Journal of Computer and Communications, 2025, 13(12), 100-114
https://www.scirp.org/journal/jcc
ISSN Online: 2327-5227
ISSN Print: 2327-5219

Explore Anomaly-Aware Transformers for Robust Financial Time Series Forecasting

Qizhao Chen

Graduate School of Information Science, University of Hyogo, Kobe, Japan
Email: af24o008@guh.u-hyogo.ac.jp

How to cite this paper: Chen, Q. (2025) Explore Anomaly-Aware Transformers for Robust Financial Time Series Forecasting. *Journal of Computer and Communications*, **13**, 100-114.
https://doi.org/10.4236/jcc.2025.1312006

Received: November 24, 2025
Accepted: December 19, 2025
Published: December 22, 2025

Abstract

This study investigates whether anomaly-aware modeling can improve stock price forecasting by incorporating signals that highlight unusual market behavior. Financial time series often contain sudden jumps, heavy tails, and irregular movements that challenge standard neural models. To address this issue, an Anomaly-Aware Transformer (AAT) is developed by combining two unsupervised anomaly detection methods—Isolation Forest and a rolling Z-score filter—to identify periods of abnormal volatility. These anomaly flags are then added as auxiliary inputs to a Transformer model and used to guide the attention mechanism toward market conditions that deviate from typical patterns. The performance of the AAT is evaluated on twenty large companies from different industries and compared with a standard Transformer baseline. Results show that while the baseline model performs better on normal trading days, the AAT provides clear advantages for several stocks during anomalous periods, especially when the abnormal movements follow detectable structures. The overall findings suggest that anomaly-aware information can enhance robustness under extreme market conditions, although the benefits depend on the volatility profile of individual assets. This highlights the potential of integrating anomaly detection with deep learning for financial forecasting and points to future directions such as adaptive state-dependent models and hybrid anomaly handling strategies.

Keywords

Financial Time Series Forecasting, Anomaly Detection, Isolation Forest, Rolling Z-Score, Transformer

1. Introduction

Financial markets are constantly changing and are shaped by diverse elements such as economic trends, individual company developments, and the mood of investors [1]. Accurately predicting stock price movements is a long-standing challenge in finance. Reliable forecasts can help investors make informed decisions, manage risk, and develop trading strategies [2]. Traditional financial models, such as generalized autoregressive conditional heteroskedasticity (GARCH) [3] models or the autoregressive integrated moving average (ARIMA) [4], often assume linear relationships and stationary time series. While these models provide some insight, they struggle to capture complex, nonlinear patterns in financial data, especially during periods of high volatility [5]. In recent years, machine learning has become increasingly popular in financial forecasting. Models such as recurrent neural networks (RNNs), long short-term memory networks (LSTMs) [6], and convolutional neural networks (CNNs) [7]-[9] have shown promise in modeling temporal dependencies and capturing nonlinear dynamics [10]. Among these, Transformer-based models have gained attention due to their ability to learn long-range dependencies and process sequences in parallel. Transformers [11] use self-attention mechanisms to assign weights to past observations, allowing the model to focus on the most relevant information for prediction. This architecture has led to significant improvements in areas such as natural language processing, and its application in finance is growing. Despite these advancements, predicting financial time series remains challenging due to market anomalies and extreme events. Anomalous days, such as those triggered by sudden news, macroeconomic shocks, or unexpected earnings reports [12], can produce abrupt price movements. Standard machine learning models often treat all data points equally, which can lead to poor performance during these abnormal periods. Ignoring anomalous events can result in models that fail to generalize, underestimating risk and generating inaccurate forecasts.

To address this problem, we propose an anomaly-aware Transformer (AAT) framework that integrates unsupervised anomaly detection with Transformer-based forecasting. The key idea is to identify days with abnormal market behavior and explicitly incorporate this information into the model. This allows the Transformer to pay more attention to normal patterns while adjusting for unusual market events. By doing so, the model can maintain predictive accuracy during stable periods while remaining robust during volatile or extreme market conditions.

The proposed framework uses two complementary anomaly detection methods. The first is Isolation Forest [13], a tree-based approach that isolates observations in feature space. Anomalies are easier to separate because they are few and differ significantly from the majority of the data. The second method is a rolling z-score, which identifies extreme deviations in the univariate return series relative to a local window. By combining these methods, we can detect both multivariate anomalies that involve multiple features, such as returns and trading volume, and univariate anomalies that reflect extreme price movements. The anomaly indicators are then used as additional inputs to the Transformer, allowing the model to adjust its attention dynamically.

This research offers three primary contributions. First, we develop a detailed pipeline for constructing features from historical stock data, including log returns, rolling volatility, volume changes, and technical indicators. Second, we propose an anomaly-aware Transformer that explicitly incorporates anomaly flags and scores. This architecture allows the model to reduce the influence of extreme events on predictions while still learning from the overall market dynamics. Third, we conduct an extensive evaluation across multiple publicly traded stocks to evaluate the performance of the model. The evaluation considers directional accuracy, which assesses the ability of the model to predict the direction of returns, along with standard metrics such as mean squared error and mean absolute error. We further evaluate performance separately for normal and anomalous days, providing insight into how the model handles extreme market conditions.

The study strives to integrate traditional forecasting models with modern machine learning approaches, focusing on the challenges arising from unusual or extreme events. Unlike conventional models that treat all data points equally, our approach acknowledges that not all days are equally informative. By combining anomaly detection with Transformer modeling, the framework provides a more robust and interpretable method for predicting financial time series. We believe this approach is particularly useful for investors, risk managers, and algorithmic traders who need reliable forecasts in both stable and volatile market periods.

