Journal of Computer Science and Software Applications
https://www.mfacademia.org/index.php/jcssa
ISSN:2377–0430
Vol. 6, No. 4, 2026

Transformer-Based Transfer Learning Method for Financial Time-Series Forecasting

Lysander Booth[1], Adrian Veldhoorn[1]
[1]University of Manitoba, Winnipeg, Canada
*Corresponding Author: Lysander Booth; ly.booth73@gmail.com

Abstract:

Financial market forecasting is widely regarded as one of the most challenging tasks in data mining. Recently, Transformer models have achieved remarkable success in improving the accuracy of Financial Time-Series Forecasting (FTSF). However, due to the implicit complexity of financial information and the limited availability of labeled data, existing approaches often exhibit insufficient generalization capability in this domain. To alleviate the overfitting problem caused by data scarcity, this paper proposes a Transformer-based deep transfer learning framework integrated with domain-adversarial adaptation, termed TADA-FTSF. The proposed framework is designed for financial time-series forecasting tasks and aims to enhance the reliability and predictive accuracy of deep forecasting models in the financial domain.

Keywords:

Financial time-series forecasting, Transformer model, domain adaptation, transfer learning

1. Introduction

The objective of time-series forecasting is to develop a model that predicts future variations based on the historical sequence of variables $X_i = \{x_1, x_2, \ldots, x_t\}$, where X_i denotes the i-th variable and t represents the length of the observation period. Financial time-series forecasting is a highly relevant research area, playing a crucial role in risk management, decision-making, and the realization of financial objectives.

Financial time-series data exhibit strong nonlinearity and are typically large in scale. Therefore, traditional time-series forecasting methods, such as Autoregression (AR) [1], Autoregressive Integrated Moving Average (ARIMA) [2], and Exponential Smoothing (ETS) [3], are constrained by their linear modeling assumptions and are insufficient for complex financial forecasting tasks.

In recent years, deep learning has demonstrated highly competitive performance across multiple computational fields [4]. Various deep learning models, including Long Short-Term Memory (LSTM) networks [5], Gated Recurrent Units (GRU) [6], and Temporal Convolutional Networks (TCN) [7], have been applied to time-series forecasting tasks [8-9]. Compared with traditional methods, deep learning models possess stronger nonlinear representation capabilities, making them more suitable for large-scale financial time-series forecasting problems.

Compared with conventional time-series data, financial time-series are sampled at fixed intervals and exhibit long-term continuity. Meanwhile, their numerical fluctuations are influenced by multiple factors, which necessitates modeling approaches that simultaneously consider long-term dependency and robustness [10]. To address this issue, this paper proposes a transfer learning-based financial time-series forecasting method built upon the RSA-Transformer framework, referred to as TADA-FTSF.

For long-term dependency modeling, the RSA-Transformer effectively mitigates the computational and parameter complexity associated with long-sequence modeling and utilizes residual structures to ensure stable gradient propagation during training. Regarding robustness, TADA-FTSF leverages source-domain datasets to guide learning in the target domain, thereby completing financial time-series forecasting tasks. On the one hand, this strategy alleviates the negative impact of limited data on model training. On the other hand, cross-domain learning reduces the influence of noise interference, thus improving the robustness of forecasting results.

Extensive experiments are conducted on the proposed TADA-FTSF architecture. Experimental results demonstrate the effectiveness and robustness of the proposed framework.

2. Related Work

Financial time-series forecasting (FTSF) has been extensively studied as a fundamental problem in data-driven financial analysis, where deep learning models have demonstrated strong capabilities in modeling nonlinear and non-stationary temporal dynamics. Early approaches primarily relied on recurrent neural networks such as Long Short-Term Memory (LSTM) and related variants, which achieved promising performance in stock market prediction and financial risk modeling due to their ability to capture sequential dependencies [11], [12]. However, these models often exhibit limitations in modeling long-term dependencies and suffer from performance degradation when handling large-scale or highly volatile financial data. To address these challenges, attention-based architectures, particularly Transformer models, have emerged as a dominant paradigm for time-series forecasting. Representative works introduce interpretable attention mechanisms for multi-horizon prediction [13], while efficient Transformer variants significantly reduce computational complexity for long sequence modeling through sparse attention mechanisms [14]. Further improvements have been achieved through decomposition-based architectures and auto-correlation mechanisms, enabling more effective modeling of long-term temporal structures [15], while subsequent studies demonstrate that Transformer-based representations can capture long-range dependencies more effectively than conventional sequence models [16].

