Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A Hybrid Deep Learning Model Based on INFO-TCN- iTransformer and Its Effectiveness in Stock Price Prediction

Huiyong Wu
Shenyang University of Chemical Technology
Lin Li
135046856430163.com

Shenyang University of Chemical Technology
Rong Wang
Shenyang University of Chemical Technology
Yanli Ding
Shenyang University of Chemical Technology

Research Article

Keywords: Stock price prediction, TCN-Transformer, Time series modeling, Deep graph neural network, Multi-scale features, Self-attention mechanism

Posted Date: February 26th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8803402/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

A Hybrid Deep Learning Model Based on INFO-TCN-iTransformer and Its Effectiveness in Stock Price Prediction

Huiyong Wu, Lin Li *,Rong Wang,Yanli Ding
College of science ,Shenyang University of Chemical Technology, CEP 110142, LN,CHN
*corresponding author: Lin Li 13504685643 @163.com

Abstract: Amid profound financial market transformations and rising uncertainty, accurately forecasting stock price volatility is essential for risk management and the high-quality development of capital markets. Stock price series often exhibit high noise, strong non-stationarity, and multi-scale fluctuation patterns, challenging traditional time series or single deep learning models to capture both local volatility and long-term dependencies. This study proposes an INFO-TCN-iTransformer framework that integrates Temporal Convolutional Networks (TCN) for extracting short- and medium-term local volatility features and an improved Transformer (iTransformer) for modeling long-term dependencies across multivariate time series, enabling effective fusion of local and global information. A Vector-weighted Average Optimization (INFO) algorithm is employed to globally optimize key hyperparameters, improving training stability and predictive accuracy. Empirical results on Chinese stock data demonstrate that the proposed model significantly outperforms baseline methods in prediction accuracy, stability, and generalization. The findings indicate that multi-scale feature learning combined with intelligent hyperparameter optimization provides an effective approach for stock price forecasting and supports quantitative investment and financial risk management.

Keywords: Stock price prediction; TCN-Transformer; Time series modeling; Deep graph neural network; Multi-scale features; Self-attention mechanism

1 Introduction

Stock price prediction is a critical task in financial markets, offering guidance for investors and facilitating risk management strategies. The highly volatile and nonlinear nature of financial time series, however, presents significant challenges for accurate forecasting. Traditional statistical methods, such as autoregressive integrated moving average (ARIMA) models, have been extensively employed for short-term financial forecasting[48], yet their linear assumptions and limited ability to capture long-range temporal dependencies restrict their applicability in complex market environments [27]. To overcome these limitations, recent studies have turned to deep learning approaches, particularly models capable of capturing nonlinear dynamics and temporal dependencies in financial data [8,20,29].

Deep learning architectures, including convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have demonstrated superior performance in time series forecasting due to their automatic feature extraction and sequence modeling capabilities. CNNs can effectively extract local temporal patterns from stock price sequences [10,33,45], while recurrent models such as long short-term memory (LSTM) and gated recurrent units (GRU) can capture sequential dependencies

across multiple time steps [2,37]. Nevertheless, RNN-based models often suffer from gradient vanishing or exploding problems and are computationally intensive for long sequences .To address these challenges, temporal convolutional networks (TCNs) have been introduced, offering exponentially large receptive fields through dilated causal convolutions while maintaining computational efficiency [3,24]. By incorporating residual connections, TCNs ensure stable gradient propagation in deep architectures, enabling robust modeling of long-term temporal dependencies [40].

In parallel, attention mechanisms and Transformer-based architectures have shown remarkable capability in modeling long-range dependencies without the recurrence constraints of RNNs [34,49]. Transformers, initially developed for natural language processing tasks, utilize multi-head self-attention to capture relationships across all time steps, making them highly suitable for financial time series with complex temporal correlations [18,38]. Several studies have explored the integration of attention mechanisms into stock price prediction, demonstrating improved prediction accuracy and adaptability to market volatility [16, 39]. However, purely Transformer-based models can exhibit high computational complexity, especially for high-frequency or multivariate financial data, motivating hybrid approaches that combine convolutional, recurrent, and attention-based structures [6,30].

The hybridization of TCNs and Transformers has emerged as a promising solution. For instance, BiMT-TCN integrates bidirectional temporal convolutional layers with improved Transformer encoders to simultaneously capture local patterns and global dependencies, enhancing prediction stability. Similarly, residual multiscale TCNs combined with Informer models have demonstrated effectiveness in modeling long-term financial sequences while preserving multiscale temporal features [4]. By exploiting the complementary strengths of convolutional feature extraction and attention-based global modeling, these hybrid frameworks address the limitations of single-architecture approaches in stock price forecasting [25].

