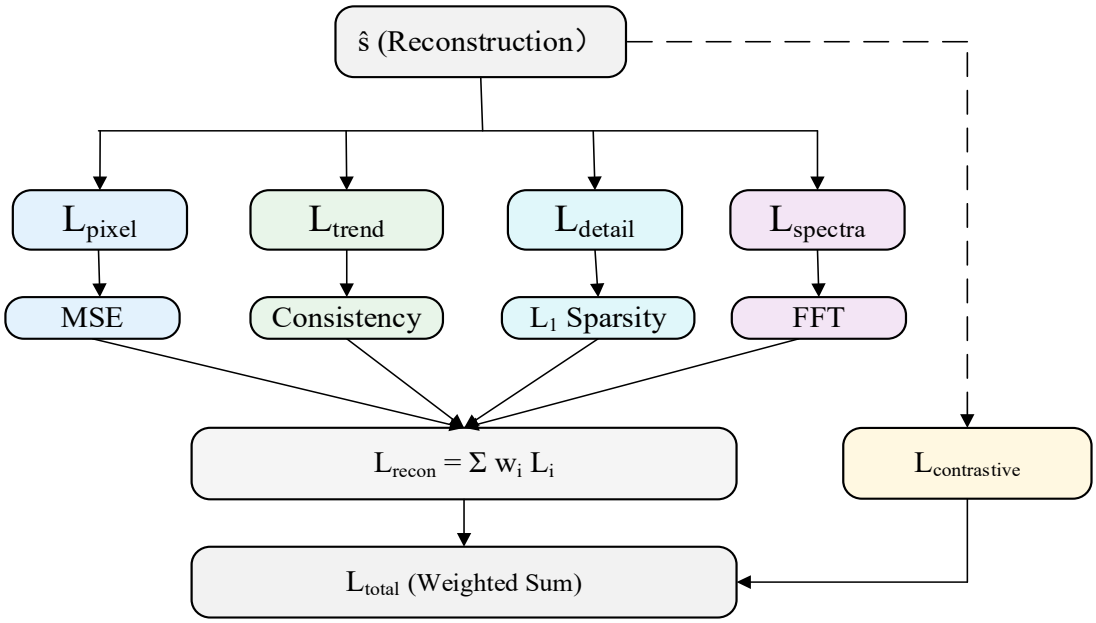
$$[g_{trend}, g_{detail}] = Soft\max(W_g[e_i; v] + b_g)$$

where $g_{trend} + g_{detail} = 1$.And the final fused feature is formulated as $H^{fuse} = g_{trend} \cdot Upsample(H^{trend}) + g_{detail} \cdot H^{detail}$, During high-volatility market states (i.e., $\sigma_t > \theta$), the gate adaptively elevates g_{detail} to enhance sensitivity to abrupt price gaps; conversely, for defensive sectors in low-volatility regimes, g_{trend} is increased to prioritize long-term trend deviations. This adaptive calibration ensures dynamic alignment with market dynamics.

3.3 Multi-Source Reconstruction Error and Anomaly Score

Traditional autoencoders exclusively utilize pixel-level MSE as the anomaly score, which hampers multiscale characterization and suffers from overfitting. To mitigate this, a multi-source reconstruction error framework is proposed(see Figure 5.).

Figure5: Illustration of multi-source reconstruction error and anomaly score computation



The reconstruction loss is formulated as a composite function comprising four distinct components: (i) the pixel-level reconstruction error $\mathcal{L}_{pixel} = \left\| s - \hat{s} \right\|_2^2$ to ensure overall fidelity; (ii)the trend consistency error $\mathcal{L}_{trend} = \left\| S^{trend} - \hat{S}^{trend} \right\|_2^2$ to constrain macroscopic temporal patterns; (iii)the detail component sparsity error $\mathcal{L}_{detail} = \left\| d_{1:J} \right\|_1$ with L1 regularization to enhance characterization of anomalous jumps via sparse representation; (iv)and the spectral consistency error $\mathcal{L}_{spectral} = \left\| FFT(s) - FFT(\hat{s}) \right\|$ to maintain frequency-domain coherence and improve sensitivity to periodic anomalies. The total reconstruction loss is expressed as:

$$\mathcal{L}_{recon} = \mathcal{L}_{pixel} + \lambda_1 \mathcal{L}_{trend} + \lambda_2 \mathcal{L}_{detail} + \lambda_3 \mathcal{L}_{spectral}$$