In financial applications, deep learning frameworks have been extended to incorporate domain-specific characteristics and complex relational structures. Hybrid architectures combining recurrent networks and attention mechanisms have been proposed for financial fraud detection and anomaly identification [17], while unified feature embedding frameworks and lightweight attention mechanisms have been developed for enterprise risk prediction tasks [18]. In addition, graph-based modeling and multi-agent systems have been introduced to capture systemic risk propagation and interdependencies among financial entities [19], [20], and causal reasoning frameworks have been applied to enhance interpretability and decision-making in enterprise financial analysis [21]. Despite these advances, most existing approaches primarily focus on improving model architectures and representation capabilities, while insufficient attention has been given to the challenges posed by limited labeled data and distribution shifts across different financial domains.

Transfer learning has been widely recognized as an effective approach to address data scarcity and improve generalization performance in machine learning tasks. The foundational framework of transfer learning provides a systematic paradigm for leveraging knowledge across domains [22], while subsequent advances in domain adaptation focus on aligning feature distributions between source and target domains. Adversarial domain adaptation methods introduce gradient reversal mechanisms to learn domain-invariant representations [23], whereas discrepancy-based approaches explicitly minimize distribution divergence through statistical measures such as Maximum Mean Discrepancy (MMD) [24]. Further developments incorporate adversarial discriminative strategies to improve transferability in complex environments [25]. In financial contexts, recent studies have explored domain adaptation and meta-learning techniques for credit fraud detection and risk prediction under dynamic environments [26], [27], demonstrating improved robustness under distribution shift and class imbalance. Nevertheless, most existing transfer learning approaches are designed for classification or anomaly detection tasks, and their application to financial time-series forecasting remains relatively underexplored.

Meanwhile, recent advances in artificial intelligence have further extended financial modeling toward more intelligent and integrated frameworks. Large language models have been introduced for financial anomaly detection and auditing through semantic reasoning and multi-document analysis [28], while graph neural networks and relational learning techniques have been widely adopted for modeling credit risk and fraud detection through complex entity interactions [29]. Causal representation learning and structural modeling approaches have also been proposed to improve interpretability and robustness in financial systems [30], [31], and anomaly ranking frameworks based on structural deviations have been developed for enterprise financial monitoring [32]. Additionally, federated learning methods have been applied to enable privacy-preserving collaborative risk detection across distributed financial environments [33]. Despite these advancements, existing methods often treat forecasting, anomaly detection, and risk modeling as separate tasks, lacking a unified framework that simultaneously addresses temporal dependency modeling, cross-domain knowledge transfer, and robustness under distribution shift.

In summary, although substantial progress has been made in Transformer-based time-series modeling, transfer learning, and AI-driven financial analysis, several critical challenges remain, including data scarcity, domain shift, and the lack of unified modeling frameworks. To address these limitations, this paper proposes a Transformer-based transfer learning framework with domain-adversarial adaptation, aiming to enhance both predictive accuracy and cross-domain generalization in financial time-series forecasting tasks.

3. TADA-FTSF Framework

In this section, we introduce the proposed framework in detail. The overall training procedure is illustrated in Figure 1. We first present a novel perspective for selecting source datasets. Then, we describe the components of the TADA-FTSF framework and its training loss functions.

3.1 Time-Series Causal Discovery for Potential Source Domain Selection

In transfer learning tasks, the selection of source datasets greatly influences the effectiveness of knowledge transfer. Traditional approaches typically rely on time-consuming manual selection or training multiple models on large datasets to determine appropriate sources. To address this issue, this section proposes a novel time-series causal discovery method, providing a new perspective for source domain selection in financial time-series forecasting (FTSF) transfer learning tasks.

Causal relationships are ubiquitous in nature, and their discovery is a central topic in interdisciplinary research. Time-series causal discovery aims to identify causal relationships between temporal sequences and is a classical problem in statistical time-series analysis with significant application value across various domains. Transfer Entropy (TE) is one of the commonly used measures for quantifying causal relationships. It measures the amount of information transferred from one variable to another and has the advantages of being model-free and distribution-free. The TE between two sequences is defined as:

$$TE = I(Y_{i+1}; X_i \mid Y^i)$$

ISSN:2377-0430
Vol. 6, No. 4, 2026

where X_i, Y_i, $i = 1,2,\ldots,T$, and $Y^i = (Y^1,\ldots,Y^T)$ denote two different time-series.

Ma proposed the concept of Copula entropy and proved its equivalence to mutual information in information theory. Let x be a random variable with marginal distribution u and Copula density $c(u)$. The Copula entropy of x is defined as:

$$H_c(x) = -\int_u c(u)\log c(u)\,du$$

Copula entropy measures independence, whereas TE measures conditional independence. Reference introduced the fundamental theoretical relationship between them. Transfer entropy can be expressed in terms of Copula entropy as follows:

$$TE(Y,X) = -H_c(Y_{i+1},Y_i,X_i) + H_c(Y_{i+1},Y_i) + H_c(X_i,Y_i)$$

In the proposed method, according to (3), the average TE from each variable in the source dataset to the target variable in the target dataset is computed using Copula entropy. The source dataset with the maximum TE value is selected as the source domain.