Moreover, model optimization plays a crucial role in enhancing predictive performance. Methods such as INFO, an optimization strategy leveraging statistical information and variance control, have been applied to tune deep learning parameters effectively [41]. The INFO framework treats model performance as a stochastic variable and guides parameter selection to balance prediction accuracy and stability, offering a principled approach to hyperparameter optimization in complex hybrid models [44]. Studies integrating INFO with stacking or ensemble techniques report improved interpretability and robustness in financial forecasting tasks [19].

Recent advances also emphasize the importance of capturing multivariate and structural dependencies in financial data. Graph neural networks (GNNs) combined with Transformer architectures, such as the TGNS model, have been proposed to exploit inter-stock relationships and temporal dependencies, further improving stock trend prediction [34]. Similarly, models integrating textual sentiment from investor communications with Transformer architectures enable volatility and risk prediction, incorporating both numerical and qualitative market indicators [15,39]. These findings underscore the necessity of leveraging diverse data sources and advanced architectures to enhance forecasting performance [21].

Furthermore, improvements in hybrid modeling strategies have demonstrated superior generalization capability across various financial instruments and market conditions. For example, hybrid CNN-Transformer frameworks with cross-attention fusion and recurrent units effectively capture short-term fluctuations and long-term trends simultaneously, outperforming conventional

CNN or Transformer-only models in wind speed and energy forecasting scenarios [11]. Such cross-domain evidence suggests that hybrid architectures can provide robust solutions for highly stochastic environments like stock markets [17].

In the context of stock price prediction, integrating INFO with TCN and Transformer architectures provides a unified approach that combines parameter optimization, local feature extraction, and global attention-based sequence modeling. The INFO-TCN-iTransformer model represents this synergy, offering three complementary advantages: (1) the INFO module guides parameter selection for stable and high-precision learning; (2) the TCN module captures local temporal dynamics and preserves long-term dependencies through dilated convolutions and residual connections [47]; (3) the iTransformer module leverages multi-head attention to model global dependencies and variable correlations among multivariate financial time series [35].

Empirical studies validate the efficacy of this hybrid approach. For instance, TCN-Transformer variants have been successfully applied to electricity price forecasting [25], wind speed prediction , and tool wear estimation [22], highlighting their capacity to handle both temporal and multivariate dependencies. Additionally, INFO-guided hybrid models demonstrate enhanced convergence stability and generalization, outperforming traditional deep learning baselines and ensemble methods [12]. Beyond predictive accuracy, these models also enable interpretable insights into variable importance and temporal influence, supporting more informed investment and risk management decisions [5].

Despite these advances, challenges remain. High-frequency financial data often exhibit abrupt structural changes, volatility clustering, and nonstationarity, requiring models that can adapt dynamically to changing patterns.Single-method approaches frequently struggle to capture these complex phenomena comprehensively. Hybrid models that combine statistical optimization, temporal convolution, and attention-based global modeling offer a viable pathway to address these limitations, ensuring both precision and robustness in stock price prediction tasks [35].

2 Theoretical Basis

2.1 Model Principles

2.1.1 Causal Convolution

Causal convolution ensures that the output at time step t depends only on the current and previous inputs $[x_0, x_1, \ldots, x_t]$, preventing information leakage from the future. Here, k denotes the convolution kernel size, and w_i represents the convolution weights. This mechanism conforms to the temporal causality inherent in stock market prediction. Formally, it can be expressed as:

$$y(t) = \sum_{i=0}^{k-1} w_i \cdot x(t-i) \tag{1}$$

2.1.2 Dilated Convolution

To effectively model long-range dependencies, TCNs introduce a dilated convolution mechanism on top of causal convolutions. By using a dilation factor d, the convolutional receptive field is expanded, enabling efficient modeling of dependencies over long time horizons. The computation can be expressed as:

$$y(t) = \sum_{i=0}^{k-1} w_i \cdot x(t - d \cdot i) \tag{2}$$

This architecture allows the model's receptive field to grow exponentially with network depth, enabling the capture of long-range, cross-period dependency patterns without significantly increasing the number of parameters or computational complexity. Moreover, due to its convolutional nature, TCNs maintain a high degree of parallelization, markedly outperforming recurrent architectures in training efficiency and scalability.

