http://dx.doi.org/10.62762/TETAI.2025.191759

RESEARCH ARTICLE

Anomaly Detection and Risk Early Warning Systemfor Financial Time Series Based on the WaveLST-Trans Model

Tian Su[1,†,*], **Runlong Li**[2,†], **Bo Liu**[3], **Xiaoxiang Liang**[4], **Xinhao Yang**[5] and **Yan Zhou**[6]

[1] Meta Platforms Inc., Seattle, WA 98109, United States
[2] Department of Electrical Engineering and Computer Science, University of California, Irvine, CA 92555, United States
[3] College of Computer Sciences, Northeastern University, Cupertino, CA 95014, United States
[4] Business Analytics, Washington University in St. Louis, St. Louis, MO 63130, United States
[5] Department of Computer Science, University of Southern California, Los Angeles, CA 90007, United States
[6] Department of Mathematics, Northeastern University, San Jose, CA 95131, United States

Abstract

Abnormal fluctuations in financial markets may signal significant risks or market manipulation, so efficient time series anomaly detection methods are crucial for risk management. However, traditional statistical methods (e.g., ARIMA, GARCH) are difficult to adapt to the nonlinear and multi-scale characteristics of financial data, while single deep learning models (e.g., LSTM, Transformer) have limitations in capturing long-term trends and short-term fluctuations. In this paper, we propose WaveLST-Trans, a financial time series anomaly detection model based on the combination of wavelet transform (WT), LSTM and Transformer. The model first uses wavelet transform to perform multi-scale decomposition, extracts low-frequency trend and high-frequency fluctuation features, and feeds them into LSTM (to learn the long-term dependence) and Transformer (to capture local mutations) respectively, and finally integrates the multi-scale information through the feature fusion layer, which improves the detection accuracy and robustness. The experiments are conducted on Binance (cryptocurrency market) and S&P 500 (stock market) datasets, and the results show that WaveLST-Trans mostly outperforms the mainstream models in terms of F1-score, recall, and precision, and improves the detection performance by 3% and 10% in high-frequency market and long-term trend market, respectively. This study provides a more accurate and stable anomaly detection method for financial market risk management, which can be widely used in market regulation, quantitative trading and financial risk control, helping to improve the security and stability of the financial system.

Keywords: financial anomaly, anomaly detection, LSTM, transformer, wavelet transform.

Submitted: 08 April 2025
Accepted: 15 May 2025
Published: 21 May 2025

Vol. 2, **No.** 2, 2025.
10.62762/TETAI.2025.191759

***Corresponding author:**
✉ Tian Su
tiansu1997@gmail.com

† Tian Su and Runlong Li contributed equally to this work

Citation

Su, T., Li, R., Liu, B., Liang, X., Yang, X., & Zhou, Y. (2025). Anomaly Detection and Risk Early Warning System for Financial Time Series Based on the WaveLST-Trans Model. *ICCK Transactions on Emerging Topics in Artificial Intelligence*, 2(2), 68–80.

© 2025 by the Authors. Published by Institute of Central Computation and Knowledge. This is an open access article under the CC BY license (https://creativecommons.org/licenses/by/4.0/).

1 Introduction

The abnormal volatility in financial markets can have a profound impact on the global economy and financial systems, such as market crashes, irregular trading, and financial fraud [1]. These unexpected events are often accompanied by drastic price fluctuations and changes in trading volume, which pose significant risks to investors, regulatory bodies, and financial institutions [2, 3]. Deep learning models such as LSTM have demonstrated strong capability in forecasting and characterizing such market dynamics [3]. Therefore, accurately identifying abnormal behaviors in financial time series is of paramount importance for risk management and market stability. However, financial market data exhibit high non-linearity, non-stationarity, and multi-scale characteristics, and abnormal events are often hidden within complex time series patterns, making traditional anomaly detection methods insufficient for practical needs [4]. The construction of efficient anomaly detection systems that can identify abnormalities in real time within the dynamic environment of financial markets remains a critical issue in the field of financial technology [1, 5].

In recent years, deep learning has made significant advancements in time series analysis, offering new approaches for financial anomaly detection [5–7]. Long Short-Term Memory (LSTM) networks have been widely adopted for financial market trend modeling due to their superior performance in handling temporal dependencies [8]. However, LSTM models still face limitations in capturing long-term dependencies and detecting sudden anomalies. In contrast, the self-attention mechanism in Transformer models enables the modeling of global information, making them particularly effective in capturing short-term fluctuations and local anomaly patterns [9]. Therefore, Transformer-based models have demonstrated strong capability in long-sequence time series modeling [10]. Combining LSTM and Transformer can create a complementary framework that addresses both long-term dependencies and short-term fluctuations simultaneously [11]. Additionally, wavelet transform (WT), as an effective signal processing tool, enables multi-scale decomposition of time series, facilitating the extraction of features across different frequencies and enhancing the model's capability in identifying both long-term market trends and short-term anomalies [12].

Despite significant progress in financial anomaly detection, several challenges remain [1]. Many traditional statistical models, such as ARIMA and GARCH, struggle to capture the nonlinear and multi-scale characteristics of financial markets, limiting their effectiveness in anomaly detection [13]. Wavelet-based methods, by contrast, offer effective multi-resolution decomposition suited to such non-stationary signals [14]. Single deep learning models often fail to simultaneously account for both long-term dependencies and short-term fluctuations in time series, making it difficult to accurately characterize the dynamic nature of financial markets [15]. Moreover, financial anomalies are often abrupt and transient, posing challenges in effectively extracting key anomaly features from time series while minimizing false positives and false negatives. To address these issues, this study introduces a model that integrates wavelet transform, LSTM, and Transformer, aiming to enhance the comprehensiveness and accuracy of financial time series anomaly detection.The main contributions of this paper are as follows:

- Proposing the WaveLST-Trans model, which integrates Wavelet Transform, LSTM, and Transformer to capture multi-scale features, enabling the separate processing of long-term trends and short-term fluctuations, thereby improving anomaly detection accuracy.
- Designing a feature fusion strategy, where Wavelet Transform decomposes the time series, feeding low-frequency components into LSTM to extract long-term dependencies, and high-frequency components into Transformer to identify localized anomalies, achieving efficient multi-scale feature integration.
- Conducting experiments on multiple financial market datasets to validate the superiority of WaveLST-Trans over traditional statistical, machine learning, and deep learning models, and performing ablation studies to analyze module contributions, ensuring model effectiveness and stability.

The structure of this paper is organized as follows: Section 2 reviews related research on financial time series anomaly detection, covering traditional statistical methods, the application of deep learning in finance, and the development of hybrid models. Section 3 provides a detailed description of the WaveLST-Trans model, including data preprocessing, wavelet transform, and the LSTM-Transformer module. Section 4 presents the experimental design, covering datasets, experimental setup, and validating the

model's performance through comparative and ablation experiments. Section 5 concludes the study and discusses future research directions.