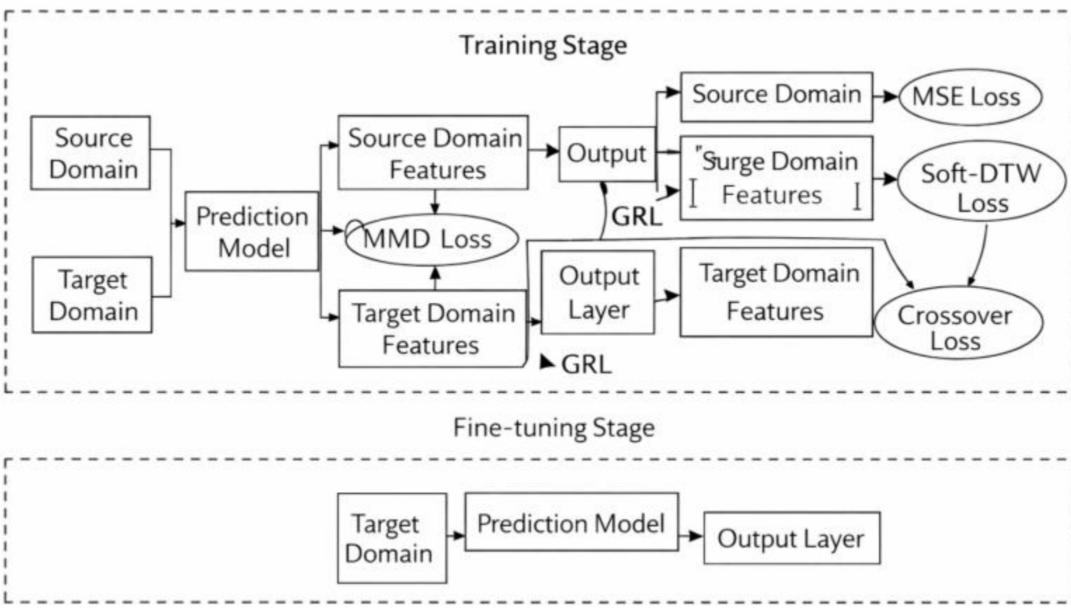

Figure 1. Overall framework of TADA-FTSF

3.2 Component Analysis of TADA-FTSF

1) *RSA-Transformer Model*

Transformer models have been widely adopted for long-sequence time-series forecasting due to their capability in modeling long-range dependencies and complex interactions. However, the standard self-attention mechanism suffers from quadratic computational complexity and high memory consumption. In addition, redundant or irrelevant information may deteriorate forecasting performance. Single-scale modeling further limits the ability to capture temporal patterns across multiple time horizons.