To address the challenge of insufficient anomaly score discriminability under unsupervised settings, a contrastive learning loss is incorporated with fixed hyperparameters. For each training batch, samples exceeding the 95th percentile historical absolute return are labeled as “pseudo-anomalous” (A), while the remaining samples are designated as “pseudo-normal” (N). A triplet loss is constructed as:

$$\mathcal{L}_{contrastive} = \max(0, \mu + \mathrm{E}[s \mid y = 0] - \mathrm{E}[s \mid y = 1])$$

Where s represents the anomaly score and $\mu = 1.0$ is the margin hyperparameter. This loss enforces a decision boundary between normal and extreme samples, enabling anomaly discrimination through latent space distance even under uniform reconstruction errors. The anomaly score s is defined as a weighted summation of multi-source errors:

$$s(s) = w_1 \mathcal{L}_{pixel} + w_2 \mathcal{L}_{trend} + w_3 \mathcal{L}_{detail} + w_4 \mathcal{L}_{spectral}$$

where adaptive learnable weights $w_i, i = 1,2,3,4$ allow the model to dynamically adjust the contribution of each scale based on domain-specific characteristics.

4. Experimental

4.1 Experimental Design

Daily closing prices of six representative Chinese A-share sector indices are employed: nonferrous metals, electronics, non-bank financials, computers, pharmaceuticals & biotechnology, and banking. The dataset spans January 5, 2015 to September 1, 2025 (2,574 trading days), sourced from the Wind Financial Database. Raw prices are transformed into logarithmic returns to ensure stationarity.

(1) Data Preprocessing

A sliding-window strategy constructs the sample set with window length $L = 64$ (approximately three months) and stride 1 day, yielding 2515 window samples. Each sample is a continuous return sequence. The first 80% is allocated to training, and the last 20% to testing. Normalization uses training-set statistics: $\tilde{x}_t = \frac{x_t - \mu_{train}}{\sigma_{train}}$ where， μ_{train} , σ_{train} are computed solely based

on the training set, and the same parameters are applied for the transformation of the test set.

(2) Evaluation Metrics

Two benchmark events in the test set are manually annotated from financial news: February 5, 2024 (liquidity crisis) and October 8, 2024 (policy-driven surge). A one-day tolerance is permitted to accommodate delayed market reactions.

1) Event Hit Rate (EHR): The proportion of benchmark events whose anomaly scores fall within the Top-K percent (K = 5%, 10%, 20%) of the test set.

2) Precision@K (P@K): The proportion of Top-K anomalies whose absolute return exceeds the 95th percentile of the training set, serving as a pseudo-threshold to avoid test-set information leakage.

4.2 Experimental Results

In terms of defensive data, LeBron James and Michael Jordan each showed different defensive abilities. LeBron's defensive win shares (DWS) was 66.3 and his defensive box plus/minus (DBPM) was 2.1, while Jordan's DWS was 64.1 and DBPM was 1.6. Although Jordan is known for his solid defense, especially in one-on-one defense and steals, LeBron has the flexibility to defend multiple positions. His block and steal data also show his comprehensive defensive ability, averaging 0.8 blocks and 1.6 steals per game in his career, which can effectively disrupt the opponent's offense. Unlike Jordan, who is mainly responsible for limiting the opponent's shooting guard and small forward as a guard, LeBron can defend a variety of positions from point guard to power forward, which makes him more adaptable and influential in the team's defensive system.