2 Related Work

2.1 Traditional Time Series Anomaly Detection Methods

In the field of financial time series anomaly detection, traditional methods mainly rely on statistical models and machine learning algorithms. These approaches have achieved certain success in early research, but often exhibit limitations when faced with the complex dynamics of financial markets [16, 17]. Among them, the Autoregressive Integrated Moving Average (ARIMA) model is a widely used statistical method for time series forecasting. It detects anomalies by modeling the linear dependencies of data, but its adaptability to nonlinear and abrupt changes is relatively poor. The Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model and its variants (such as EGARCH and TGARCH) focus on volatility modeling in financial markets. These models effectively describe the heteroscedasticity of financial time series, but still suffer from a delay in identifying sudden anomalies.In addition to statistical models, Isolation Forest, an unsupervised anomaly detection method, identifies outliers by random partitioning of data points. Deep variants of Isolation Forest further extend this framework using neural representations [18]. The Local Outlier Factor (LOF) calculates local abnormality based on data density distributions. A comprehensive review of such density-based approaches can be found in [19]. Support Vector Machines (SVM) and its extended versions, such as One-Class SVM, distinguish between normal and anomalous data by constructing hyperplanes in high-dimensional space. While these methods are suitable for static data, they perform poorly when handling dynamic time series anomalies [20].

These limitations motivate the development of hybrid approaches that can overcome the constraints of individual statistical or machine learning methods when applied to complex financial time series.

2.2 Deep Learning in Financial Risk Management

In recent years, deep learning has been widely applied in financial time series anomaly detection and risk management, demonstrating superior performance compared to traditional methods. Long Short-Term Memory (LSTM) networks, which can model long-term dependencies, have shown excellent results in financial market trend prediction and anomaly detection [21]. However, LSTM is susceptible to the vanishing gradient problem when processing long time series and has limited ability to capture short-term abrupt anomalies. Gated Recurrent Units (GRU), as a simplified version of LSTM, reduce computational complexity and provide comparable performance in some financial applications [8], but they still struggle to effectively model local mutations in time series. Convolutional Neural Networks (CNNs) have also been applied to financial time series feature extraction, such as learning market patterns through one-dimensional convolutional layers [22]. However, CNNs rely on fixed-size receptive fields, making it difficult to capture long-range dependencies, and they perform poorly with non-stationary time series data. Recently, Transformer models have gained attention due to their self-attention mechanism, which excels at globally modeling the relationships between data points and is particularly adept at capturing local anomaly patterns. However, due to the lack of implicit modeling of temporal information, Transformer models may have limitations in modeling long-term trends [23]. Moreover, some hybrid models have emerged that combine LSTM and CNN in parallel or sequential architectures for time series feature extraction and temporal dependency modeling [24], or employ variant Transformers such as Informer [10] to optimize long-sequence time series modeling efficiency. Graph-based deep learning methods have also shown promise in anomaly detection by modeling relational dependencies among entities [25]. Nonetheless, these models still face issues such as high computational complexity and limited generalization capabilities, motivating the development of more comprehensive multi-scale hybrid approaches.

3 Methodology

3.1 Overall Model Architecture

The WaveLST-Trans model consists of a wavelet transform preprocessing module, an LSTM long-term dependency modeling module, a Transformer short-term fluctuation capturing module, and a feature fusion layer. Financial time series typically exhibit both long-term trends and short-term fluctuations, making it challenging for a single model to capture these features simultaneously. This limitation affects the accuracy of anomaly detection. To address this issue, WaveLST-Trans first applies wavelet transform for multi-scale decomposition of

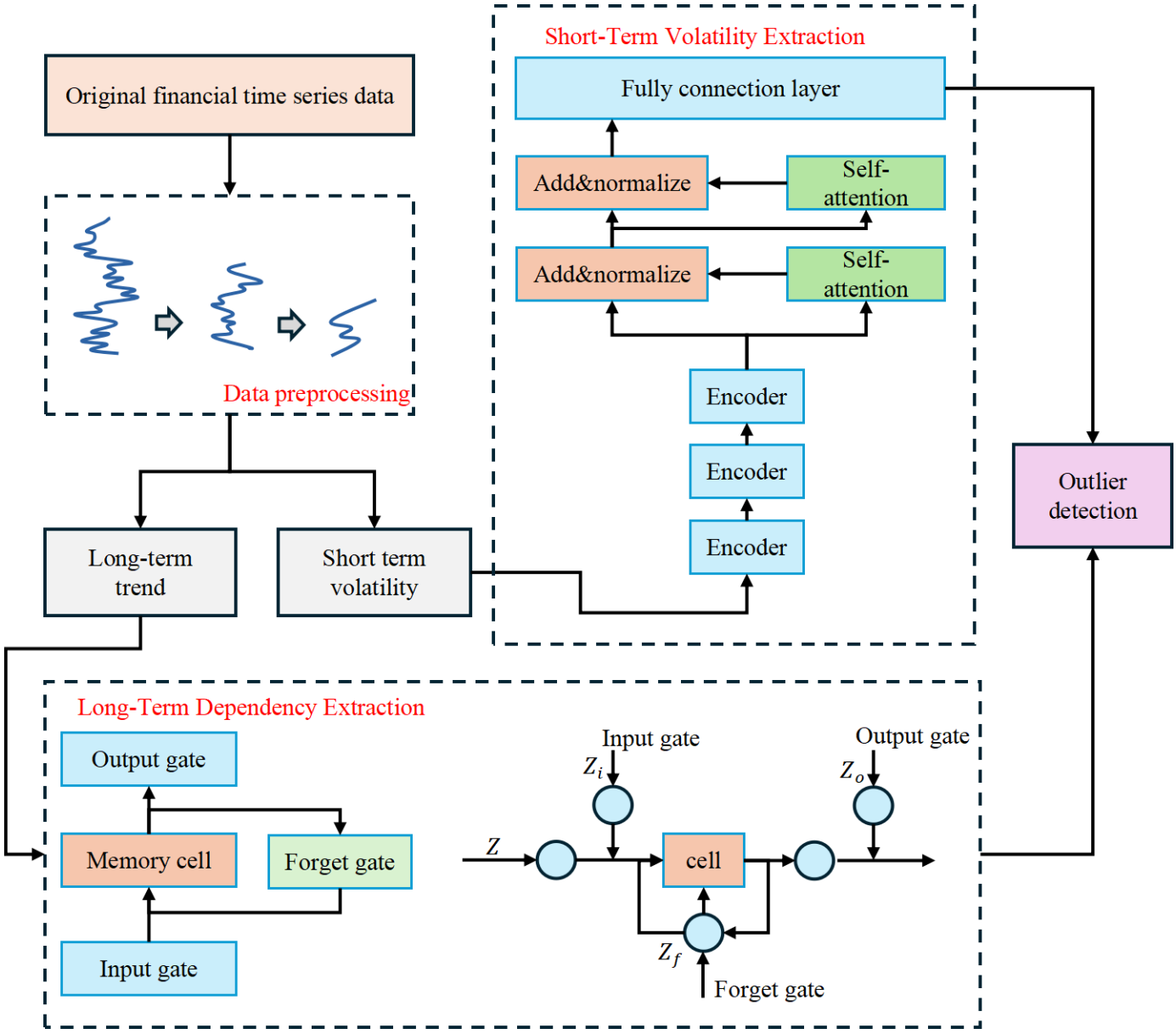

Figure 1. WaveLST Trans model architecture.