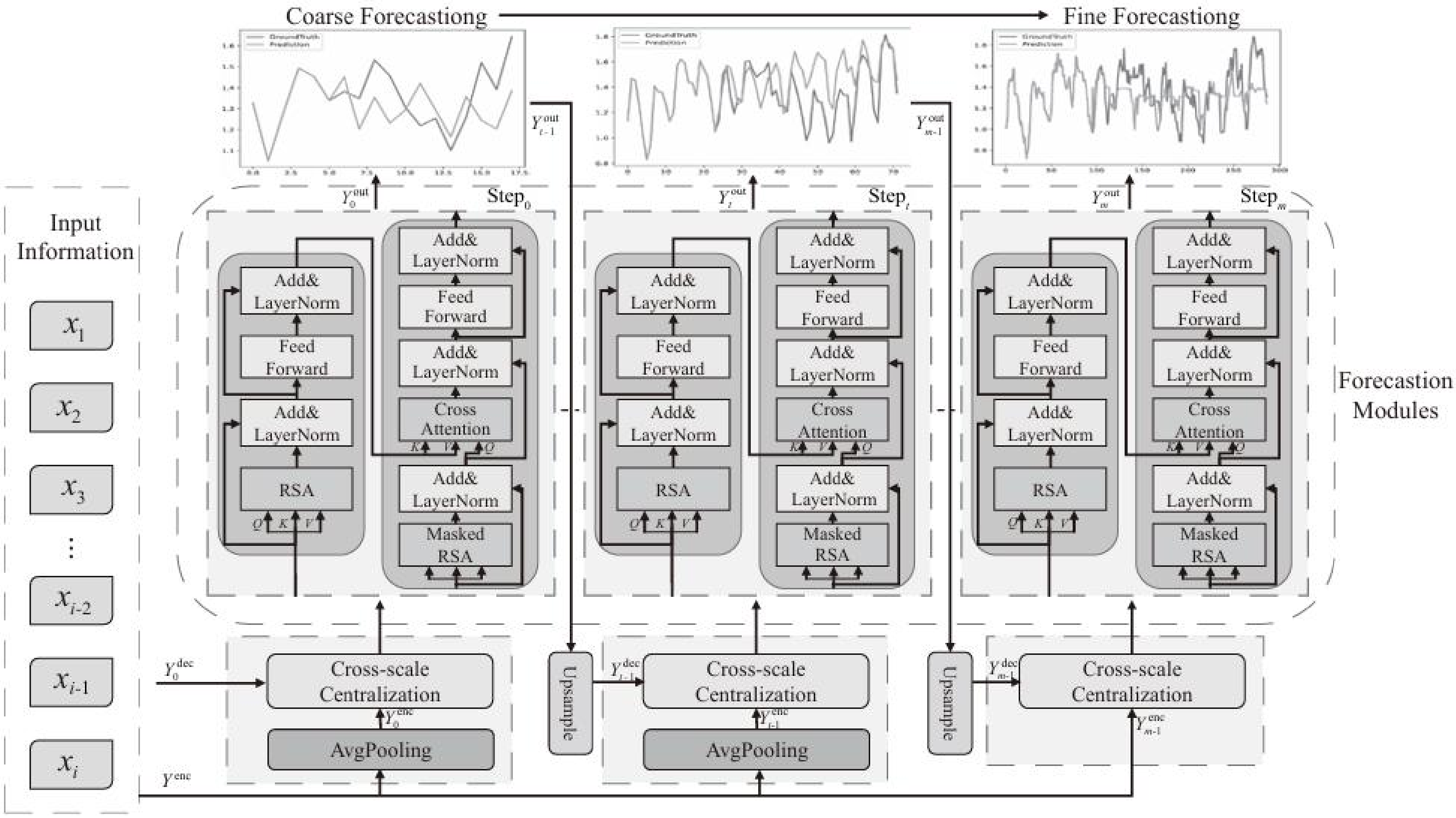

Figure 2. Network architecture of RSA-Transformer

To address these challenges, an RSA-Transformer is introduced to construct an efficient multi-scale framework for financial time-series forecasting (FTSF). The overall architecture follows an encoder-decoder paradigm, where the embedded input sequence is processed through stacked RSA blocks to reduce computational complexity while preserving long-term dependency modeling.

In the standard Transformer, attention is computed based on query Q, key K, and value V matrices. The scaled dot-product attention is defined as:

$$\mathrm{Atten}(Q,K,V)=\mathrm{Softmax}\left(\frac{QK^T}{\sqrt{d}}\right)V$$

where d denotes the feature dimension. Although effective, this formulation incurs $O(L^2)$ complexity.

To reduce redundancy, a sparsity criterion is introduced following [21]. Instead of computing full attention, a sparsity score is constructed to measure the importance of each query-key pair. For numerical stability and efficiency, the sparsity indicator is approximated as:

$$\tilde{\Gamma}(Q_i,K)=\max_j\left(\frac{Q_iK_j^T}{\sqrt{d}}\right)-\frac{1}{L_K}\sum_{j=1}^{L_K}\frac{Q_iK_j^T}{\sqrt{d}}$$

Based on this approximation, sparse attention can be expressed as:

$$S_{\text{sparse}}\text{Atten}(Q,K,V) = \text{Softmax}\left(\frac{\tilde{Q}K^T}{\sqrt{d}}\right)V$$

where $\widetilde{Q}$ denotes the selected informative queries. A residual accumulation strategy is further introduced to stabilize gradient propagation and enhance long-term modeling capability:

$$A_n = \frac{\tilde{Q}_n K_n^T}{\sqrt{d}} + A_{n-1}$$

Through sparse attention and residual aggregation, RSA-Transformer effectively reduces computational complexity while maintaining predictive performance for long financial sequences.

2) Hybrid Loss Function Construction

Within the TADA-FTSF framework, a hybrid loss function is designed for the pretraining stage to promote effective alignment between the source and target domains. The objective function consists of four complementary components: supervised regression loss, distribution alignment loss, adversarial domain loss, and shape-aware alignment loss.

First, the supervised term corresponds to the mean squared error (MSE) computed on labeled source-domain samples. This term ensures that the RSA-Transformer learns accurate predictive mappings from input sequences to target values under fully supervised conditions.