The rest of this paper is organized as follows. Section 2 provides a review of related research in financial forecasting, Transformer models, and anomaly detection. Section 3 details the proposed methodology, covering feature construction, anomaly detection, model architecture, and evaluation procedures. Section 4 presents the results and discusses their significance. Finally, Section 5 concludes the study and suggests directions for future research.

2. Related Work

Forecasting financial time series has been a central topic in quantitative finance for decades. Traditional statistical models, such as autoregressive and moving average models [14], have been widely applied due to their simplicity and interpretability. These models generally assume linear relationships and rely on stationary time series, which can limit their effectiveness in capturing the complex dynamics of financial markets. For example, Almaafi *et al.* [15] compare ARIMA and XGBoost for forecasting Saudi Telecom's weekly closing prices, finding that XGBoost consistently delivers superior performance. The results underscore the limitations of traditional statistical models and highlight the advantages of machine learning in capturing complex patterns in financial data.

In financial forecasting, machine learning methods have gained prominence for their ability to capture intricate and nonlinear patterns. Techniques like support vector machines [16], random forests [17], and neural networks [18] have demonstrated the capacity to enhance predictive performance by extracting patterns from historical data. LSTM [19] and GRU [20] networks, as types of recurrent neural networks, have been extensively applied for modeling time-dependent sequential relationships. Convolutional neural networks [21] have also been applied to capture local patterns in time series data. Despite these advances, conventional machine learning models may still underperform in extreme market conditions, as they typically treat all data points equally and do not explicitly consider anomalous or rare events that can significantly affect returns.

Originally, Transformers were created to address challenges in natural language processing. In the context of financial data, Transformers have been applied to predict stock returns, volatility, and market trends [22] [23]. For example, Ji *et al.* [24] propose a new Transformer-based model called Galformer to improve multi-step stock index forecasting by introducing a generative decoder for faster long-horizon prediction and a hybrid loss that captures both numerical precision and trend direction. Experiments on four major indices show that this design enhances forecasting performance compared with conventional approaches. Hu [25] examines stock price forecasting by comparing a Temporal Fusion Transformer (TFT) with SVR and LSTM, highlighting the value of incorporating static metadata alongside time-series inputs. Evaluation using MSE and SMAPE shows that TFT delivers the most accurate predictions among the tested models. Olorunnimbe and Viktor [26] propose an ensemble strategy that boosts the accuracy of temporal Transformer models for multi-horizon financial forecasting by training multiple models on sliding windows and combining them through averaging and a stacking meta-learner. By decomposing long time series and optimizing model weighting, the approach simplifies training while delivering substantially better performance across varying market conditions.

However, in previous study, Transformer models do not explicitly account for abnormal market events, which can limit their robustness during periods of high volatility or sudden shocks.

Anomaly detection has been widely studied in financial applications to identify unusual patterns that may indicate market stress or rare events [27]. Various methods, including statistical techniques and machine learning algorithms, have been developed to detect anomalies in time series data. For example, Zhang *et al.* [28] introduce an anomaly-aware domain generalization method for financial time series, designed to handle the diverse and irregular feature distributions created by differing accounting standards. By leveraging RNN-derived representations, adapting them with a classifier, and refining forecasts through a latent variable autoregressive model, the approach improves both trend and risk prediction, demonstrating strong potential for managing complex financial data. Cao and Guo [29] present a curiosity-driven model selection framework for multivariate financial anomaly detection that overcomes the limits of single-assumption and label-dependent methods. By exploring a search space of diverse anomaly models, applying reinforcement learning with self-imitation and experience replay, the system chooses the best model at each time step and delivers stronger detection accuracy and interpretability across multiple datasets.

Despite the growing interest in combining anomaly detection with deep learning, there remains a lack of comprehensive studies applying this approach to Transformer-based financial forecasting. Most existing work focuses either on improving forecasting accuracy using standard Transformers or on detecting anomalies as a separate preprocessing step. Few studies explicitly integrate anomaly information into the model architecture to guide learning. Addressing this gap is essential for developing robust financial forecasting systems that can perform well under both normal and extreme market conditions.

3. Methodology

3.1. Data Collection and Processing

Using the Python library yfinance, daily historical stock prices for 20 companies—such as Apple, Microsoft, Tesla, Alibaba, Pepsi, and HSBC—were collected from Yahoo Finance. The analysis period spans from January 1, 2016, to December 31, 2024, covering multiple market regimes, including both calm periods and highly volatile periods. The retrieved fields include Open, High, Low, Close, and Volume. All subsequent features and calculations are derived from the Close price, while volume is incorporated as an additional input. Unlike some studies that rely on adjusted Close prices, the present framework directly computes returns and technical indicators from the raw Close series, and the model is tasked with forecasting next-day log returns.

Table 1. Description of retrieved stock price features and technical indicators.

Feature/Indicator	Description
Open	The stock price at the beginning of the trading session.
High	The highest price reached during the trading session.
Low	The lowest price observed during the trading session.
Close	The final stock price at the end of the trading session.
Volume	The total number of shares traded during the trading session.
SMA	Simple Moving Average over window size n: $\text{SMA}_t = \frac{1}{n}\sum_{i=0}^{n-1}\text{Close}_{t-i}$.
EMA	Exponential Moving Average with smoothing factor α: $\text{EMA}_t = \alpha \cdot \text{Close}_t + (1-\alpha)\cdot \text{EMA}_{t-1}$.
Return	Simple return: $R_t = \frac{\text{Close}_t - \text{Close}_{t-1}}{\text{Close}_{t-1}}$.
Log Return	Logarithmic return: $r_t = \ln\frac{\text{Close}_t}{\text{Close}_{t-1}}$.