time series data, breaking down the raw data into low-frequency components (long-term trends) and high-frequency components (short-term fluctuations). The low-frequency components are fed into the LSTM network to capture long-term dependencies, while the high-frequency components are input into the Transformer module, where the self-attention mechanism extracts short-term fluctuation features. The outputs of LSTM and Transformer are integrated in the feature fusion layer, ensuring that the model effectively learns anomaly patterns across different time scales. After passing through a fully connected layer, the anomaly detection module computes anomaly scores, ultimately identifying anomalous time steps. The overall architecture of the model is illustrated in Figure 1.

WaveLST-Trans integrates LSTM and Transformer to achieve precise multi-scale modeling, enhancing the accuracy and robustness of financial time series anomaly detection. LSTM is responsible for capturing long-term trends, leveraging its gating mechanism to model temporal dependencies, making it suitable for detecting structural market changes such as sustained price increases or declines. In contrast, Transformer utilizes self-attention to efficiently extract short-term fluctuations, identifying localized sudden anomalies such as market shocks or short-term trading irregularities. The feature fusion layer combines the outputs of LSTM and Transformer, enabling the model to learn both long-term dependency patterns and short-term mutation features. Machine learning methods for time series have increasingly adopted

multi-model fusion strategies to improve robustness and generalization [11]. Compared to traditional methods, WaveLST-Trans offers stronger multi-scale feature representation, overcoming the limitations of ARIMA and GARCH on non-stationary data and surpassing Isolation Forest and Local Outlier Factor (LOF) in modeling time dependencies. Additionally, compared to standalone LSTM or Transformer models, WaveLST-Trans combines long-term memory capabilities with short-term anomaly detection, making it suitable for various market environments.

In the computational process, the wavelet transform module first decomposes the raw time series into different frequency components, where low-frequency signals capture long-term market trends, and high-frequency signals reflect short-term fluctuations and local anomalies [26]. The low-frequency signals are fed into LSTM to learn long-term dependencies, while the high-frequency signals are input into Transformer to extract short-term fluctuation features. In the feature fusion layer, the feature vectors from LSTM and Transformer are concatenated along the feature dimension and subsequently passed through a fully connected layer to produce a unified representation, ensuring the model effectively leverages multi-scale information to enhance anomaly detection accuracy. A predefined threshold is then applied to the resulting anomaly scores to identify anomalous time steps, enabling precise financial market anomaly detection.

The design of WaveLST-Trans not only enhances financial time series anomaly detection but also strengthens its applicability in risk warning systems. In market regulation, quantitative trading, and credit evaluation, this model can promptly identify potential market risks, providing more precise decision support for financial institutions [11]. Anomaly detection in time series has been systematically studied [16], and the financial domain presents unique challenges due to data non-stationarity and multi-scale characteristics. Compared with traditional methods, WaveLST-Trans offers greater adaptability to the complexities of financial markets, serving as both a theoretical foundation and practical reference for future advancements in financial risk management tools.

3.2 Data Preprocessing

WaveLST-Trans employs Wavelet Transform for data preprocessing to enhance the multi-scale feature representation of time series data [12, 26]. Multi-scale modeling with adaptive denoising has demonstrated effectiveness in extracting stable features from noisy financial signals [27]. Financial time series typically exhibit both long-term trends and short-term fluctuations, and directly modeling the raw data may make it difficult for the model to distinguish between these components, thereby affecting the accuracy of anomaly detection. Wavelet transform has been widely studied as a multi-scale signal analysis tool with strong adaptability to non-stationary financial data [12, 14]. The Wavelet Transform is particularly suitable for this task due to its ability to decompose a time series into multiple frequency bands, allowing for the extraction of both low-frequency components (representing long-term trends) and high-frequency components (representing short-term fluctuations).However, there are some trade-offs associated with this approach. While the Wavelet Transform effectively decomposes the time series into meaningful components, there is a computational cost involved in performing the transform and the subsequent model training. Additionally, the choice of wavelet function and decomposition level can impact the model's performance, and selecting inappropriate parameters may lead to overfitting or underfitting. Despite these challenges, this preprocessing approach ensures that the model effectively learns abnormal patterns across different time scales, while simultaneously reducing noise interference and improving the stability and accuracy of anomaly detection. Figure 2 illustrates the process of the data preprocessing module.

In data preprocessing, the core idea of wavelet transform is to perform time-frequency decomposition on the time series, transforming it into multi-scale feature representations. Given a discrete time series $X = \{x_1, x_2, \ldots, x_T\}$, this study adopts the Discrete Wavelet Transform (DWT) to improve computational efficiency for high-dimensional, non-stationary financial data. DWT applies a scaling factor s and time translation factor τ derived from a mother wavelet ψ at dyadic scales, decomposing the time series into low-frequency components X_L (long-term trends) and high-frequency components X_H (short-term fluctuations) as follows:

$$W_X(s,\tau) = \int X(t)\,\psi^*\left(\frac{t-\tau}{s}\right) dt \qquad (1)$$

Due to the high dimensionality and non-stationarity of financial time series, Discrete Wavelet Transform (DWT) is adopted in this paper to improve

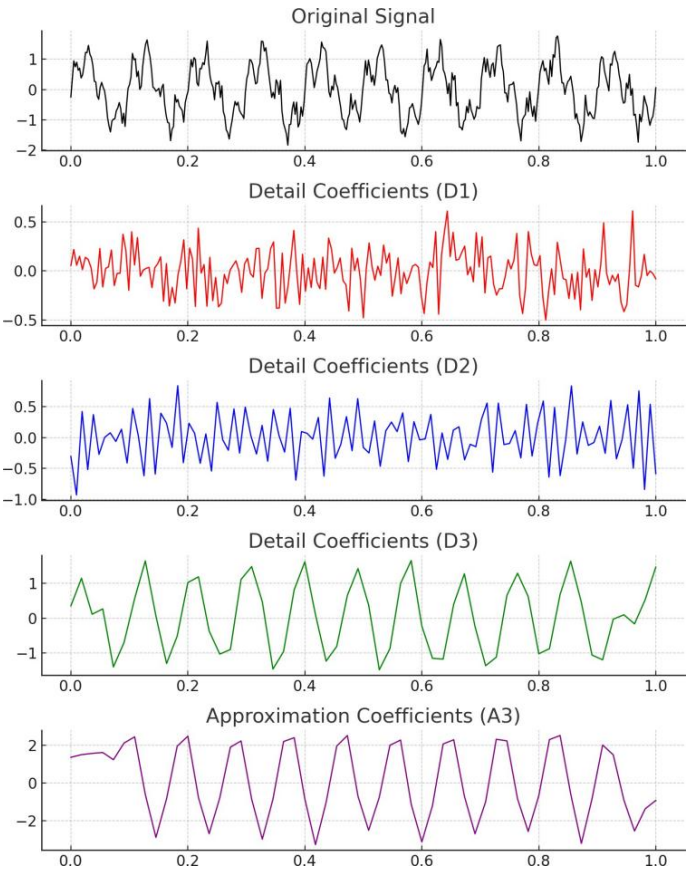

Figure 2. Schematic diagram of Wavelet Transform and data preprocessing module.