Second, to mitigate distribution discrepancy between source and target domains, Maximum Mean Discrepancy (MMD) is adopted as a statistical distance metric. MMD measures the difference between two probability distributions in a reproducing kernel Hilbert space and is widely used in domain adaptation to align feature representations across domains.

Third, an adversarial loss is introduced through a Gradient Reversal Layer (GRL). Under the assumption of domain-invariant feature learning, a domain classifier is trained to distinguish between source and target features, while the feature extractor is optimized adversarially to confuse the classifier. This min-max optimization mechanism encourages the learned representations to be domain-agnostic, thereby improving transferability.

Finally, to better align temporal structures between predicted and true sequences, a soft-DTW loss is incorporate. Unlike classical Dynamic Time Warping (DTW), which is non-differentiable due to the min operator, soft-DTW introduces a smooth soft-min formulation that enables gradient-based optimization. This mechanism enhances structural consistency between sequences and facilitates the transfer of shape-related knowledge.

The overall training objective is formulated as a weighted combination of the above components, whee balancing hyperparameters control the relative contribution of each term. During optimization, model parameters are updated via backpropagation, and AdamW is employed as the optimizer to improve convergence stability and generalization performance.

Journal of Computer Science and Software Applications
https://www.mfacademia.org/index.php/jcssa
ISSN:2377-0430
Vol. 6, No. 4, 2026

4. Experimental Results

4.1 Datasets

The experimental data consist of financial index time-series collected from internationally recognized financial information platforms. The datasets include the Shanghai Composite Index (SHI), Shenzhen Component Index (SZI), Dow Jones Industrial Average (DJI), S&P 500 Index (SP), Hang Seng Index (HSI), and NASDAQ Composite Index (NCI).

The data span from January 5, 2010 to December 31, 2023 and are collected at a daily frequency. Due to differences in trading calendars, the data scales are not fully aligned. To avoid information leakage in transfer learning, future information is strictly excluded. All time-series are processed through data expansion and linear interpolation to achieve temporal alignment. Each target dataset is divided chronologically into 80% training data and 20% testing data. Four variables, including opening price and highest price, are selected and transformed into inputs suitable for supervised learning. A Vector Autoregression (VAR) model is employed to determine the lag order based on the BIC criterion before conversion to model inputs.

4.2 Experimental Settings

To validate the effectiveness of the proposed transfer learning framework and the RSA-Transformer model, several competitive baseline methods are selected for comparison, including vanilla LSTM, TCN, GRU, Support Vector Regression (SVR), and Random Forest (RF).

To further evaluate the generality of the proposed transfer strategy, the framework is also applied to TCN, LSTM, and GRU, forming ADA-TCN, ADA-LSTM, and ADA-GRU, respectively. In these configurations, ADA-LSTM and ADA-GRU stack two recurrent layers as feature extractors, while ADA-TCN stacks four causal convolutional layers. Each model includes a fully connected layer as the prediction module.

To ensure fair comparison, classical transfer learning strategies are also implemented. Specifically, models are first pretrained on the source domain and then fine-tuned on the target domain. These variants are denoted as TL-LSTM, TL-TCN, and TL-GRU.

Based on extensive experimental results, the hyperparameters of ADA-LSTM, ADA-GRU, and ADA-TCN are set as $k_1 = k_2 = 1$ and $k_3 = 0.1$. The hidden layer sizes are selected from {16, 32, 64} according to prediction performance. During training, the learning rate is adaptively adjusted based on validation loss. The initial learning rate is set to 0.1, and the minimum learning rate is set to 1e-6. If the training loss does not decrease for five consecutive epochs, the learning rate is reduced to three quarters of its current value. The maximum number of training iterations is set to 10,000. For traditional machine learning models such as SVR and RF, hyperparameters are optimized using random search.

All datasets are normalized using min-max normalization to eliminate scale differences among variables. As evaluation metrics, Root Mean Squared Error (RMSE) and Symmetric Mean Absolute Percentage Error (sMAPE) are adopted to measure predictive accuracy.

4.3 Comparative Experiments

In this section, suitable source domains are first selected using transfer entropy-based causal analysis. Each dataset is treated as the target domain in turn, while the remaining datasets serve as candidate source domains. According to the computed transfer entropy values, the dataset with the largest value is selected as the source domain for each target. The selected source domains correspond to NCI, SHI, SHI, DJI, DJI, and SHI for the respective targets.