In financial time series modeling, this mechanism enables the model to simultaneously capture short-term fluctuation patterns and long-term trend evolution within a limited number of layers, thereby constructing a unified multi-scale representation space in Fig. 1.

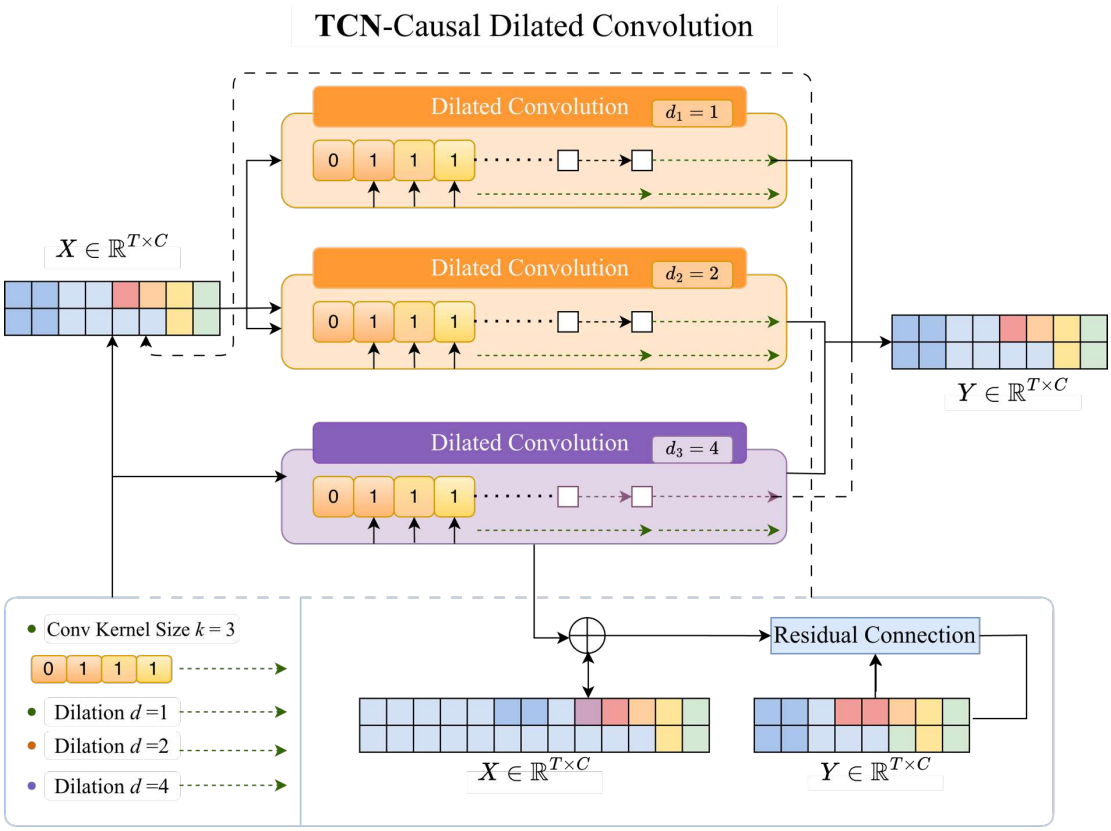


Fig. 1 Illustration of the computation mechanism of TCN causal dilated convolutions

2.1.3 Residual Connection

Residual connections directly link the input to the output, allowing training data to propagate unobstructed through multiple network layers, which enhances model trainability and stability. To further improve training stability and gradient flow in deep networks, TCNs incorporate a residual

connection structure that enables cross-layer information to pass through via identity mapping. Its formulation can be expressed as:

$$y = F(x) + x \tag{3}$$

Among them, F(x) denotes a nonlinear mapping module composed of convolutional layers, normalization, and nonlinear activation functions. The residual structure effectively mitigates the gradient degradation problem during the training of deep networks, enhances the stability of model convergence and generalization capability, while preserving the joint expressive power of original input features and high-level abstract features.

Through the synergistic combination of causal convolutions, dilated convolutions, and residual connections, TCN achieves efficient modeling of the multi-scale dynamic structures in financial time series while adhering to strict temporal causality constraints. This provides a stable, low-noise, and highly expressive foundation of temporal feature representations for subsequent high-level structural modeling modules.