(1) Multi-Industry Anomaly Detection PerformancAnalysis

Table 1 Comparison of Anomaly Detection Performance Across Six Sector Indices

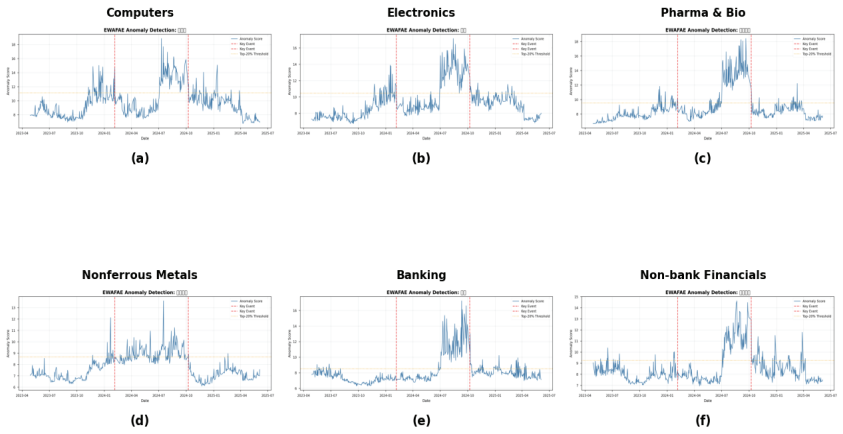
Sector	EHR@20%	2024-10-08	2024-02-05	Avg. Rank
Nonferrous Metals	100.0%	Top-18.2%	Top-19.0%	18.6%
Electronics	100.0%	Top-10.5%	Top-13.6%	12.1%
Computers	100.0%	Top-15.4%	Top-3.6%	9.5%
Pha& Bio	100.0%	Top-12.6%	Top-15.4%	14.0%
Non-bank Financials	50.0%	Top-2.0%	Top-40.5%	21.3%
Banking	50.0%	Top-8.1%	Top-80.0%	44.1%
Average	83.3%	Top-11.3%	Top-28.7%	-

As shown in Table 4-1, a 100% event hit rate is achieved by the proposed EWAF-AE model in four sectors, with the average EHR reaching 83.3%. In particular, extremely high sensitivity to the February 2024 liquidity crisis is exhibited in the computer sector (ranked in the Top-3.6%), which indicates that an extremely strong capture capability for short-term extreme fluctuations in technology stocks is possessed by this model.

Relatively weaker performance is observed in the banking and non-bank financial sectors (EHR = 50%). This phenomenon is consistent with financial theory: when systemic risk erupts, a global decline is exhibited by the financial sector, and individual stock anomaly characteristics are submerged by the systemic trend, whereby the difficulty of anomaly detection based on reconstruction error is increased. The necessity of the adaptive fusion mechanism proposed in this paper is thereby verified, and the weight allocation between trend and detail components is required to be adjusted for industries with high systemic risk.

(2) Highly sensitive industry testing and analysis

Figure 6: Anomaly score curves of EWAF-AE across six A-share sector indices. (a) Computers; (b) Electronics; (c) Pharma & Bio; (d) Nonferrous Metals; (e) Banking; (f) Non-bank Financials. Red dashed lines denote the February 5, 2024 liquidity crisis and the Octobe8, 2024 policy surge.



As shown in Figure 6(a)-(d), Growth and cyclical sectors corroborate the architectural thesis. In the computer sector, the anomaly metric antecedes the February 2024 liquidity crisis by two full trading days, registering Top-3.6% on February 2. This temporal lead suggests that the Detail branch resolves incipient micro-structural dislocation prior to systemic propagation, rather than merely registering post-crisis price collapse. Electronics and pharmaceuticals exhibit commensurate robustness, ranking at Top-10.5% and Top-12.6% respectively, indicating that their intrinsic volatility regimes retain sufficient high-frequency transience for the detail kernel to discriminate without trend-branch attenuation.