Table 1: Transfer Entropy Between Potential Source and Target Domains

Source \ Target	DJI	HSI	NCI	SHI	SZI	SP
DJI	-	0.1925	0.3037	0.265	0.2488	0.2045
HSI	0.3013	-	0.2239	0.3386	0.3131	0.3
NCI	0.2444	0.1816	-	0.2614	0.2547	0.2136
SHI	0.4127	0.315	0.3569	-	0.25	0.4078
SZI	0.4069	0.3589	0.3419	0.2798	-	0.3883
SP	0.2047	0.1919	0.2349	0.2749	0.2738	-

The RMSE and sMAPE values of the financial index datasets are summarized in Tables 2 and 3. The lowest RMSE and sMAPE values are highlighted in bold, and underlined values indicate improvements achieved by the proposed framework over the corresponding baseline models.

Table 2: RMSE results of different models

Model	DJI (NCI)	HSI (SHI)	NCI (SHI)	SHI (DJI)	SZI (DJI)	SP (SHI)
LSTM	1370.838	343.5236	1928.231	51.0194	244.224	282.81
TL-LSTM	815.0399	338.228	269.1494	33.6984	183.3301	150.0077
GRU	1655.494	408.2342	1200.668	43.8331	199.9866	565.0848
TL-GRU	1573.448	332.4222	406.7413	34.7303	181.7557	62.5489
TCN	395.4291	2733.936	4686.874	539.9791	186.2049	1678.715
TL-TCN	385.9888	2517.005	163.8922	531.3895	189.4721	1616.89
SVR	5740.819	532.6851	5117.484	104.9741	378.6515	756.075
RF	4730.451	341.1819	4269.791	37.6168	240.2685	867.1969
ADA-LSTM	433.8843	329.5171	244.5611	32.7906	180.4096	53.925
ADA-GRU	381.8586	329.811	455.1392	32.5512	174.2237	49.932
ADA-TCN	399.3176	327.1486	2691.162	32.4291	180.7958	58.7484
Proposed	365.155	308.136	157.37	30.9674	169.853	46.3578

Table 3: sMAPE (%) results of different models

Model	DJI (NCI)	HSI (SHI)	NCI (SHI)	SHI (DJI)	SZI (DJI)	SP (SHI)
LSTM	3.6205	0.9938	12.3227	1.1564	1.4387	5.6696
TL-LSTM	2.0615	0.9677	1.7777	0.7617	1.0628	3.3109
GRU	4.4901	1.2052	7.3165	1.0054	1.1633	10.8492
TL-GRU	4.1648	0.9589	2.4836	0.7836	1.0531	1.2794
TCN	0.9529	8.4511	36.9682	14.7183	1.0997	53.1106
TL-TCN	0.9012	7.6639	1.0738	14.3926	1.1206	50.183
SVR	16.9256	1.6135	43.49	2.9722	2.6406	19.6238
RF	11.5521	0.9843	31.4771	0.8618	1.3745	18.4583
ADA-LSTM	1.0798	0.9437	1.4816	0.7338	1.0454	1.0686
ADA-GRU	0.8918	0.945	3.0917	0.7275	1.0086	1.0396
ADA-TCN	0.9521	0.9338	20.9765	0.7278	1.0523	1.3479
Proposed	0.0693	0.7194	1.0856	0.6983	0.9974	0.9632

From the analysis of Tables 2 and 3, it can be observed that compared with ADA-based approaches, the proposed TADA-FTSF achieves the best overall performance. This demonstrates that the RSA-Transformer model exhibits stronger adaptability to financial data characteristics.

Moreover, experimental observations show that ADA-based methods outperform the original models and the fine-tuned TL variants in most cases, verifying the effectiveness of the proposed transfer learning framework. However, the performance of ADA-TCN is relatively unstable, which may be attributed to the structural complexity of TCN and the limited size of the datasets. TCN models are generally more suitable for large-scale data. Additionally, most financial time-series used in this study are non-stationary. The experimental results confirm that the proposed framework maintains robust modeling capability under non-stationary conditions and demonstrates end-to-end learning advantages.

4.4 Ablation Study

In this subsection, ablation experiments are conducted on the TADA-FTSF framework to verify the effectiveness of each component corresponding to the four loss terms. Due to space limitations, the ablation analysis is performed only on the SZI dataset.

The results indicate that incorporating any transfer-based component yields better performance than training the model from scratch. However, the combination of L_1 and L_4 produces a negative transfer effect on SZI, which may be attributed to changes in the value range of the dataset. The hybrid loss function demonstrates competitive performance overall. In particular, the strong coupling ability introduced by the shape-aware

ISSN:2377-0430
Vol. 6, No. 4, 2026

component improves global performance, while negative effects among individual components offset each other, resulting in improved generalization balance. Detailed ablation results for different components are summarized in the corresponding table.