computational efficiency. The time series is decomposed into low-frequency components X_L and high-frequency components X_H as follows:

$$X_L[k] = \sum_n x[n]\, g[n-2k], \quad X_H[k] = \sum_n x[n]\, h[n-2k] \tag{2}$$

In this process, $g(n)$ and $h(n)$ represent the low-pass and high-pass filters, respectively, which are used to decompose the time series into different scale-dependent features. The low-frequency component X_L primarily captures long-term trend information, whereas the high-frequency component X_H reflects short-term fluctuations and local anomalies. After the initial decomposition, the low-frequency component can undergo recursive decomposition to extract smoother long-term trends, where J represents the number of decomposition levels. In this paper, we select Daubechies (db4) wavelet as the mother wavelet, as experimental results demonstrate its ability to effectively adapt to both the smoothness and fluctuation characteristics of financial time series data. Finally, after wavelet transform processing, the low-frequency component X_L is fed into the LSTM module to model long-term dependencies, while the high-frequency component X_H is input into the Transformer module to learn short-term fluctuation features:

$$X_L^{(j)} = \sum_n X_L^{(j-1)} g(n-2k), \quad j = 1, 2, \ldots, J \tag{3}$$

Prior to sliding window segmentation, this study applies z-score normalization to the entire training set to ensure consistent scale across time steps, thereby improving model convergence speed and stability. The mean μ and standard deviation σ are computed exclusively from the training set and applied to the validation and test sets to prevent data leakage. The normalization is performed as follows:

$$\tilde{x}_t = \frac{x_t - \mu}{\sigma} \tag{4}$$

Considering the non-stationarity of financial market data, this study first employs a Sliding Window mechanism to segment the raw time series into fixed-size, overlapping subsequences. Given a window size W, the model input at time step t is defined as $X^{(t)} = \{x_{t-W+1}, x_{t-W+2}, \ldots, x_t\}$. Each windowed subsequence is then independently subjected to DWT decomposition, yielding low-frequency component $X_L^{(t)}$ and high-frequency component $X_H^{(t)}$. This order ensures that the model processes locally stationary segments, reducing boundary effects and preventing information leakage between non-overlapping time intervals. The low-frequency and high-frequency components are subsequently fed into the LSTM and Transformer modules respectively for feature extraction and anomaly detection.

Mathematically, the Wavelet Transform provides a multi-resolution analysis of the data, where each resolution corresponds to a different scale of the time series. This is achieved by convolving the signal with a wavelet function, which is localized in both time and frequency, making it well-suited for capturing time-varying behaviors in financial data. By decomposing the time series into these components, we are able to isolate the long-term trends and short-term anomalies, which enhances the model's ability to detect abnormal patterns at different time scales. The low-frequency components are fed into the LSTM network, which learns the long-term dependencies of the time series, while the high-frequency components are input into the Transformer module, where the self-attention mechanism extracts short-term fluctuation features.

3.3 LSTM-Transformer Hybrid Model for Multi-Scale Temporal Dependency Learning

Financial market time series data often exhibit both long-term trend variations and short-term extreme fluctuations, making it challenging for a single model to effectively capture both characteristics. Traditional LSTM, leveraging its gating mechanism, excels at learning long-term dependencies in time series but is limited in detecting short-term market fluctuations and local anomalies. On the other hand, Transformer, with its self-attention mechanism, can effectively model global information, making it well-suited for short-term anomaly detection. However, using Transformer alone may lead to the neglect of long-term dependencies. This module integrates both LSTM and Transformer, combining their strengths to enhance multi-scale time series modeling. The architecture of this module is illustrated in Figure 3.

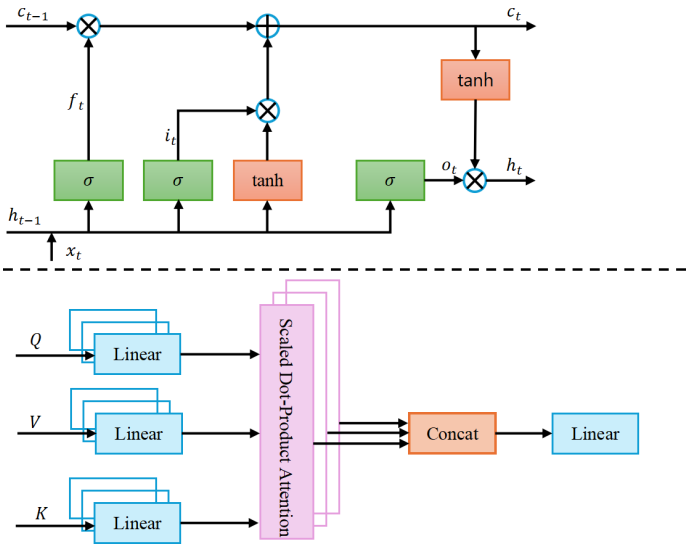

Figure 3. Modeling of multi-scale temporal features.

LSTM is primarily used to capture long-term dependency patterns in financial markets, making it suitable for detecting price trends, market cyclicality, and other structural features. Traditional statistical methods, such as ARIMA and GARCH, struggle to effectively model the nonlinear dynamics of long-term time series, whereas LSTM, through its gating mechanism, controls information storage and forgetting, enabling the learning of long-term temporal features. In this study, Wavelet Transform (WT) is applied to decompose financial time series into low-frequency and high-frequency components. The low-frequency component X_L is fed into the LSTM layer to extract long-term dependency features, which are then passed to the feature fusion layer, where they are integrated with the short-term features extracted by Transformer. LSTM consists of three gates: Forget Gate, Input Gate, and Output Gate, with their computations as follows:

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \tag{5}$$

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \tag{6}$$

$$\tilde{C}_t = \tanh(W_C[h_{t-1}, x_t] + b_C) \tag{7}$$

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \tag{8}$$

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \tag{9}$$

$$h_t = o_t \odot \tanh(C_t) \tag{10}$$

The hidden state h_t serves as the output feature of the LSTM module, which is then passed to the feature fusion layer to be combined with short-term features extracted by Transformer, forming the final time series representation. By leveraging its gating mechanism, LSTM mitigates the vanishing gradient problem, enabling the model to capture long-term trends and enhancing the stability of anomaly detection.