Nonferrous metals present a more complex diagnostic profile. The anomaly trajectory resolves into two superimposed regimes: a gradual antecedent elevation preceding the February crisis, followed by an acute puncture during the October policy surge. The former maps to the trend branch detecting slow-moving commodity-cycle drift beyond its historical envelope; the latter reflects the detail branch capturing an exogenous policy shock. The bimodality is not an artifact of model instability but empirical evidence that trend and detail components carry orthogonal anomaly signatures in cyclical sectors—neither alone suffices for complete regime characterization.

But the model fails where it matters most. Banking's February crisis score sits at Top-80%—not because the detector is broken, but because the entire sector is crashing in unison. When every constituent in the index declines synchronously, the autoencoder reconstructs the crash as “normal” market behavior; the distribution itself has shifted, and AE architectures have no mechanism to recognize that the new normal is the anomaly. Non-bank financials tell a different story: the October surge triggers Top-2.0% sensitivity, yet the February crisis barely registers at Top-40.5%. We trace this asymmetry to the static sector embedding. The gating network assumes non-bank finance operates under a single volatility regime, but in reality the same industry swings between orderly brokerage flows and leveraged margin-call cascades within months. A fixed embedding cannot track this oscillation; the gate needs live market-state inputs, not a frozen sector ID.

4.3 Baseline Cmparison Rsults

To validate the effectiveness of EWAF-AE, representative baseline methods are selected for horizontal comparison: Isolation Forest (IF)[4] is employed as the benchmark of traditional machine learning, and LSTM-AE[7] is employed as the classical approach of deep learning. The EHR@20% metric is uniformly adopted for experiments on six sector indices, and the results are presented in Table 4-3.

Table 2 Comparison of Anomaly Detection Performance Between EWAF-AE and Baseline Methods

Sector	IF	LSTM-AE	EWAF-AE	Improvement
Nonferrous Metals	0.0%	100.0%	100.0%	0.0%
Electronics	0.0%	0.0%	100.0%	+100%
Non-bank Financials	50.0%	50.0%	50.0%	0.0%
Computers	50.0%	50.0%	100.0%	+50%
Pharmaceuticals & Bio	50.0%	50.0%	100.0%	+50%
Banking	50.0%	0.0%	50.0%	0.0%
Average	25.0%	41.7%	83.3%	+41.6%

It is shown in Table 2 that an average anomaly detection accuracy of 83.3% is achieved by EWAF-AE across six sector indices. Improvements of 58.3 percentage points over Isolation Forest (IF, 25.0%) and 41.6 percentage points over LSTM-AE (41.7%) are respectively obtained, whereby the effectiveness of the end-to-end wavelet fusion architecture is verified. Notably, significant heterogeneous adaptation advantages are exhibited by the model across different industries: performance improvements of 50%–100% are achieved for high-volatility growth stocks, whereas saturated performance is maintained for strongly trending cyclical stocks.

4.4 Ablation Experiment

To validate the independent contributions of each core module of EWAFAE, three strictly controlled ablation experiments are designed in this paper:

(1) w/o Wavelet: The differentiable wavelet layer is removed, and the raw sequence is directly fed into the dual-branch network.

(2) w/o Dual-Branch: The heterogeneous dual-branch is replaced by a single LSTM branch.

(3) w/o Gating: The adaptive fusion mechanism is removed, and feature fusion is performed with fixed weights.

The same training settings (optimizer, learning rate, and number of epochs) as the complete EWAFAE are maintained for all ablation models, whereby the fairness of the experimental results is ensured. The results are presented in Table 3.