To capture both price dynamics and local volatility patterns, the Close price is used to compute daily returns, defined as simple percentage changes, and log returns, defined as the natural logarithm of the ratio between consecutive Close prices. Rolling volatility measures are computed over 5-day and 20-day windows to reflect short- and mid-term market variability, and changes in trading volume are also included. Additionally, simple technical indicators, including a 5-day simple moving average and a 5-day exponential moving average of the Close price, are calculated. The resulting multivariate time series is standardized based on the training set statistics to prevent data leakage, and a sliding window of twenty days is applied to generate sequences for Transformer input. Each sequence consists of twenty days of historical features and is aligned to predict the next-day log return, providing a one-step-ahead forecasting target. Table 1 shows the description of each feature. In general, the features used in prediction include Log Return, 5-day volatility, 20-day volatility, volume, 5-day SMA, and 5-day EMA.

3.2. Anomaly Detection Methods

To strengthen the Transformer's ability to deal with extreme market conditions, we include an unsupervised anomaly detection step to flag days with unusual price or volume movements. We use two complementary methods: Isolation Forest and the rolling Z-score. Isolation Forest is effective for capturing multivariate and nonlinear abnormal patterns because it separates rare observations based on how easily they can be isolated in a random tree structure. In contrast, the rolling Z-score focuses on simple, fast, and interpretable deviations in each individual series, allowing the model to detect sudden jumps or drops relative to recent history. Using both methods together helps us identify a wider range of anomalies in financial time series, covering complex cross-feature irregularities as well as clear univariate outliers.

The first method, *Isolation Forest* (IF), is a tree-based ensemble algorithm specifically designed to isolate anomalous observations. Unlike traditional distance-based anomaly detection methods, IF algorithm works on the principle that outliers are infrequent and differ significantly from the majority, allowing them to be isolated more effectively. Formally, given a feature matrix $\boldsymbol{X} \in \mathbb{R}^{T \times F}$, where T is the number of days and F is the number of features (in this study, log returns, short-term volatility, and volume change), Isolation Forest works by forming multiple binary trees, where the data is split recursively using randomly selected features and thresholds. For a given data point $\boldsymbol{x}_i$, its path length $h(\boldsymbol{x}_i)$ refers to the count of edges traveled from the root node to the leaf where the observation $\boldsymbol{x}_i$ ends. The expected path length across a forest of n trees is then used to compute the anomaly score:

$$s(\boldsymbol{x}_i, n) = 2^{-\frac{E[h(\boldsymbol{x}_i)]}{c(n)}},$$

where $E[h(\boldsymbol{x}_i)]$ is the average path length of $\boldsymbol{x}_i$ over the ensemble of trees, while $c(n)$ estimates the average path length for unsuccessful searches in a binary search tree containing n data points, approximated by:

$$c(n) = 2H_n - \frac{2(n-1)}{n},$$

with H_n denoting the n-th harmonic number. The anomaly score $s(\boldsymbol{x}_i)$ lies in the interval $[0,1]$, where values close to 1 indicate a high likelihood of being an anomaly. In this framework, an Isolation Forest contamination rate of 2% is set, meaning the algorithm is tuned to flag approximately 2% of the data as anomalous, consistent with the expectation that extreme market events are rare. A binary flag is generated by thresholding the anomaly score:

$$\text{flag}^{\text{IF}}(\boldsymbol{x}_i) = \begin{cases} 1, & s(\boldsymbol{x}_i) > \tau_{\text{IF}}, \\ 0, & \text{otherwise}, \end{cases}$$

where τ_{IF} is determined by the contamination level.

The second method, the *rolling z-score*, is a classical univariate approach applied to the log return series to detect statistically significant deviations from the local mean. For each day t, the rolling mean μ_t and rolling standard deviation σ_t over a window of w days (here $w = 60$) are computed as:

$$\mu_t = \frac{1}{w}\sum_{i=t-w+1}^{t} r_i, \quad \sigma_t = \sqrt{\frac{1}{w}\sum_{i=t-w+1}^{t} (r_i - \mu_t)^2},$$

where r_i denotes the log return on day i. The corresponding z-score is then defined as:

$$z_t = \frac{r_t - \mu_t}{\sigma_t}.$$

Days with absolute z-scores exceeding a threshold θ (here $\theta = 3$) are flagged as anomalous:

$$\text{flag}_t^{Z} = \begin{cases} 1, & |z_t| > \theta, \\ 0, & \text{otherwise}. \end{cases}$$

The absolute z-score can also be normalized to the interval $[0,1]$ for use as an anomaly score, providing a continuous measure of abnormality analogous to the Isolation Forest score.

Finally, to combine the two complementary approaches, the binary anomaly flags are merged using a logical OR operation, ensuring that days flagged by either method are considered anomalous. The anomaly score used as input to the Transformer is defined as the element-wise maximum of the normalized Isolation Forest and rolling z-score outputs. This integration allows the model to simultaneously account for multivariate anomalies captured by Isolation Forest and univariate extremes captured by the rolling z-score, providing a robust mechanism to highlight days with abnormal market behaviors.

3.3. Prediction Models

The core predictive model is a Transformer encoder that processes sequences of 20 time steps. Each input window is first projected into a 64-dimensional embedding space, and positional embeddings are added to preserve the temporal order. The encoder contains three layers with four attention heads and a feed-forward dimension of 128, with a dropout rate of 0.1 to reduce overfitting.