Table 4: ablation results of different components

Component	RMSE	sMAPE
$L_1 + L_2$	175.7263	1.0247
$L_1 + L_3$	176.1868	0.0467
$L_1 + L_4$	184.9452	1.0725
$L_1 + L_2 + L_3$	175.5098	1.0216
$L_1 + L_2 + L_3 + L_4$	169.853	0.9974

4.5 DM Test Results

The Diebold-Mariano (DM) test is employed to compare the predictive performance between two forecasting models. Essentially, it is a statistical significance test based on the difference in forecasting errors. Under the hypothesis testing framework, the null hypothesis states that there is no significant difference between the predictive accuracies of two models, whereas the alternative hypothesis indicates that their predictive performances differ.

Let model A denote the target model and model B denote the baseline model. The null hypothesis assumes that the expected loss values of the two models are equal. The DM statistic is computed based on the time-series of loss differentials and their estimated variance across the testing period.

To conserve space, the DM test is applied to evaluate the effectiveness of the TADA-FTSF framework by comparing TL-LSTM vs ADA-LSTM, TL-GRU vs ADA-GRU, and TL-TCN vs ADA-TCN. The approximate results for three datasets are reported in Table 5. Notably, on the SZI dataset, all three comparisons pass the significance test with nearly 100% confidence, demonstrating the practical effectiveness and reliability of the proposed framework.

Table 5: dm test results

Metric	ADA-LSTM	ADA-GRU	ADA-TCN
DM statistic	-1.46	-3.36	-5.38
p value	0	0	0

5. Conclusion

This paper proposes the TADA-FTSF framework to address the challenge of high-precision forecasting in financial time-series. The main contributions can be summarized in three aspects.

First, to capture the long-term characteristics of financial data, the RSA-Transformer is employed as the feature extractor, enhancing the modeling capability for complex temporal dependencies and improving the fitting performance on financial sequences.

Second, a time-series causal discovery strategy is integrated for source domain selection. By computing transfer entropy to determine the most relevant source domain, the proposed approach satisfies the requirement of avoiding information leakage while providing a simpler and more interpretable alternative to traditional selection methods.

Third, an adversarial domain adaptation strategy is incorporated, and the soft-DTW loss is introduced to embed shape-related knowledge into the learning process. This design improves the generalization ability of the framework and enhances robustness across domains.

Experimental results demonstrate that the proposed framework achieves superior predictive performance and strong generalization capability on multiple financial time-series datasets. Future work may explore source domain selection strategies based on statistical similarity measures to address the problem of incorrect source selection in multi-source transfer scenarios and further refine the overall framework.

References

[1] G. Janacek, "Time series analysis, forecasting and control," J. Time Ser. Anal., 2010.

[2] Y. S. Lee and L. I. Tong, "Forecasting time series using a methodology based on autoregressive integrated moving average and genetic programming," Knowl.-Based Syst., vol. 24, no. 1, pp. 66-72, 2011.

[3] A. Corberán-Vallet, J. D. Bermúdez, and E. Vercher, "Forecasting correlated time series with exponential smoothing models," Int. J. Forecast., vol. 27, no. 2, pp. 252-265, 2011.

[4] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, no. 7553, pp. 436-444, 2015.

[5] T. Fischer and C. Krauss, "Deep learning with long short-term memory networks for financial market predictions," Eur. J. Oper. Res., vol. 270, no. 2, pp. 654-669, 2017.

[6] S. Aryal, D. Nadarajah, P. L. Rupasinghe, et al., "Comparative analysis of deep learning models for multi-step prediction of financial time series," J. Comput. Sci., vol. 16, no. 10, pp. 1401-1416, 2020.

[7] O. B. Sezer, M. Gudelek, and A. M. Ozbayoglu, "Financial time series forecasting with deep learning: A systematic literature review: 2005-2019," Appl. Soft Comput., vol. 90, p. 106181, 2020.

[8] A. W. Li and G. S. Bastos, "Stock market forecasting using deep learning and technical analysis: A systematic review," IEEE Access, vol. 8, pp. 185232-185242, 2022.

[9] T. Niu, J. Z. Wang, H. Y. Lu, et al., "Developing a deep learning framework with two-stage feature selection for multivariate financial time series forecasting," Expert Syst. Appl., vol. 148, p. 113237, 2020.

[10]G. Li, A. Zhang, Q. Zhang, et al., "Pearson correlation coefficient-based performance enhancement of broad learning system for stock price prediction," IEEE Trans. Circuits Syst. II, Exp. Briefs, vol. 69, no. 5, pp. 2413-2417, 2022.