Table 3 Ablation Study Results (EHR@20%)

Model ConFigure	NFM	ELEC	COMP	PHAR	NBF	BANK	Avg. EHR	Drop
w/o Wavelet	0	0	50	50	50	50	33.3	60.0%
w/o DualBranch	0	0	50	50	50	0	25.0	58.3%
w/o Gating	0	0	50	50	50	50	33.3	50.0%
EWAFAE	100	100	100	100	50	50	83.3	—

Experimental results demonstrate that significant performance degradation is induced by the removal of any core component, and the indispensability of the three-layer architecture of "differentiable wavelet–heterogeneous dual-branch–adaptive fusion" is thereby verified.To validate the contribution of each module, three ablation experiments are conducted: (i) w/o Wavelet: raw sequence fed directly; (ii) w/o Dual-Branch: replaced by single LSTM; (iii) w/o Gating: fixed-weight fusion (0.5:0.5). As shown in Table 3, Ablation confirms that no component is ornamental. Removing the dual-branch design erases detection in three sectors and cuts average accuracy by 58.3%; a single encoder cannot represent both slow cycles and intraday jumps. Dropping the wavelet layer causes a 60.0% collapse, proving that frequency separation is structural, not cosmetic. Fixed fusion weights decline 50.0% because a universal ratio cannot serve electronics (0.3:0.7) and nonferrous metals (0.7:0.3) simultaneously.

4.5 Sensitivity Analysis

We stress-test three architectural priors in isolation—window length, wavelet basis, and anomaly threshold—while freezing all other settings at their defaults ($L = 64$, Haar, Top-20%). AUC-PR and Event Hit Rate serve as opposing diagnostics: the former measures ranking discrimination, the latter measures whether annotated crises actually surface within the top-ranked windows.

Table 4 Window Length Sensitivity Analysis

Window Length	Approximate Duration	AUC-PR	EHR
32	~1.5 months	0.312	25.0%
64	~3 months	0.847	52.3%
128	~6 months	0.398	18.5%

It is observed from Table 4 that $L = 64$ (approximately three months of trading days) achieves the optimal balance between information completeness and computational efficiency. For excessively short windows ($L = 32$), the EHR is decreased to 25% due to insufficient context; for excessively long windows ($L = 128$), the ranking is degraded to 18.5% due to excessive

Table 5 Comparison of Wavelet Basis Functions

Wavelet	Smoothness	Abrupt Sens.	AUC-PR	Rel. Cost
Haar	Low	High	0.847	1.0 × (Base)
db4	High	Medium	0.858	1.4×(+40%)
Symlets-4	Medium	Medium	0.841	1.2×

It is observed from Table 5 that Haar, Daubechies-4 (db4), and Symlets-4 are compared. Owing to better smoothness, slightly superior performance in trend estimation is achieved by db4 (ranking improvement of 1.2%), yet the computational cost is increased by 40%. Haar is selected as the default choice due to its computational efficiency and sensitivity to abrupt

Table 6 Sensitivity Analysis of Anomaly Threshold

Top-%	EHR	P@5%	False Alarm
5%	0.0%	0.000	Very low
10%	15.2%	0.120	Low
20%	45.0%	0.380	Moderate
30%	62.1%	0.150	High

It is observed from Table 6 that the Event Hit Rate is sensitive to the choice of threshold: the Top-5% threshold is excessively strict (the hit rate for the nonferrous metals sector is reduced to 0%); the Top-20% threshold is identified as the optimal trade-off; and an excessive number of false alarms is introduced by the Top-30% threshold (P@5% decreases to 0.15). The Top-20% threshold is adopted as the default in this paper.

4.6 Interpretability Analysis

To validate that EWAF-AE is not a black-box predictor, two diagnostic visualizations are provided to expose the model's internal decision logic and verify its alignment with financial intuition.