To incorporate anomaly information, we introduce the anomaly flags and scores as an additional input channel that directly modulates the attention weights. At each attention layer, the anomaly score for time step t is transformed into a penalty term and subtracted from the attention logits before the softmax operation. This mechanism lowers the attention weights assigned to time steps with higher anomaly scores, reducing the model's reliance on observations that reflect abnormal or unstable market conditions.

The final sequence representations are aggregated using mean pooling and passed to a fully connected prediction head to generate the next-day log-return forecast.

A baseline Transformer model with the same architecture is trained without anomaly features, enabling a direct comparison to evaluate the contribution of anomaly awareness to forecasting performance. Both models are trained to minimize the mean squared error between predicted and actual log returns. The model is trained in batches of 64, employing the AdamW optimizer with a learning rate of 1×10^{-4} and a weight decay parameter set to 1×10^{-5}. Training continues for up to 60 epochs, and early stopping is activated when there is no improvement in validation metrics over eight successive epochs.

3.4. Evaluation

The dataset is split in temporal order into training, validation, and test sets for model assessment. Seventy percent of the samples, corresponding to the period from 2016 to 2021, are used for training, allowing the model to learn underlying patterns and calibrate anomaly detection. The following fifteen percent, covering 2022 to 2023, serve as the validation set for hyperparameter tuning and early stopping, while the remaining fifteen percent, corresponding to 2024, are reserved for out-of-sample testing. This chronological splitting ensures that future information does not contaminate the model's learning process.

The evaluation of performance is conducted with three distinct metrics. Mean Squared Error (MSE) measures the average of the squared differences between predicted and actual log returns, emphasizing larger errors and penalizing extreme deviations. Mean Absolute Error (MAE) quantifies the average absolute differences between predictions and actual values, providing a linear measure of prediction accuracy that is less sensitive to outliers than MSE. Directional Accuracy (DirAcc) evaluates the model's ability to correctly predict the sign of the return, i.e., whether the next-day price will increase or decrease, which is particularly relevant in financial applications where correctly anticipating the market direction can guide trading decisions. It is computed as:

$$\text{DirAcc} = \frac{1}{N}\sum_{i=1}^{N} \mathbb{I}\left[\text{sign}\left(\hat{y}_i\right) = \text{sign}\left(y_i\right)\right],$$

where $\mathbb{I}(\cdot)$ is the indicator function, $\hat{y}_i$ is the predicted value, and y_i is the true value.

To provide a clearer view of how market irregularities affect model behavior, the evaluation separates forecasting performance into normal days and anomalous days. Normal days refer to trading periods in which price movements fall within typical volatility ranges and are not flagged by the anomaly detection module. In contrast, anomalous days represent periods where the market exhibits unusually large or irregular movements, as identified jointly by the Isolation Forest scores and the rolling z-score filter. For every prediction, the anomaly status is determined by the label of the target day being forecasted, ensuring that the evaluation reflects the actual conditions the model is attempting to predict.

4. Results and Discussion

Table 2 shows the model performances between normal and anomalous days in terms of MSE and Figure 1 to Figure 3 visualize the results from the table. As MAE exhibits patterns consistent with those of MSE, we present only the MSE results for clarity. Table 3 demonstrates the model performances between normal and anomalous days in terms of DirAcc.

Table 2. Comparison of normal and anomalous day forecasting MSE for multiple stocks.

Stocks	Baseline Norm. MSE	AAT Norm. MSE	Baseline Anom. MSE	AAT Anom. MSE
Apple	0.000197	0.000261	0.000215	0.000344
Microsoft	0.000257	0.000260	0.000599	0.000042
Tesla	0.001418	0.001733	0.001494	0.002989
Alibaba	0.000541	0.001055	0.000475	0.001009
Pepsi	0.000117	0.000114	0.000399	0.000321
HSBC	0.000306	0.000474	0.000596	0.000738
Google	0.000387	0.000434	0.000861	0.000892
Amazon	0.000928	0.000371	0.000682	0.000941
Johnson & Johnson	0.000198	0.000118	0.000041	0.000102
JPMorgan	0.000269	0.000199	0.000321	0.000397
ExxonMobil	0.003049	0.000226	0.004583	0.000168
Walmart	0.000167	0.000233	0.000046	0.000074
Disney	0.000387	0.000284	0.000624	0.000622
Visa	0.000213	0.000124	0.000181	0.000088
Netflix	0.000470	0.000379	0.002637	0.002288
Nvidia	0.001142	0.001077	0.002102	0.002298
Procter & Gamble	0.000108	0.000307	0.000075	0.000428
Mastercard	0.000232	0.000733	0.000216	0.000548
Coca-Cola	0.000092	0.000346	0.000065	0.000360
Intel	0.001126	0.006126	0.002120	0.015363

Table 3. Comparison of normal and anomalous day forecasting DirAcc for multiple stocks.