2.2 Variable Structure Modeling Module: iTransformer

In the conventional Transformer, each timestep serves as the unit for attention modeling, and its self-attention computation is formulated as[1,13]:

$$Attention(Q, K, V) = softmax\left(\frac{QK^{\top}}{\sqrt{d_k}}\right) \tag{2}$$

Here, Q, K, and Vrepresent the query, key, and value matrices, respectively, and $d_k$denotes the dimension of the key vectors. This modeling paradigm implicitly assumes that the temporal dimension is the dominant factor for predictive decisions. In financial time series, however, it is susceptible to non-stationary fluctuations and noise, which can lead to instability in learning structural dependencies, as illustrated in Fig. 2.

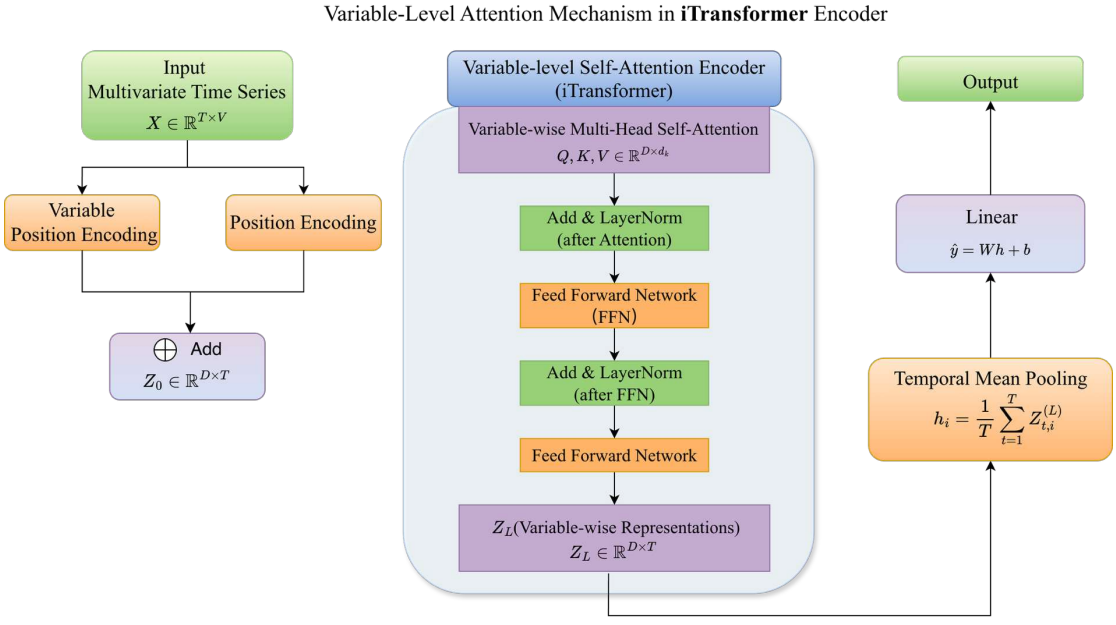


Fig. 2 Schematic Diagram of iTransformer Variable Structure Modeling and Attention Mapping Mechanism

iTransformer reconstructs the input representation structure by shifting the attention focus from the temporal dimension to the variable dimension, thereby enabling explicit modeling of inter-variable structural relationships[28,32]. Let the feature representation processed by TCN be $\mathbf{Z} \in \mathbb{R}^{D\times T}$, where D denotes the number of variables and T the sequence length[42]. iTransformer treats each variable as a token and employs a variable-level self-attention mechanism to learn structural dependencies among variables, systematically modeling interactions among multi-source financial variables such as prices, trading volumes, and technical indicators at the structural level.

2.3 Information-Driven Optimization Mechanism: INFO Outer-Layer Optimization Mode

The INFO outer-layer optimization mechanism focuses on statistical stability of predictive performance as its core optimization objective[7,23,26]. It treats model performance as a random variable and guides the parameter search process using statistical measures, simultaneously balancing prediction accuracy and model stability during optimization[43,50]. Let the model parameter configuration be θ, and let its predictive performance on the validation set be evaluated multiple times to obtain a sequence of determination coefficients $\{R_k^2(\theta)\}_{k=1}^{K}$. Based on this, the INFO scoring function is defined as:

$$S(\theta) = E!\left[R^2(\theta)\right] - \lambda\, Std!\left[R^2(\theta)\right] \tag{3}$$

Here, $\mathbb{E}(\cdot)$ and $\mathrm{Std}(\cdot)$ denote the expectation and standard deviation of predictive performance, respectively, and λ is the stability penalty coefficient. By incorporating a stability constraint into the optimization objective, this function guides the parameter search process to not only pursue high prediction accuracy but also mitigate performance volatility risk, thereby steering the optimization toward parameter regions with stronger generalization and robustness.