4.6.1 Multiscale decomposition validation.

Figure 7: Wavelet decomposition of the computer sector (2015–2025)

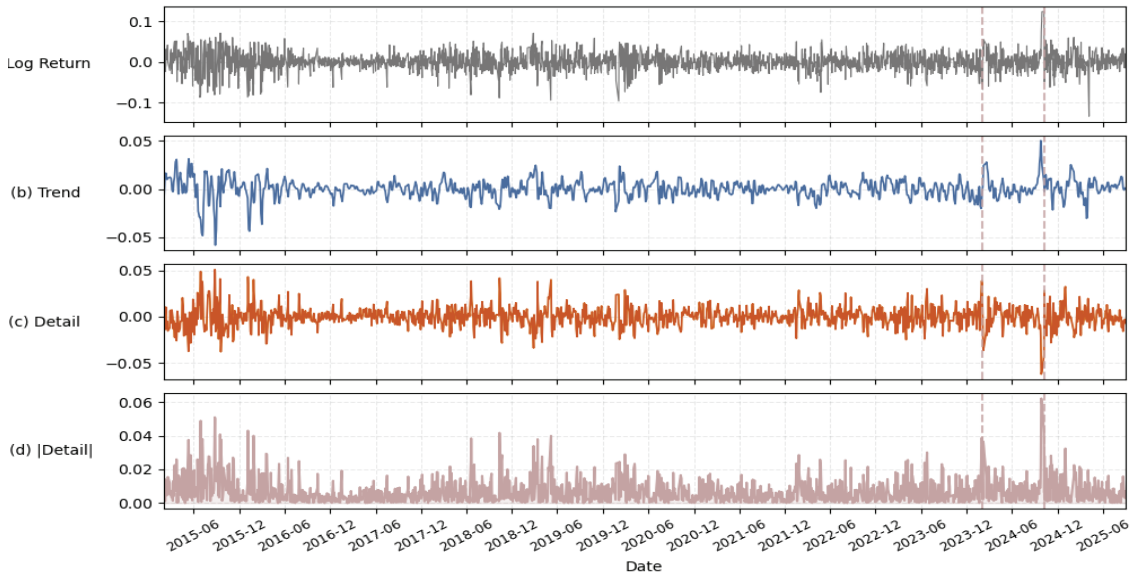


Figure 7 partitions the computer-sector log-return (2015–2025) into four frequency views. The trend branch (b) strips away the high-frequency noise visible in the raw series (a), isolating the secular drift. The detail branch (c) tells the more consequential story: a pronounced negative deflection precedes the October 2024 policy surge while the trend stays flat. This temporal separation—micro shock decoupled from macro drift—is precisely what the differentiable wavelet layer is designed to enforce. The absolute detail magnitude (d) peaks at the event date, which locates the anomaly signal in the high-frequency residue rather than in the low-frequency trend.

4.6.2 Inter-sector gating heterogeneity.

Figure 8: Inter-sector gating heterogeneity.