Stocks	Baseline Norm. DirAcc	AAT Norm. DirAcc	Baseline Anom. DirAcc	AAT Anom. DirAcc
Apple	0.5619	0.5196	0.5000	0.3333
Microsoft	0.4448	0.5612	0.0000	1.0000
Tesla	0.5227	0.4864	0.6667	0.3333
Alibaba	0.4939	0.5091	0.2857	0.4286
Pepsi	0.5333	0.4939	0.4286	0.4286
HSBC	0.4230	0.4169	0.6667	0.5000
Google	0.4164	0.4681	0.6250	0.3750
Amazon	0.4428	0.4819	0.6000	0.2000
Johnson & Johnson	0.4955	0.5552	0.5000	0.0000
JPMorgan	0.4727	0.4939	0.7143	0.4286
ExxonMobil	0.5015	0.5075	0.5000	0.5000
Walmart	0.4713	0.5076	0.6667	0.6667
Disney	0.4743	0.4562	0.5000	0.5000
Visa	0.4726	0.5427	0.5556	0.4444
Netflix	0.4985	0.4924	0.3333	0.3333
Nvidia	0.4341	0.4341	0.3333	0.3333
Procter & Gamble	0.5045	0.4653	0.6667	0.3333
Mastercard	0.5196	0.5317	0.5000	0.6667
Coca-Cola	0.5075	0.4939	0.5000	0.2500
Intel	0.4896	0.5014	0.4666	0.4906

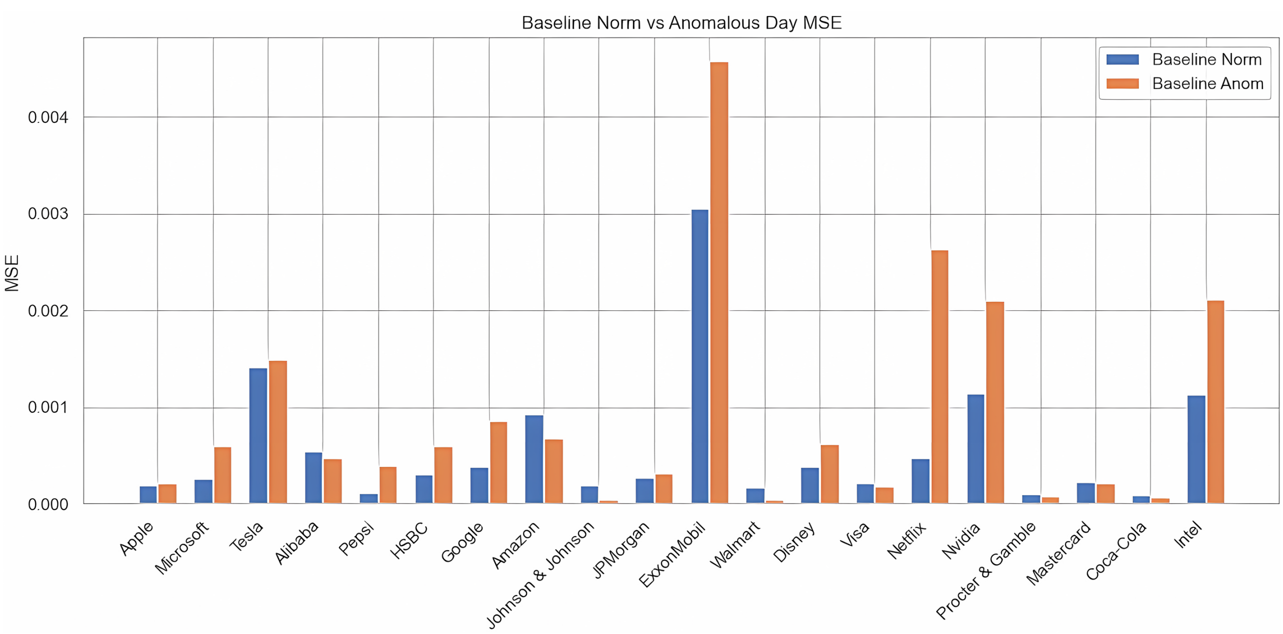

Figure 1. Baseline norm vs anomalous day MSE.

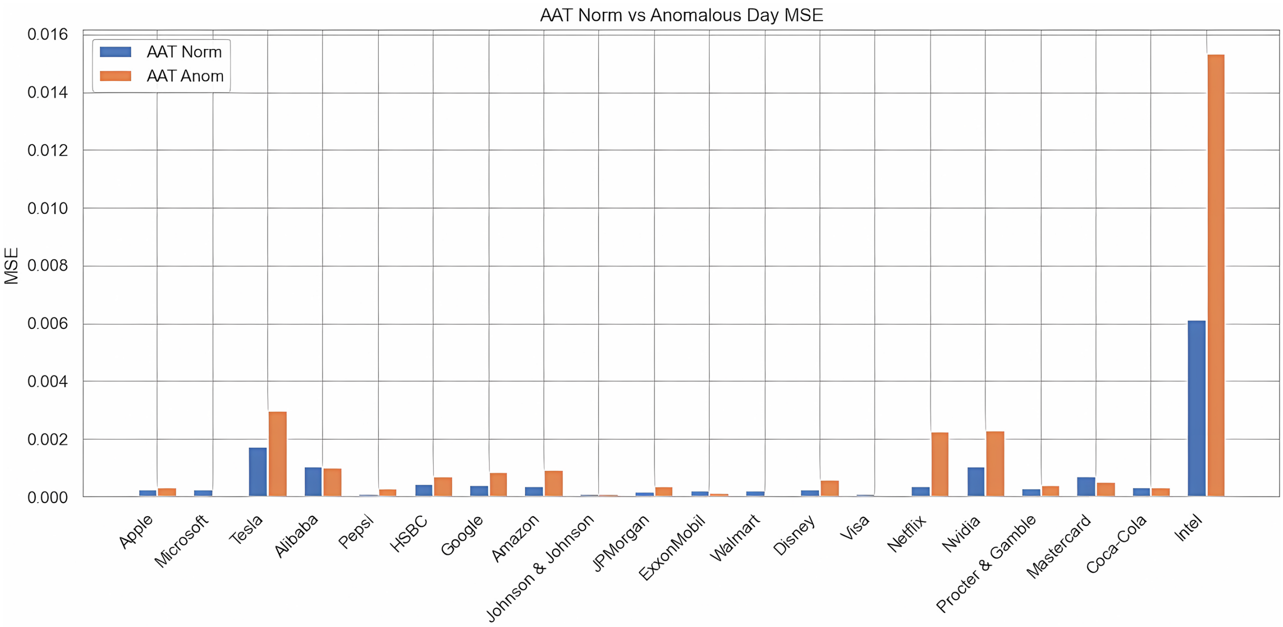

Figure 2. AAT norm vs anomalous day MSE.