Within the overall framework, INFO serves as an outer-layer optimization mechanism nested outside the INFO–TCN–iTransformer model structure. It conducts a global, information-driven search over key structural parameters and training hyperparameters, realizing a dual-level collaborative modeling mechanism of “model structure modeling + information-driven optimization,” as illustrated in Fig. 3：

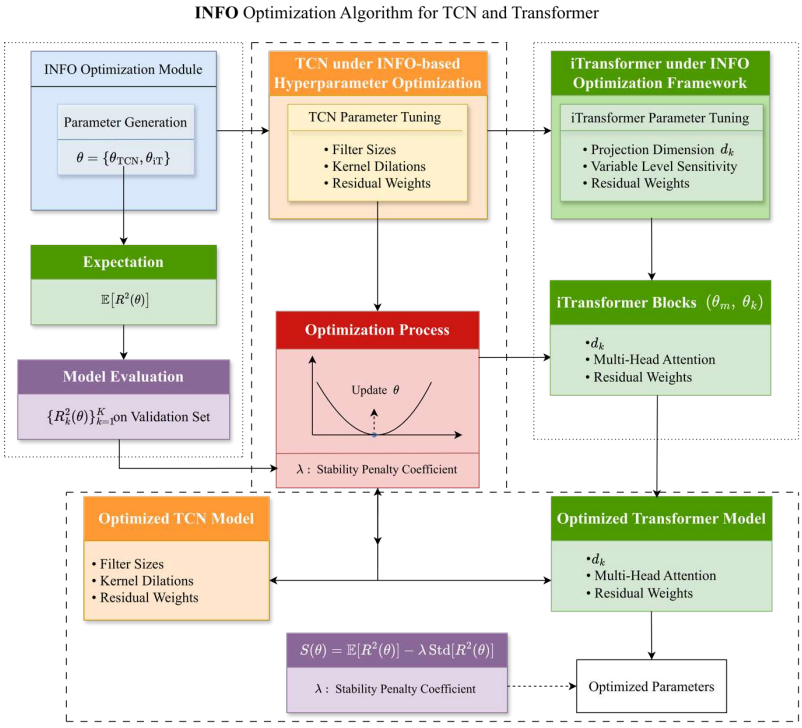


Fig. 3 Framework of INFO-driven joint optimization mechanism for TCN and iTransformer

3 Model Construction

3.1 INFO–TCN–iTransformer Model Architecture and Principles

The proposed INFO–TCN–iTransformer framework consists of three components: a TCN-based temporal modeling branch, an iTransformer-based variable-structure modeling branch, and the INFO outer-layer optimization mechanism.

Specifically, the TCN models local dynamic structures of time series through causal dilated convolutions, while the iTransformer applies self-attention along the variable dimension to capture structural dependencies among multiple variables. INFO, as an outer-layer optimization framework, operates solely on the global search of model structural parameters and key without participating in forward propagation or gradient updates, as illustrated in Fig. 4.

This architecture achieves a functional decoupling between temporal dynamics modeling and variable-structure modeling at the modeling level, while simultaneously enhancing prediction accuracy and generalization performance under a unified stability constraint.