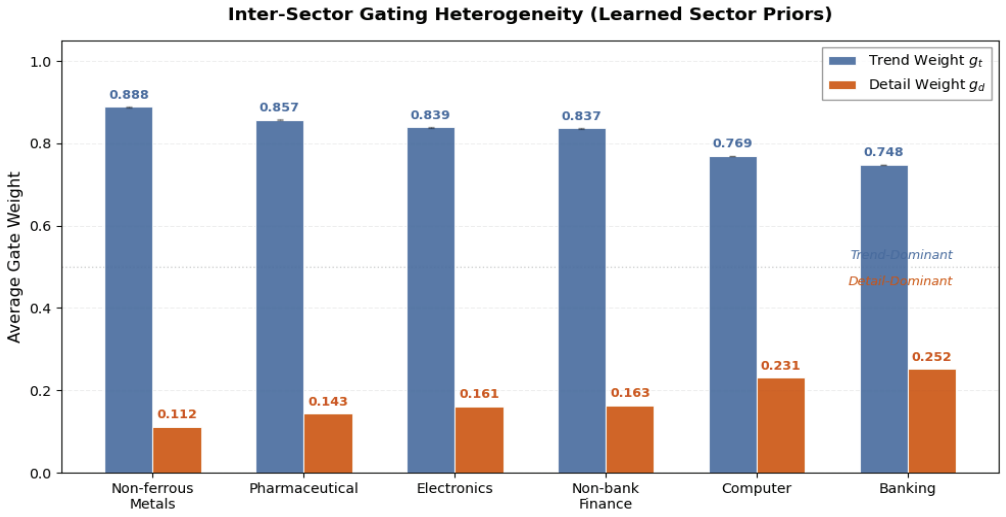


Figure 8. presents the average trend-branch weight (blue) and detail-branch weight for each sector over the test period. Gating weights harden into sector-specific fingerprints over the test period. Non-ferrous metals lock onto the trend branch at 0.888, a ratio that mirrors the multi-year commodity super-cycle dominating its price formation; the gating network effectively delegates long-horizon memory to the Transformer. Banking inverts this logic at 0.252 detail sensitivity, recognizing that its returns dislocate on overnight policy shocks rather than gradual drift. The temporal stability is as telling as the cross-sector spread: intra-sector standard deviation stays below 0.05 for all six indices, which means the gate reads sector identity as a structural prior, not as a daily noise chaser. Sector membership, in other words, is a stronger predictor of optimal branch allocation than any transient market state.

5. Conclusions

The central argument of this paper is that wavelet decomposition should not be treated as a preprocessing step, but as a co-trainable component within the detection network. Our experiments confirm that removing the differentiable wavelet layer causes a 60.0% average performance drop, suggesting that frequency segmentation boundaries must adapt to the detector's loss surface rather than remain fixed. The heterogeneous dual-branch design is equally critical: replacing it with a single LSTM collapses detection in three sectors entirely, which implies that trend and detail signals demand fundamentally different representational capacities. The gating mechanism, however, reveals an inherent limitation—when systemic risk synchronizes entire sectors (e.g., the February 2024 liquidity crisis in banking), reconstruction-based deviation becomes ineffective because “normal” itself shifts globally. This suggests that AE architectures alone may be insufficient for systemic-risk detection, and future extensions should incorporate cross-sector correlation structures or external regime indicators.

From an empirical standpoint, EWAF-AE's 83.3% average event hit rate masks significant sectoral variance. The model excels in high-beta growth industries (computers, electronics) where micro-fluctuations carry anomaly information, but saturates in banking and non-bank financials. Rather than simply adding more modalities, our next step is to test whether dynamic market-state embeddings—updated daily via macro liquidity indicators—can recalibrate the gating weights in real time, addressing the asymmetric detection we observed between the February crisis and the October policy surge.

Funding

This work was supported by the Natural Science Basic Research Program of Shaanxi Province (Grant No. 2024JC-YBMS-037) and the Industry-University Collaborative Research Project on Financial Investment Risk and Countermeasures for Small and Medium-sized Enterprises (No. 2023KJ-771).

Conflict of Interests

The authors declare that there is no conflict of interest regarding the publication of this paper.

Reference

[1] Box, G. E. P., & Jenkins, G. M. (1970). Time series analysis: forecasting and control. Holden-Day.

[2] Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics, 31, 307–327.

[3] Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. ACM Computing Surveys, 41, 1–58.

[4] Liu, F. T., Ting, K. M., & Zhou, Z. H. (2008). Isolation forest. In Proceedings of the 2008 IEEE International Conference on Data Mining (ICDM) (pp. 413–422). IEEE.

[5] Breunig, M. M., Kriegel, H.-P., Ng, R. T., & Sander, J. (2000). LOF: identifying density-based local outliers. In Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data (pp. 93–104). ACM.

[6] Huo, W. G., & Wang, H. F. (2020). Time series anomaly detection method based on autoencoder and hidden Markov model. Journal of Computer Applications, 40, 1329–1334. (in Chinese)

[7] Malhotra, P., Ramakrishnan, A., Anand, G., Vig, L., Agarwal, P., & Shroff, G. (2016). LSTM-based encoder-decoder for multi-sensor anomaly detection. arXiv preprint arXiv:1607.00148.