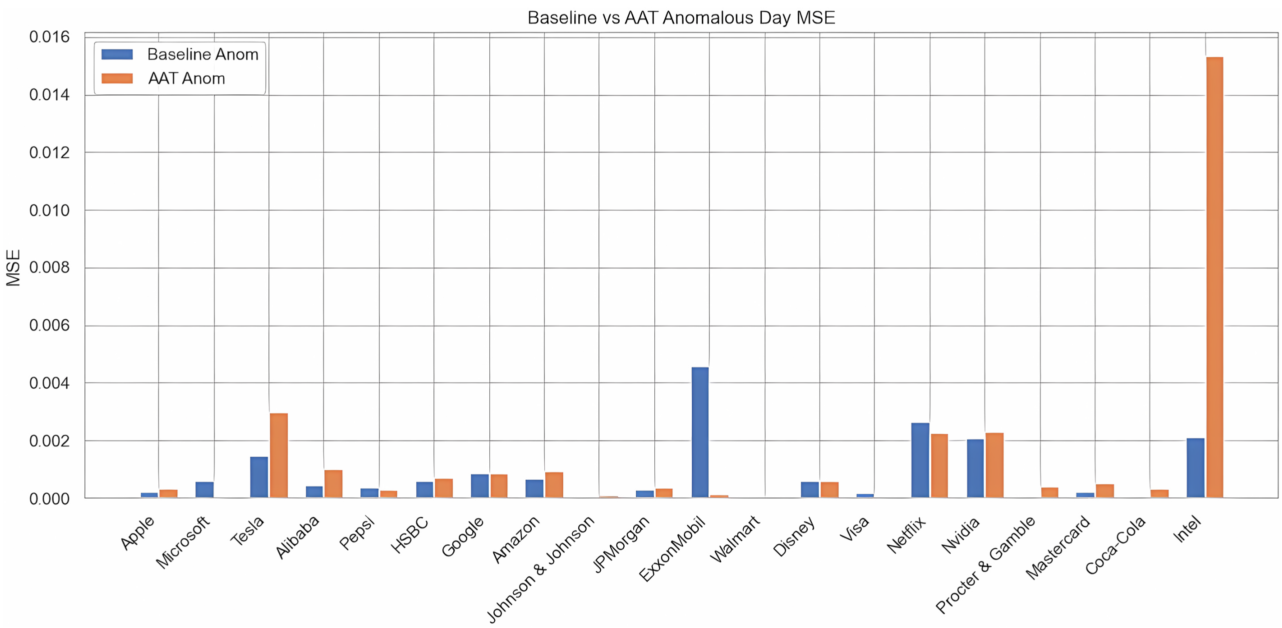

Figure 3. Baseline vs AAT anomalous day MSE.

The comparison of error metrics across the 20 stocks reveals several clear patterns regarding how the Baseline Transformer and the Anomaly-Aware Transformer (AAT) behave under normal and anomalous market conditions. For most stocks, the baseline model achieves lower MSE on both standard trading days and anomalous days, indicating that it captures market dynamics more consistently. The AAT, which incorporates anomaly-related features into its learning process, often shows higher overall MSE for many stocks, reflecting its greater sensitivity to irregular fluctuations that may not occur frequently enough to guide stable parameter updates.

However, when attention is restricted to anomalous days, the behavior of the two models diverges more clearly. For a group of stocks—such as Microsoft, Pepsi, Google, Walmart, Disney, Visa, and especially ExxonMobil—the AAT achieves lower MSE during anomalous periods. This pattern suggests that the model benefits when anomalies reflect structural or repeatable signals, such as volume surges tied to earnings, sector-wide shocks, or meaningful technical breaks. These are "learnable" anomalies in the sense that they have some consistent relation to past market behavior.

In contrast, for highly volatile stocks like Tesla, Alibaba, Nvidia, and Intel, the AAT produces much higher errors on anomalous days. These cases often involve sharp price reversals, liquidity-driven spikes, or news shocks that do not resemble past patterns. Such movements are closer to "erratic" swings—large but unpredictable jumps that provide little stable structure for the model to learn from. For a few stable consumer goods stocks, such as Pepsi, the two models perform similarly, reinforcing the idea that anomaly-aware modeling is most beneficial when the unusual movements have some degree of regularity rather than being purely noise-driven.

Furthermore, Table 3 shows that both Baseline and AAT models achieve moderate accuracy on normal days, with AAT slightly improving or slightly underperforming depending on the stock. On anomalous days, accuracy varies widely: the Baseline model often drops, while AAT sometimes improves for stocks such as Intel and Mastercard. Overall, directional accuracy is generally higher for normal days, highlighting the difficulty of predicting extreme market movements, and AAT captures complex patterns in some high-volatility cases but not consistently across all stocks.