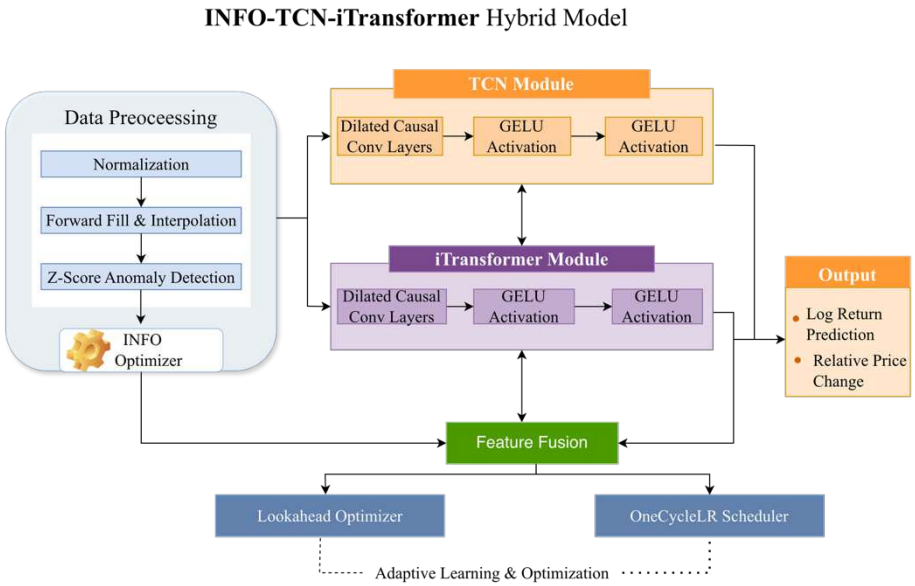


Fig. 4 INFO–TCN–iTransformer Schematic of the hybrid prediction model optimization framework

3.1.1 Input Representation and Sample Construction

Let the input sample tensor be $\mathbf{X} \in \mathbb{R}^{B\times T\times D}$, where B denotes the batch size, $T = 96$represents the length of the time window, and $D = 7$ is the feature dimension, including the opening price, highest price, lowest price, closing price, return, amplitude, and turnover rate.

3.1.2 TCN Branch: Temporal-Constrained Modeling

The TCN branch constructs a temporal-constrained modeling structure through multi-layer causal dilated convolution modules. The convolution operation at the l-th layer is defined as:

$$y_{t^{(l)}} = \sum_{k=0}^{K-1} b\Pi_{k^{(l)}} \cdot h_{t-d_l\cdot k}^{(l-1)}, \quad d_l = 2^{l-1} \tag{4}$$

Each layer consists of convolution, layer normalization, ReLU activation, and a residual connection. When the channel dimensions are inconsistent, a 1×1 convolution is used for alignment. After stacking L_{tcn}layers, the hidden state at the final time step is taken as the temporal feature representation.

$$\boldsymbol{h}^{TCN} = \boldsymbol{H}_{:,-1}^{(L_{tcn})} \tag{5}$$

The hidden state corresponding to the last time step is adopted as the temporal feature representation.

3.1.3 iTransformer Branch: Variable-Structure Modeling

The iTransformer branch is designed to capture structural dependencies in multivariate financial time series. It organizes input features along the variable dimension so that the self-

attention mechanism operates on variable-level tokens, thereby explicitly modeling structural dependencies among multiple variables.

Let the TCN output features be $\mathbf{Z} \in \mathbb{R}^{D \times T}$. After encoding through multiple layers of variable-level attention, aggregation along the temporal dimension yields the variable-structure representation.

$$\boldsymbol{h}^{\mathrm{i}} = \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{Z}_{\boldsymbol{t}}^{(\boldsymbol{L_i})} \tag{6}$$

This design achieves a decoupling of temporal dynamics and variable structures: the temporal dimension provides local dynamic information, while global predictive decisions are primarily governed by the variable structure.

3.1.4 Feature Fusion and Multi-Objective Prediction

The feature vectors from the TCN and iTransformer branches are concatenated along the feature dimension and mapped into a shared latent space.

$$\boldsymbol{h}_f = \phi([\boldsymbol{h}^{TCN}; \boldsymbol{h}^{iT}]) \tag{7}$$

Based on the fused features, multiple prediction heads are constructed to model log returns and relative price changes separately. The price-level prediction is then recovered via a change-rate recursive mechanism, ensuring structural consistency among financial variables.

3.2 Model Optimization Objectives and Training Strategy

Model training is performed using a multi-objective joint loss function.

$$\mathcal{L} = \lambda_r \mathcal{L}_{ret} + \lambda_p \mathcal{L}_{pct} + \lambda_s \mathcal{L}_{price} \tag{8}$$

Specifically, the Log-Cosh loss is used to enhance robustness against anomalous fluctuations, the mean squared error (MSE) constrains the prediction accuracy of change rates, and the Smooth L1 loss is applied to the price-level reconstruction error. Model parameters are updated using the AdamW optimizer, while the learning rate and key hyperparameters are adaptively searched by INFO based on validation set performance, thereby reducing the uncertainty associated with manual tuning.