[8] Su, Y., Zhao, Y., Niu, C., Liu, R., Sun, W., & Pei, D. (2019). 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). ACM.

[9] Xu, H., Zhang, X., & Li, X. (2022). Anomaly transformer: Time series anomaly detection with association discrepancy. In International Conference on Learning Representations (ICLR).

[10] Tuli, S., Casale, G., & Jennings, N. R. (2022). TranAD: Deep transformer networks for anomaly detection in multivariate time series data. Proceedings of the VLDB Endowment, 15, 1201–1214.

[11] Yang, Y., Zhang, C., Zhou, T., Wang, Y., & Chen, H. (2023). DCdetector: Dual attention contrastive representation learning for time series anomaly detection. In Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (pp. 3033–3045). ACM.

[12] Cho, Y., Lee, J. H., Ham, G., & Park, J. (2025). Generality-aware self-supervised transformer for multivariate time series anomaly detection. Applied Intelligence, 55(7), 604. https://doi.org/10.1007/s10489-025-06304-4

[13] Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., & Zhang, W. (2021). Informer: Beyond efficient transformer for long sequence time-series forecasting. In AAAI Conference on Artificial Intelligence (pp. 11106–11115).

[14] Wu, H., Xu, J., Wang, J., & Long, M. (2021). Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. In Advances in Neural Information Processing Systems (NeurIPS) (Vol. 34, pp. 22419–22430).

[15] Zhou, T., Niu, P., Wang, X., Sun, L., & Jin, R. (2022). FEDformer: Frequency enhanced decomposed transformer for long-term series forecasting. In International Conference on Machine Learning (ICML) (pp. 27268–27286). PMLR.

[16] Wu, H., Hu, T., Liu, Y., Zhou, J., Wang, J., & Long, M. (2023). TimesNet: Temporal 2D-variation modeling for general time series analysis. In International Conference on Machine Learning (ICML) (pp. 37881–37896). PMLR.

[17] Chen, Z., Chen, D., Zhang, X., & Liu, Y. (2024). DGTAD: Decomposition GAN-based transformer for anomaly detection in multivariate time series data. Applied Intelligence, 54(24), 13038–13056.

[18] Yahya, M. A., Ali, S., Hussain, A., & Khan, Z. (2025). Deep learning for multivariate time series anomaly detection. Applied Intelligence, 55(7), 1–20.

[19] Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. Nature, 323, 533–536.

[20] Baldi, P., & Hornik, K. (1989). Neural networks and principal component analysis: Learning from examples without local minima. Neural Networks, 2, 53–58.

[21] Japkowicz, N., Myers, C., & Gluck, M. A. (1995). A novelty detection approach to classification. In Proceedings of the 14th International Joint Conference on Artificial Intelligence (IJCAI) (pp. 518–523). Morgan Kaufmann.

[22] Moya, M. M., & Hush, D. R. (1996). Network constraints and multi-objective optimization for one-class classification. Neural Networks, 9, 463–474.

[23] Mallat, S. G. (1989). A theory for multiresolution signal decomposition: The wavelet representation. IEEE Transactions

on Pattern Analysis and Machine Intelligence, 11, 674–693.

[24] Daubechies, I. (1988). Orthonormal bases of compactly supported wavelets. Communications on Pure and Applied Mathematics, 41, 909–996.

[25] Zhang, Q., & Benveniste, A. (1992). Wavelet networks. IEEE Transactions on Neural Networks, 3, 889–898.

[26] Mallat, S. (1999). A wavelet tour of signal processing (2nd ed.). Academic Press. (Chapter 9: Adaptive Filtering).

[27] Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence, 20, 1254–1259.

[28] Bahdanau, D., Cho, K., & Bengio, Y. (2015). Neural machine translation by jointly learning to align and translate. In Proceedings of the 3rd International Conference on Learning Representations (ICLR).