Overall, the pattern suggests that the baseline model remains strong and reliable in normal regimes, while the AAT offers advantages in certain anomalous environments, particularly for stocks whose anomalies follow identifiable or learnable patterns. At the same time, the AAT may struggle when anomalies are too abrupt or unpredictable. This highlights a trade-off: anomaly-aware modeling enhances robustness for some assets during turbulent conditions but does not universally surpass baseline performance, pointing toward the potential value of adaptive or hybrid methods that can switch between modeling strategies depending on market state.

Conclusions

This study examined whether incorporating anomaly-aware signals can enhance stock price forecasting, particularly during periods of irregular market behavior. By integrating Isolation Forest and rolling Z-score-based anomaly indicators into a Transformer architecture, the proposed Anomaly-Aware Transformer (AAT) explicitly distinguishes between normal and anomalous trading regimes. The empirical evaluation across twenty large-cap stocks shows a clear asymmetry in model behavior: the standard Transformer remains strong on normal days where price dynamics follow stable patterns, whereas the AAT exhibits noticeable advantages for several stocks during anomalous intervals characterized by sudden volatility shifts or structured irregular movements. These results indicate that anomaly-aware inputs can help a model adapt its attention to extreme conditions, improving robustness when markets deviate from typical behavior. At the same time, the gains are not uniform across all stocks, suggesting that the usefulness of anomaly conditioning depends on each asset's volatility structure and the detectability of its abnormal patterns.

Overall, this work demonstrates the promise of integrating unsupervised anomaly detection with deep sequence models for financial forecasting and encourages future research on adaptive or hybrid methods. For instance, a state-dependent model could use a gating mechanism to switch between the baseline Transformer and the AAT framework depending on the current market regime. Other possibilities include dynamic thresholds or hybrid anomaly-handling strategies that adjust the model's attention or prediction behavior in response to unusual market conditions.

This study has several limitations. First, the current study relies on an encoder-only Transformer, which may restrict the model's ability to capture longer-term dependencies or complex temporal patterns that alternative architectures, such as encoder-decoder Transformers or temporal convolutional networks, could better exploit. Second, the anomaly detection methods, while effective, are static and may not fully adapt to evolving market dynamics, limiting performance in periods of unprecedented volatility. Third, the model treats all detected anomalies similarly, without differentiating between "learnable" structural anomalies and highly erratic or noise-driven price movements, which may reduce forecasting accuracy for highly volatile stocks. Finally, the approach assumes that past patterns provide reliable signals for future anomalies, which may not always hold in rapidly changing markets. Addressing these limitations through more flexible architectures, adaptive anomaly-handling, or regime-aware designs represents a promising direction for future research.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

[1] Chen, Q. and Kawashima, H. (2024) Stock Price Prediction Using LLM-Based Sentiment Analysis. 2024 *IEEE International Conference on Big Data* (*BigData*), Washington DC, 15-18 December 2024, 4846-4853. https://doi.org/10.1109/bigdata62323.2024.10825946

[2] Chen, Q. (2025) A Two-Stage Framework for Stock Price Prediction: LLM-Based Forecasting with Risk-Aware PPO Adjustment. *Journal of Computer and Communications*, **13**, 120-139. https://doi.org/10.4236/jcc.2025.134008

[3] Mutinda, J.K. and Langat, A.K. (2024) Stock Price Prediction Using Combined

GARCH-AI Models. *Scientific African*, **26**, e02374.
https://doi.org/10.1016/j.sciaf.2024.e02374

[4] Xia, P., Feng, J. and Tian, T. (2025) Prediction of Financial Time-Series Data Using ARIMA Model and Machine Learning Algorithm. *Proceedings of the* 2*nd Guangdong-Hong Kong-Macao Greater Bay Area International Conference on Digital Economy and Artificial Intelligence*, Dongguan, 28-30 March 2025, 1619-1622. https://doi.org/10.1145/3745238.3745492

[5] Chen, Q. (2025) Stock Price Prediction with LLM-Guided Market Movement Signals and Transformer Model. *FinTech and Sustainable Innovation*, **1**, 1-7. https://doi.org/10.47852/bonviewfsi52025703

[6] Chavhan, S., Raj, P., Raj, P., Dutta, A.K. and Rodrigues, J.J.P.C. (2024) Deep Learning Approaches for Stock Price Prediction: A Comparative Study of LSTM, RNN, and GRU Models. 2024 9*th International Conference on Smart and Sustainable Technologies* (*SpliTech*), Bol and Split, 25-28 June 2024, 1-6. https://doi.org/10.23919/splitech61897.2024.10612666

[7] Chen, Q. (2025) Comparing Vision-Instruct LLMs, Vision-Based Deep Learning, and Numeric Models for Stock Movement Prediction. *International Journal of Advanced Computer Science and Applications*, **16**, 11-18. https://doi.org/10.14569/ijacsa.2025.0160402

[8] Sayavong, L., Wu, Z. and Chalita, S. (2019) Research on Stock Price Prediction Method Based on Convolutional Neural Network. 2019 *International Conference on Virtual Reality and Intelligent Systems* (*ICVRIS*), Jishou, 14-15 September 2019, 173-176. https://doi.org/10.1109/icvris.2019.00050

[9] Chen, Q. (2025) Image-Driven Stock Price Prediction with Llama: A Prompt-Based Approach. *International Journal of Modeling and Optimization*, **15**, 17-24. https://doi.org/10.7763/ijmo.2025.v15.867