The Transformer module is primarily used for modeling short-term market fluctuations, making it suitable for detecting sudden market changes caused by policy adjustments, market manipulations, or breaking news. Unlike LSTM, which relies on recursive computations to handle temporal dependencies, Transformer utilizes the self-attention mechanism to model global time dependencies and precisely capture short-term fluctuation features. Temporal pattern attention mechanisms have demonstrated strong capability in capturing dependencies across multiple time steps in multivariate series [28]. The input to Transformer is the high-frequency component X_H extracted by Wavelet Transform, which is processed through a multi-head self-attention mechanism to compute the weighted relationships between time steps, effectively focusing on short-term anomaly patterns.

The Transformer module first computes Query (Q), Key (K), and Value (V), where $W_Q, W_K, W_V \in \mathbb{R}^{d \times d_k}$ are learnable parameter matrices, where d is the input embedding dimension and d_k is the dimension of each attention head. It then calculates the Scaled Dot-Product Attention, where $\sqrt{d_k}$ acts as a scaling factor to keep attention scores within an appropriate range, preventing vanishing or exploding gradients:

$$Q = X_H W_Q, \quad K = X_H W_K, \quad V = X_H W_V \tag{11}$$

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^\top}{\sqrt{d_k}}\right) V \tag{12}$$

Table 1. Summary of dataset characteristics.

Dataset	Market Type	Time Granularity	Features	Application
Binance	Cryptocurrency	Minute-level	Price, Volume, Order book	High-frequency anomalies, Market manipulation
SP 500	Stock Market	Daily-level	Price, Volume, Adjusted Close	Long-term trend anomalies, Market crashes

To enhance modeling capability, the Multi-Head Attention mechanism is employed to compute multiple independent attention mappings, allowing the model to capture diverse short-term anomaly patterns more effectively. Finally, the output is passed through a Feed-Forward Network (FFN) for feature transformation, further refining the learned representations

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \ldots, \text{head}_h)W_O \tag{13}$$

$$\text{FFN}(X) = \max(0, XW_1 + b_1)W_2 + b_2 \tag{14}$$

Since Transformer lacks inherent sequential modeling capability, this study employs Positional Encoding to explicitly introduce temporal position information, where t denotes the time step index, i denotes the dimension index, and d denotes the embedding dimension of the model:

$$PE(t, 2i) = \sin\left(\frac{t}{10000^{2i/d}}\right) \tag{15}$$

$$PE(t, 2i+1) = \cos\left(\frac{t}{10000^{2i/d}}\right) \tag{16}$$

The hybrid architecture of WaveLST-Trans integrates LSTM's capability in long-term trend modeling with Transformer's ability to capture short-term fluctuations, while leveraging Wavelet Transform's multi-scale signal decomposition for feature extraction. This design enables the model to efficiently identify complex anomaly patterns in financial time series.

4 Experiments

4.1 Datasets

This study selects two publicly available datasets, Binance cryptocurrency market data [29] and S&P 500 index constituent stock data [30], to comprehensively evaluate the performance of WaveLST-Trans in financial time series anomaly detection. These datasets represent high-frequency trading markets and traditional stock markets, covering financial anomalies across different time scales, allowing the model's applicability to be validated in various market environments. The Binance dataset is sourced from Binance, one of the world's largest cryptocurrency exchanges, and includes minute-level trading data for major cryptocurrencies such as Bitcoin (BTC) and Ethereum (ETH), with features including open price, close price, highest price, lowest price, trading volume, and taker buy volume. The cryptocurrency market is highly volatile, with frequent anomalous trades, making it well-suited for studying short-term extreme fluctuations. This dataset provides the Transformer module with rich short-sequence information, enabling a thorough evaluation of the model's short-term anomaly detection capabilities.

In contrast, the S&P 500 dataset reflects long-term trends in traditional stock markets. This study selects daily-level data for the S&P 500 index and its constituent stocks, including open price, close price, highest price, lowest price, trading volume, and adjusted close price, which collectively represent the overall performance of the U.S. stock market. Unlike cryptocurrency markets, where anomalies manifest as short-term extreme fluctuations, stock market anomalies typically occur over longer time horizons, influenced by financial crises, policy changes, or industry shifts. Additionally, historical stock market data is more comprehensive, covering multiple economic cycles, allowing the experiment to assess the model's robustness and generalization across different market conditions. Table 1 summarizes their key characteristics, including data sources, time granularity, feature dimensions, and applicable scenarios.

For the WaveLST-Trans model, we selected a learning rate of 0.001, a batch size of 64, and the Adam optimizer based on extensive preliminary testing, achieving a balance between model complexity and performance for both Binance and S&P 500 datasets. The Binance dataset was chosen for its high-frequency, volatile nature, ideal for testing the model's ability to detect short-term extreme anomalies such as flash crashes and abnormal trading volume spikes. The S&P 500

Table 2. Comparison of experimental results between WaveLST-Trans and mainstream models on Binance and S&P 500 datasets.

Model	Dataset	Accuracy	Recall	Precision	F1-score	MAE
WaveLST-Trans	Binance	**0.92**	**0.90**	0.87	**0.88**	0.15
	S&P 500	**0.91**	**0.89**	**0.85**	**0.87**	0.18
TFT[31]	Binance	0.90	0.88	**0.88**	0.87	0.14
	S&P 500	0.89	0.85	0.84	0.84	0.19
CSDI[32]	Binance	0.89	0.86	0.85	0.85	**0.13**
	S&P 500	0.88	0.84	0.82	0.83	**0.17**
GDN[33]	Binance	0.88	0.85	0.86	0.85	0.16
	S&P 500	0.87	0.83	0.81	0.82	0.20
TS2Vec[34]	Binance	0.86	0.84	0.82	0.83	0.18
	S&P 500	0.85	0.80	0.79	0.79	0.21
DeepAnT[35]	Binance	0.85	0.82	0.80	0.81	0.19
	S&P 500	0.84	0.79	0.78	0.78	0.23

dataset, with its daily granularity and longer historical coverage, was selected to evaluate the model's capacity to identify structural anomalies and long-term trend deviations across multiple market cycles.