4 Empirical Design and Results Analysis

4.1 Data Sources and Sample Construction

4.1.1 Data Acquisition and Sample Overview

This chapter presents an empirical analysis of the predictive performance, structural effectiveness, and interpretability of the INFO–TCN–iTransformer model. Comparative experiments are conducted on the A-share index and individual stocks from different industries, such as Huajian Group and Fuyao Glass, to evaluate the model's prediction accuracy, stability, and generalization capability under various market conditions and risk states. The data are daily frequency and sourced from the Tonghuashun financial terminal. Input features include opening

price, highest price, lowest price, closing price, return, amplitude, and turnover rate, with the closing price serving as the prediction target.

4.1.2 Data Preprocessing and Sliding Window Construction

Prior to the empirical analysis, missing values in the raw series are handled, and numerical normalization is applied using a Min–Max method based on the training set. Subsequently, supervised learning samples are constructed using a sliding window approach, with a window length of $T = 96$. Each sample contains a multivariate feature sequence of 96 consecutive trading days, with the input representation given by:

$$\mathbf{X} \in \mathbb{R}^{B \times T \times D},$$

The samples are divided into training, validation, and test sets in chronological order, strictly adhering to the causality constraints inherent in time series forecasting.

4.2 Evaluation Metrics

To comprehensively evaluate the model performance, four types of metrics are selected:

1. Root Mean Squared Error (RMSE): Sensitive to large deviations, used to assess predictive risk and the model's ability to handle extreme fluctuations.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{9}$$

2. Mean Absolute Error (MAE): Reflects the average level of prediction error.

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \tag{10}$$

3. Mean Absolute Percentage Error (MAPE): Converts absolute errors into percentages, facilitating comparisons across securities with different price levels.

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100\% \tag{11}$$

4. Coefficient of Determination (R^2): Measures the model's ability to explain price fluctuations, with values closer to 1 indicating more accurate capture of the main trends.

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} \tag{12}$$

4.3 Model Training and Parameter Configuration

The key model structures and training hyperparameters are adaptively searched and scheduled by the INFO outer-layer optimization mechanism under validation set performance constraints, balancing complexity, stability, and generalization capability. Training is performed using an adaptive optimization strategy with weight decay, combined with dynamic learning rates and an early stopping mechanism to improve convergence efficiency and reduce overfitting.

The local temporal features from the TCN and the global structural features from the iTransformer are fused through feature concatenation and nonlinear mapping to form the final latent representation. The core structural parameters and search settings are listed in Table 1.

Table 1 Key Parameters of the INFO Model and Their Functions

Parameter	Value	Design Rationale
Optimization Objective	Validation Set Price-Level R^2	Using price-level prediction performance as the direct optimization objective avoids bias introduced by evaluating only in return or differential spaces
Search Parameters	Learning Rate, TCN Depth, Hidden Dimension, Regularization Strength	Collaborative search of structural and training parameters achieved via INFO
Hidden Dimension	64	Balances expressive capability with computational efficiency
Number of Attention Heads	4	Enhances multi-perspective feature modeling capability
Activation Function	GELU	Maintains consistent nonlinear expression with the TCN branch
Dropout	0.1	Improves model robustness

Through validation set feedback, the INFO mechanism effectively regulates model parameters, achieving robust configurations across different asset classes and volatility regimes, thereby enhancing prediction accuracy and improving generalization performance.

4.4 Training Mechanism and Optimization Strategy

To enhance the training stability and out-of-sample generalization capability of the INFO–TCN–iTransformer model in nonlinear financial time series forecasting tasks[9,14,36], this study establishes a unified training mechanism from three aspects: activation function configuration, loss function design, and optimization strategy.

In terms of activation function selection, the TCN branch adopts the ReLU function：

$$ReLU(x) = max(0, x) \tag{13}$$

To improve the efficiency of local temporal feature extraction and alleviate the vanishing gradient problem in deep networks; the iTransformer branch, in contrast, adopts the GELU activation function.

$$GELU(x) = 0.5x\left(1 + tanh\left(\sqrt{\frac{2}{\pi}}(x + 0.044715x^3)\right)\right) \tag{14}$$

In the loss function design, to balance strong penalization of significant prediction deviations with robustness in overall error measurement, a weighted combination of MSE and MAE is adopted.

$$HybridLoss = \alpha \cdot MSE + (1 - \alpha) \cdot MAE \tag{15}$$

This hybrid loss exhibits smoother convergence behavior and more stable predictive performance across multiple experimental settings.

Model training employs the RAdam optimizer combined with the Lookahead mechanism to mitigate the instability of adaptive learning rate updates in small-sample scenarios. In addition, the OneCycleLR learning rate scheduling strategy is introduced to accelerate convergence through dynamic learning rate adjustment, reduce training oscillations, and enhance optimization efficiency and training robustness.