[10] Chen, Q. and Kawashima, H. (2025) A Novel Sentiment Correlation Based Method with Dual Transformer Model for Stock Price Prediction. *International Journal of Data Science and Analytics*, **21**, Article No. 39. https://doi.org/10.1007/s41060-025-00932-7

[11] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L. and Polosukhin, I. (2023) Attention Is All You Need. *Proceedings of the* 31*st International Conference on Neural Information Processing Systems*, Long Beach, 4-9 December 2017, 6000-6010. https://arxiv.org/abs/1706.03762

[12] Chen, Q. (2025) A Framework for Measuring How News Topics Drive Stock Movement. https://arxiv.org/abs/2510.06864

[13] Liu, F.T., Ting, K.M. and Zhou, Z. (2008) Isolation Forest. 2008 8*th IEEE International Conference on Data Mining*, Pisa, 15-19 December 2008, 413-422. https://doi.org/10.1109/icdm.2008.17

[14] Sharma, R., Sharma, R. and Hariharan, S. (2024) Stock Price Prediction Using ARIMA: A Study on AAPL, MSFT, NFLX, and GOOGL Stocks. 2024 *International Conference on Computational Intelligence and Computing Applications* (*ICCICA*), Vol. 1, 47-50. https://doi.org/10.1109/iccica60014.2024.10584587

[15] Almaafi, A., Bajaba, S. and Alnori, F. (2023) Stock Price Prediction Using ARIMA versus Xgboost Models: The Case of the Largest Telecommunication Company in the Middle East. *International Journal of Information Technology*, **15**, 1813-1818. https://doi.org/10.1007/s41870-023-01260-4

[16] Mohammed, S.A.S.A. (2025) Support Vector Machines in Stock Price Prediction. In: *Financial Innovation for Global Sustainability*, John Wiley & Sons, Ltd., 561-578.

[17] Du, S., Hao, D. and Li, X. (2022) Research on Stock Forecasting Based on Random Forest. 2022 *IEEE 2nd International Conference on Data Science and Computer Application* (*ICDSCA*), Dalian, 28-30 October 2022, 301-305. https://doi.org/10.1109/icdsca56264.2022.9987903

[18] Turchenko, V., Beraldi, P., De Simone, F. and Grandinetti, L. (2011) Short-Term Stock Price Prediction Using MLP in Moving Simulation Mode. *Proceedings of the 6th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems*, Vol. 2, 666-671. https://doi.org/10.1109/idaacs.2011.6072853

[19] Istiake Sunny, M.A., Maswood, M.M.S. and Alharbi, A.G. (2020) Deep Learning-Based Stock Price Prediction Using LSTM and Bi-Directional LSTM Model. 2020 *2nd Novel Intelligent and Leading Emerging Sciences Conference* (*NILES*), Giza, 24-26 October 2020, 87-92. https://doi.org/10.1109/niles50944.2020.9257950

[20] Rahmadeyan, A. and Mustakim (2024) Long Short-Term Memory and Gated Recurrent Unit for Stock Price Prediction. *Procedia Computer Science*, **234**, 204-212. https://doi.org/10.1016/j.procs.2024.02.167

[21] Wang, J.-W. and Chen, J.-S. (2025) An Integrated Fuzzy Convolutional Neural Network Model for Stock Price Prediction. *Engineering Proceedings*, **98**, Article No. 3.

[22] Chen, Q. (2025) Comparing Different Transformer Model Structures for Stock Prediction. https://arxiv.org/abs/2504.16361

[23] Chen, Q. and Kawashima, H. (2025) Sentiment-Aware Stock Price Prediction with Transformer and LLM-Generated Formulaic Alpha. https://arxiv.org/abs/2508.04975

[24] Ji, Y., Luo, Y., Lu, A., Xia, D., Yang, L. and Wee-Chung Liew, A. (2024) Galformer: A Transformer with Generative Decoding and a Hybrid Loss Function for Multi-Step Stock Market Index Prediction. *Scientific Reports*, **14**, Article No. 23762. https://doi.org/10.1038/s41598-024-72045-3

[25] Hu, X. (2021) Stock Price Prediction Based on Temporal Fusion Transformer. 2021 *3rd International Conference on Machine Learning, Big Data and Business Intelligence* (*MLBDBI*), Taiyuan, 3-5 December 2021, 60-66. https://doi.org/10.1109/mlbdbi54094.2021.00019

[26] Olorunnimbe, K. and Viktor, H. (2024) Ensemble of Temporal Transformers for Financial Time Series. *Journal of Intelligent Information Systems*, **62**, 1087-1111. https://doi.org/10.1007/s10844-024-00851-2

[27] Crépey, S., Lehdili, N., Madhar, N. and Thomas, M. (2022) Anomaly Detection in Financial Time Series by Principal Component Analysis and Neural Networks. *Algorithms*, **15**, Article No. 385. https://doi.org/10.3390/a15100385

[28] Zhang, J., Hong, Z., Yan, X. and Wu, S. (2024) Anomaly Detection in Financial Time Series Based on Recurrent Neural Networks and Latent Variable Autoregressive Models. *Proceedings of the International Conference on Digital Economy, Blockchain and Artificial Intelligence*, Guangzhou, 23-25 August 2024, 391-395. https://doi.org/10.1145/3700058.3700119

[29] Cao, F. and Guo, X. (2024) Automated Financial Time Series Anomaly Detection via Curiosity-Guided Exploration and Self-Imitation Learning. *Engineering Applications of Artificial Intelligence*, **135**, Article ID: 108663. https://doi.org/10.1016/j.engappai.2024.108663