4.2 Comparison Experiments and Analysis

In this experiment, to comprehensively evaluate the performance of WaveLST-Trans in financial time series anomaly detection, we selected TFT (Temporal Fusion Transformer), CSDI (Conditional Score-based Diffusion Models for Time Series), GDN (Graph Deviation Network), TS2Vec (Self-Supervised Learning for Time Series), and DeepAnT as benchmark models. Experiments were conducted on both the Binance cryptocurrency market dataset and the S&P 500 stock market dataset.These benchmark models represent mainstream approaches in recent years for time series modeling, anomaly detection, and forecasting, offering strong capabilities in capturing long-and-short-term dependencies. The evaluation was based on five key metrics: Accuracy, Recall, Precision, F1-score, and Mean Absolute Error (MAE). The experimental results are summarized in Table 2. From the experimental results, it is evident that WaveLST-Trans demonstrates strong stability and superiority across all evaluation metrics. Particularly in terms of accuracy, recall, and F1-score, the model outperforms the benchmark models, indicating its strong generalization capability in financial time series anomaly detection.On the Binance dataset (cryptocurrency market), WaveLST-Trans achieves a 6% improvement in recall compared to TS2Vec and an 8% increase compared to DeepAnT, proving its effectiveness in capturing short-term extreme fluctuations. In contrast, TS2Vec, based on self-supervised learning, and DeepAnT, a CNN-based unsupervised anomaly detection model, exhibit weaker performance in detecting short-period anomalies. Additionally, in F1-score, WaveLST-Trans surpasses GDN by 3%, demonstrating its ability to balance precision and recall effectively in anomaly detection tasks.

On the S&P 500 dataset (traditional stock market), WaveLST-Trans also outperforms other models in precision and F1-score, achieving 3% and 5% higher F1-score than TFT and CSDI, respectively. This suggests that the model can effectively reduce false positive rates when detecting long-term trend deviations. Although TFT performs slightly better in some metrics (e.g., precision), its lower recall indicates that while it models long-term trends effectively, it struggles to identify abnormal transactions or market anomalies. Additionally, CSDI achieves the best MAE scores (0.13 on the Binance dataset and 0.17 on the S&P 500 dataset), highlighting the strength of score-based diffusion models in reconstruction quality, but showing no significant advantage in discrete anomaly classification tasks.

As illustrated in Figure 4, WaveLST-Trans combines the long-term dependency modeling capability of LSTM with the short-term fluctuation capturing ability of Transformer, allowing it to perform well across different market environments. Compared to CNN-based unsupervised models (e.g., DeepAnT), it provides more stable detection results; compared to self-supervised learning methods (e.g., TS2Vec), it offers better adaptability to both short-term and long-term patterns; and compared to score-based diffusion models (e.g., CSDI), it outperforms in

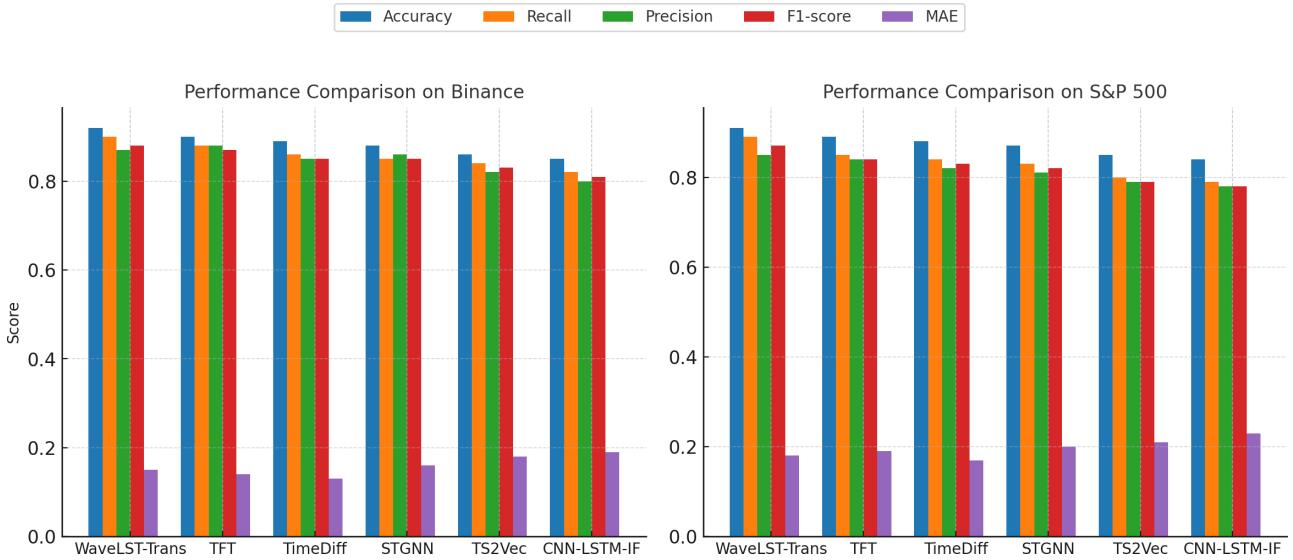

Figure 4. Performance comparison on Binance and S&P 500 datasets.

Table 3. Ablation study results on Binance and S&P 500 datasets.

Model Variant	Dataset	Accuracy	Recall	Precision	F1-score	MAE
WaveLST-Trans (Full Model)	Binance	**0.92**	**0.90**	**0.87**	**0.88**	**0.15**
	S&P 500	**0.91**	**0.89**	**0.85**	**0.87**	**0.18**
w/o Wavelet Transform (WT)	Binance	0.89	0.86	0.85	0.85	0.17
	S&P 500	0.88	0.85	0.83	0.84	0.20
w/o LSTM	Binance	0.87	0.84	0.83	0.83	0.18
	S&P 500	0.86	0.82	0.81	0.82	0.21
w/o Transformer	Binance	0.86	0.82	0.84	0.83	0.19
	S&P 500	0.85	0.81	0.80	0.80	0.22
w/o WT + LSTM	Binance	0.84	0.80	0.78	0.79	0.22
	S&P 500	0.82	0.78	0.76	0.77	0.24
w/o WT + Transformer	Binance	0.83	0.78	0.77	0.78	0.24
	S&P 500	0.83	0.76	0.74	0.75	0.26
w/o LSTM + Transformer	Binance	0.82	0.77	0.75	0.76	0.25
	S&P 500	0.80	0.75	0.72	0.74	0.27

anomaly classification tasks. Overall, WaveLST-Trans consistently delivers superior performance in financial anomaly detection across various market conditions, demonstrating strong applicability and robustness.

4.3 Ablation Study and Analysis

In this experiment, to verify the effectiveness of each module in WaveLST-Trans, we designed a series of ablation experiments by removing different key components of the model and evaluating their impact on performance. The experiments were conducted on the datasets to assess how each module contributes to the overall model performance. The primary objective of the ablation study is to analyze the individual contributions of each component to the model's effectiveness. The experimental results are summarized in Table 3.

The ablation study results presented in Table 3 confirm the effectiveness of each component within the WaveLST-Trans model. On the Binance dataset, removing the Wavelet Transform (WT) led to a drop in accuracy from 0.92 to 0.89 and in F1-score from 0.88 to 0.85, underscoring its role in capturing multi-scale features. Excluding either the LSTM or Transformer module resulted in an F1-score of 0.83 in both cases, which highlights their complementary strengths in modeling long-term trends and short-term fluctuations. When both LSTM and Transformer were removed, the F1-score decreased to 0.76, approaching the performance of traditional machine learning methods. On the S&P 500 dataset, similar patterns were observed. The absence of WT led to a decrease in F1-score from 0.87 to 0.84, while removing LSTM or Transformer caused F1-scores of 0.82 and 0.80,

respectively. Notably, removing both WT and LSTM resulted in an F1-score of 0.77, while removing both LSTM and Transformer produced the lowest F1-score of 0.74, reinforcing the critical role of these components in both long-term pattern extraction and short-term anomaly detection. Overall, these results demonstrate that the full integration of WT, LSTM, and Transformer is critical to the model's superior performance across different types of financial time series.