[11]T. Fischer and C. Krauss, "Deep learning with long short-term memory networks for financial market predictions," European Journal of Operational Research, vol. 270, no. 2, pp. 654-669, 2018.

[12]D. M. Q. Nelson, A. C. M. Pereira and R. A. de Oliveira, "Stock market's price movement prediction with LSTM neural networks," 2017 International Joint Conference on Neural Networks (IJCNN), pp. 1419-1426, 2017.

[13]B. Lim, S. Ö. Arık, N. Loeff and T. Pfister, "Temporal fusion transformers for interpretable multi-horizon time series forecasting," International Journal of Forecasting, vol. 37, no. 4, pp. 1748-1764, 2021.

[14]H. Zhou, S. Zhang, J. Peng, S. Zhang, J. Li, H. Xiong and W. Zhang, "Informer: Beyond efficient transformer for long sequence time-series forecasting," Proceedings of the AAAI Conference on Artificial Intelligence, vol. 35, no. 12, pp. 11106-11115, 2021.

[15]H. Wu, J. Xu, J. Wang and M. Long, "Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting," Advances in Neural Information Processing Systems, vol. 34, pp. 22419-22430, 2021.

[16]Y. Nie, N. H. Nguyen, P. Sinthong and J. Kalagnanam, "A time series is worth 64 words: Long-term forecasting with transformers," arXiv preprint arXiv:2211.14730, 2022.

[17]P. Feng, "Hybrid BiLSTM-transformer model for identifying fraudulent transactions in financial systems," Journal of Computer Science and Software Applications, vol. 5, no. 3, 2025.

[18]R. Fang, "A machine learning framework for enterprise risk prediction: Unified feature embedding and lightweight attention," 2026.

[19]C. Chiang, D. Li, R. Ying, Y. Wang, Q. Gan and J. Li, "Deep learning-based dynamic graph framework for robust corporate financial health risk prediction," Proceedings of the 2025 3rd International Conference on Mathematics and Machine Learning, pp. 98-105, 2025.

[20]Y. Wang, "Multi-agent collaborative modeling for systemic risk propagation in financial markets: A game-theoretic framework," 2026.

[21]R. Ying, Q. Liu, Y. Wang and Y. Xiao, "AI-based causal reasoning over knowledge graphs for data-driven and intervention-oriented enterprise performance analysis," 2025.

[22]S. J. Pan and Q. Yang, "A survey on transfer learning," IEEE Transactions on Knowledge and Data Engineering, vol. 22, no. 10, pp. 1345-1359, 2009.

[23]Y. Ganin and V. Lempitsky, "Unsupervised domain adaptation by backpropagation," Proceedings of the International Conference on Machine Learning, pp. 1180-1189, 2015.

[24]M. Long, Y. Cao, J. Wang and M. Jordan, "Learning transferable features with deep adaptation networks," Proceedings of the International Conference on Machine Learning, pp. 97-105, 2015.

[25]E. Tzeng, J. Hoffman, K. Saenko and T. Darrell, "Adversarial discriminative domain adaptation," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7167-7176, 2017.

[26]S. Huang, Y. Zheng, Y. Zhao, R. Ying, K. Cao and X. Liang, "A unified meta learning and domain adaptation framework for credit fraud detection in dynamic environments," 2026.

[27]R. Yan, Y. Ou, S. Sun, N. Chen, K. Zhou and Y. Shu, "DualShiftNet: Joint class-imbalance and distribution-shift aware learning for business risk prediction," 2026.

[28]Q. Gan, "Large language model framework for multi-document financial anomaly detection in intelligent auditing via semantic mapping and risk reasoning," Transactions on Computational and Scientific Methods, vol. 4, no. 12, 2024.

[29]K. Cao, Y. Zhao, H. Chen, X. Liang, Y. Zheng and S. Huang, "Multi-hop relational modeling for credit fraud detection via graph neural networks," 2025.

[30]J. Li, Q. Gan, R. Wu, C. Chen, R. Fang and J. Lai, "Causal representation learning for robust and interpretable audit risk identification in financial systems," 2025.

[31]R. Ying, Q. Liu, Y. Wang and Y. Xiao, "AI-based causal reasoning over knowledge graphs for data-driven and intervention-oriented enterprise performance analysis," 2025.

[32]H. Chen, R. Wu, C. Chen, H. Feng, Y. Nie and Y. Lu, "Anomaly ranking for enterprise finance using latent structural deviations and reconstruction consistency," 2026.

[33]H. Feng, Y. Wang, R. Fang, A. Xie and Y. Wang, "Federated risk discrimination with siamese networks for financial transaction anomaly detection," Proceedings of the 2025 2nd International Conference on Digital Economy and Computer Science, pp. 231-236, 2025.