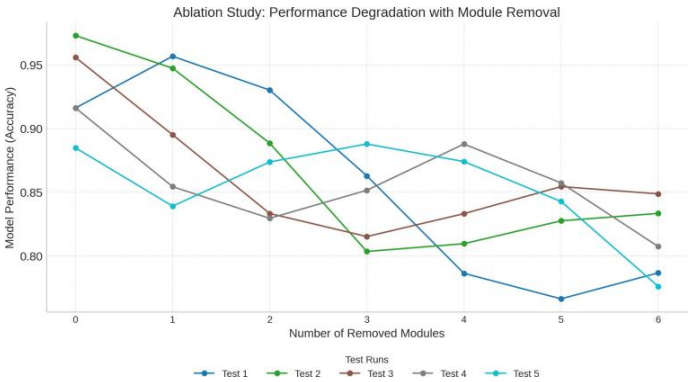

Figure 5. Ablation study: performance degradation with module removal.

As illustrated in Figure 5, the ablation study results validate the synergy among WT, LSTM, Transformer, and the feature fusion layer in WaveLST-Trans. Compared to single-model approaches, the full architecture excels in capturing short-term anomalies, long-term trend changes, and multi-scale information fusion. Regardless of whether in high-frequency trading markets or long-term investment markets, the complete model outperforms all individual module variants, further verifying the rationality of the proposed model architecture.

5 Conclusion And Discussion

This paper proposes WaveLST-Trans, a financial time series anomaly detection model based on an LSTM-Transformer hybrid architecture with wavelet transform (WT). The model first applies WT for multi-scale decomposition, feeding low-frequency components into LSTM to model long-term dependencies and high-frequency components into Transformer to capture short-term market fluctuations. These extracted features are then integrated in the feature fusion layer, and the anomaly detection module outputs an anomaly score. Compared to existing methods, WaveLST-Trans effectively models anomalies across different time scales, and ablation studies confirm the necessity of WT, LSTM, Transformer, and the feature fusion layer.Experiments on Binance (cryptocurrency market) and S&P 500 (stock market) datasets show that WaveLST-Trans outperforms mainstream models, improving F1-score and recall by 3%–10% in high-frequency markets and accuracy by 5%–9% in long-term trend markets.

The model maintains strong generalization and robustness across different financial environments. Future work will focus on optimizing the Transformer structure, adaptive anomaly thresholds, and integrating external market events to enhance adaptability. Additionally, further studies are needed to assess its performance under extreme financial events to ensure stability in sudden risk management.

Data Availability Statement

This study uses publicly available datasets. The Binance cryptocurrency market data are accessible via the Binance API (https://developers.binance.com/docs/binance-spot-api-docs). The S&P 500 index data are publicly available via Yahoo Finance (https://finance.yahoo.com/quote/%5EGSPC/history/).

Funding

This work was supported without any funding.

Conflicts of Interest

Tian Su is affiliated with the Meta Platforms Inc., Seattle, WA 98109, United States. The authors declare that this affiliation did not influence the study design, data collection, analysis, interpretation of the results, or the decision to publish this work. The authors further declare that they have no other competing interests.

Ethical Approval and Consent to Participate

Not applicable.

References

[1] Ahmed, M., Choudhury, N., & Uddin, S. (2017, July). Anomaly detection on big data in financial markets. In *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017* (pp. 998-1001). [CrossRef]

[2] Chatzis, S. P., Siakoulis, V., Petropoulos, A., Stavroulakis, E., & Vlachogiannakis, N. (2018). Forecasting stock market crisis events using deep and statistical machine learning techniques. *Expert systems with applications*, *112*, 353-371. [CrossRef]

[3] Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market

predictions. *European journal of operational research, 270*(2), 654-669. [CrossRef]

[4] Cont, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative finance, 1*(2), 223. [CrossRef]

[5] Yıldız, K., Dedebek, S., Okay, F. Y., & Şimşek, M. U. (2022, September). Anomaly detection in financial data using deep learning: A comparative analysis. In *2022 Innovations in Intelligent Systems and Applications Conference (ASYU)* (pp. 1-6). IEEE. [CrossRef]

[6] Alghofaili, Y., Albattah, A., & Rassam, M. A. (2020). A financial fraud detection model based on LSTM deep learning technique. *Journal of Applied Security Research, 15*(4), 498-516. [CrossRef]

[7] Mubalaike, A. M., & Adali, E. (2018, September). Deep learning approach for intelligent financial fraud detection system. In *2018 3rd International Conference on Computer Science and Engineering (UBMK)* (pp. 598-603). IEEE. [CrossRef]

[8] Torres, J. F., Hadjout, D., Sebaa, A., Martínez-Álvarez, F., & Troncoso, A. (2021). Deep learning for time series forecasting: A survey. *Big Data, 9*(1), 3–21. [CrossRef]

[9] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems, 30*.

[10] Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., & Zhang, W. (2021, May). Informer: Beyond efficient transformer for long sequence time-series forecasting. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 35, No. 12, pp. 11106-11115). [CrossRef]

[11] Masini, R. P., Medeiros, M. C., & Mendes, E. F. (2023). Machine learning advances for time series forecasting. *Journal of Economic Surveys, 37*(1), 76–111. [CrossRef]

[12] Yu, H., Ming, L. J., Sumei, R., & Shuping, Z. (2020). A hybrid model for financial time series forecasting—integration of EWT, ARIMA with the improved ABC optimized ELM. *IEEE Access, 8*, 84501-84518. [CrossRef]

[13] Yang, T., Li, A., Xu, J., Su, G., & Wang, J. (2024). Deep learning model-driven financial risk prediction and analysis. *Applied and Computational Engineering*, 77, 196–202. [CrossRef]

[14] Guo, T., Zhang, T., Lim, E., Lopez-Benitez, M., Ma, F., & Yu, L. (2022). A review of wavelet analysis and its applications: Challenges and opportunities. *IEEE Access, 10*, 58869–58903. [CrossRef]

[15] Bao, W., Yue, J., & Rao, Y. (2017). A deep learning framework for financial time series using stacked autoencoders and long-short term memory. *PloS one, 12*(7), e0180944. [CrossRef]

[16] Wu, H. S. (2016, December). A survey of research on anomaly detection for time series. In *2016 13th international computer conference on wavelet active media technology and information processing (ICCWAMTIP)* (pp. 426-431). IEEE. [CrossRef]

[17] Blázquez-García, A., Conde, A., Mori, U., & Lozano, J. A. (2021). A review on outlier/anomaly detection in time series data. *ACM computing surveys (CSUR), 54*(3), 1-33. [CrossRef]

[18] Xu, H., Pang, G., Wang, Y., & Wang, Y. (2023). Deep isolation forest for anomaly detection. *IEEE Transactions on Knowledge and Data Engineering, 35*(12), 12591–12604. [CrossRef]

[19] Samariya, D., & Thakkar, A. (2023). A comprehensive survey of anomaly detection algorithms. *Annals of Data Science, 10*(3), 829–850. [CrossRef]

[20] Qiao, Y., Wu, K., & Jin, P. (2021). Efficient anomaly detection for high-dimensional sensing data with one-class support vector machine. *IEEE Transactions on Knowledge and Data Engineering, 35*(1), 404–417. [CrossRef]

[21] Tang, Q., Fan, T., Shi, R., Huang, J., & Ma, Y. (2021). Prediction of financial time series using LSTM and data denoising methods. *arXiv preprint arXiv:2103.03505*. [CrossRef]

[22] Reddy, N. M., Sharada, K. A., Pilli, D., Paranthaman, R. N., Reddy, K. S., & Chauhan, A. (2023, June). CNN-Bidirectional LSTM based Approach for Financial Fraud Detection and Prevention System. In *2023 International Conference on Sustainable Computing and Smart Systems (ICSCSS)* (pp. 541-546). IEEE. [CrossRef]

[23] Xu, J., Wu, H., Wang, J., & Long, M. (2021). Anomaly transformer: Time series anomaly detection with association discrepancy. *arXiv preprint arXiv:2110.02642*. [CrossRef]

[24] Karim, F., Majumdar, S., Darabi, H., & Chen, S. (2017). LSTM fully convolutional networks for time series classification. *IEEE Access, 6*, 1662-1669. [CrossRef]

[25] Ma, X., Wu, J., Xue, S., Yang, J., Zhou, C., Sheng, Q. Z., ... & Akoglu, L. (2021). A comprehensive survey on graph anomaly detection with deep learning. *IEEE Transactions on Knowledge and Data Engineering, 35*(12), 12012–12038. [CrossRef]

[26] Wu, D., Wang, X., & Wu, S. (2022). A hybrid framework based on extreme learning machine, discrete wavelet transform, and autoencoder with feature penalty for stock prediction. *Expert Systems with Applications, 207*, 118006. [CrossRef]

[27] Li, J., Liu, Y., Gong, H., & Huang, X. (2024). Stock price series forecasting using multi-scale modeling with boruta feature selection and adaptive denoising. *Applied Soft Computing, 154*, 111365. [CrossRef]

[28] Shih, S. Y., Sun, F. K., & Lee, H. Y. (2019). Temporal pattern attention for multivariate time series forecasting. *Machine Learning, 108*, 1421-1441. [CrossRef]

[29] Binance. (2024). Binance API documentation: Market

data endpoints. Binance Developer Portal. Retrieved from https://developers.binance.com/docs/binance-spot-api-docs

[30] Yahoo Finance. (2024). S&P 500 historical data. Yahoo Finance Market Data. Retrieved from https://finance.yahoo.com/quote/%5EGSPC/history/

[31] Lim, B., Arık, S. Ö., Loeff, N., & Pfister, T. (2021). Temporal fusion transformers for interpretable multi-horizon time series forecasting. *International Journal of Forecasting*, *37*(4), 1748–1764. [CrossRef]

[32] Audibert, J., Michiardi, P., Guyard, F., Marti, S., & Zuluaga, M. A. (2020, August). Usad: Unsupervised anomaly detection on multivariate time series. In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining* (pp. 3395-3404). [CrossRef]

[33] Deng, A., & Hooi, B. (2021, May). Graph neural network-based anomaly detection in multivariate time series. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 35, No. 5, pp. 4027-4035). [CrossRef]

[34] Yue, Z., Wang, Y., Duan, J., Yang, T., Huang, C., Tong, Y., & Xu, B. (2022, June). Ts2vec: Towards universal representation of time series. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 36, No. 8, pp. 8980-8987). [CrossRef]

[35] Munir, M., Siddiqui, S. A., Dengel, A., & Ahmed, S. (2018). DeepAnT: A deep learning approach for unsupervised anomaly detection in time series. *IEEE access*, *7*, 1991-2005. [CrossRef]

Tian Su received the B.Eng. degree in Computer Science and Technologies from Southeast University, Chien-Shiung Wu College, Nanjing, China in 2019 and the M.Sc. degree in Information Networking from Carnegie Mellon University, Pittsburgh, PA, in 2021. She is currently working on building creator video optimization and data analytics systems with Meta Platform Inc. (Email: tiansu1997@gmail.com)

Runlong Li received the B.Eng. degree from Nanjing Agricultural University, Nanjing, China, in 2020, and the M.Sc. degree in Computer Engineering from the University of California, Irvine, in Los Angeles, United States, in 2022. He is currently engaged in the development of anomaly detection systems at Onward Search, Inc., focusing on practical applications of machine learning in real-world data environments. (Email: runlonl@uci.edu)

Bo Liu received her M.Sc. degree in Computer Software Engineering from Northeastern University, Boston, USA, in 2023. She is currently working at Harness, Inc., an AI-powered software delivery platform, where she focuses on developing machine learning and data analytics systems. (Email: liu.bo4@northeastern.edu)

Xiaoxiang Liang received his Master's degree in Business Analytics from Washington University in St. Louis, Missouri, in 2022. Currently, he is working at Capital One, Inc., where he focus on innovating, developing, and maintaining robust data solutions to address complex business challenges. Leveraging cutting-edge machine learning technologies, Xiaoxiang works to uncover valuable opportunities that empower everyday individuals to save money, time, and reduce the complexities in their financial lives. (Email: l.xiaoxiang@wustl.edu)

Xinhao Yang received his B.Eng. and M.Eng. degrees in Geodesy from Wuhan University, China, in 2017 and 2021, respectively, and an M.Sc. degree in Computer Science from the University of Southern California in 2023. He is currently working on Generative AI and Machine Learning, focusing on AI agents and their applications. His expertise includes building scalable AI infrastructure, developing and optimizing ML models, and integrating GenAI solutions to enhance automation and decision-making processes. (Email: xinhaoya@usc.edu)

Yan Zhou received her Bachelor of Science degree in Mathematics from Ocean University of China in 2016 and her Master of Science degree in Mathematics from Northeastern University in Boston, USA in 2018. With over five years of experience as a data scientist, she has successfully tackled complex business challenges in areas such as pricing, underwriting, marketing, and sales, driving substantial revenue incremental growth for various companies.Furthermore, Yan Zhou is currently preparing to venture into entrepreneurship, supplementing her expertise with knowledge in entrepreneurship and finance to establish a technology company. (Email: zhou.yan2@northeastern.edu)