4.5 Experiments and Interpretability Analysis

4.5.1 Prediction Experiments and Results Analysis

As shown in the results in Table 2 the INFO–TCN–iTransformer demonstrates significant predictive advantages across the three stocks. Taking the A-share index as an example, INFO–TCN–iTransformer achieves MAE = 8.9717, RMSE = 9.6929, and R^2 = 0.9984, showing substantial improvements over traditional models such as LSTM, Transformer, and CNN. While the single-structure models (TCN and iTransformer) also exhibit strong performance, they remain inferior to the complete integrated framework.

Similar patterns are observed in the Fuyao Glass and Huajian Group datasets. In particular, the INFO–TCN–iTransformer model achieves outstanding performance on Huajian Group, with MAE = 0.0524, RMSE = 0.1314, and R^2 = 0.9989, significantly outperforming other ablation models. These results demonstrate the model's robustness and high predictive accuracy under high-volatility conditions and industry-specific constraints, highlighting its stability and effectiveness in complex financial environments.

Table 2 Comparison of Evaluation Results of Different Models on Various Datasets

Stock	Model	MAE	MAPE%	R^2	RMSE
A-share Index	BP	65.5647	1.8676	0.9199	78.4785
	LSTM	33.8371	0.9709	0.9745	44.3869
	CNN	84.9984	2.4073	0.8802	96.0037
	Transformer	32.1366	0.9287	0.9759	43.0514
	TCN	25.7349	0.83	0.9763	30.0367
	iTransformer	28.7383	0.9214	0.9730	38.4762
	INFO-TCN	31.7077	1.0061	0.9810	32.9868
	INFO-iTransformer	63.1184	1.9933	0.9158	69.3894
	TCN-iTransformer	23.9520	0.7731	0.9792	34.4921
	INFO-TCN-iTransformer	**8.9717**	**0.2861**	**0.9984**	**9.6929**
Fuyao Glass	BP	1.1148	2.28	0.9752	1.5307
	LSTM	0.8052	1.71	0.9865	1.1300
	CNN	1.1096	2.44	0.9775	1.4565
	Transformer	3.6476	6.67	0.7033	5.2944
	TCN	1.0580	2.23	0.9801	1.3763
	iTransformer	0.9621	1.9255	0.9819	1.3112
	INFO-TCN	1.3746	2.8389	0.9769	1.4959
	INFO-iTransformer	0.7883	1.7159	0.9889	1.0283
	TCN-iTransformer	2.6976	5.73	0.9191	2.7976
	INFO-TCN-iTransformer	**0.4259**	**0.8378**	**0.9968**	**0.5592**

Huajian Group	BP	0.5012	6.83	0.9567	0.8089
	LSTM	0. 3954	3.01	0.9836	0.4983
	CNN	0.3954	5. 12	0.9641	0.7365
	Transformer	0. 2385	2.71	0.9807	0.5398
	TCN	0.2040	2.38	0.9876	0.4353
	iTransformer	0.2040	2.3873	0.9877	0.4328
	INFO-TCN	0.2043	2.2540	0.9794	0.5724
	INFO-iTransformer	0.2050	2.3690	0.9877	0.4426
	TCN-iTransformer	0. 2052	2.4132	0.9882	0.4325
	INFO-TCN-iTransformer	**0.0524**	**0.6469**	**0.9989**	**0.1314**

In the Fuyao Glass sample, the INFO–TCN–iTransformer model demonstrates superior performance in terms of MAE (0.4259), RMSE (0.5592), and R^2 (0.9968), indicating strong complementarity between the TCN and iTransformer branches for stocks with clear industry characteristics and relatively stable volatility, while the INFO optimization mechanism enhances training adaptability and stability.

The training loss curves are smooth and converge rapidly, with no signs of overfitting or performance degradation. Out-of-sample prediction errors are substantially reduced, validating the effectiveness of the INFO mechanism in improving generalization capability. Compared with LSTM, BP, CNN, and Transformer models, the INFO–TCN–iTransformer outperforms traditional approaches in convergence speed, out-of-sample stability, and fitting accuracy.

4.5.2 Analysis of Experimental Results for Mainstream Comparative Models

Compared with traditional models such as BP neural networks and CNNs, predictive performance on both index and individual stock samples is noticeably weaker, with larger errors, indicating that shallow architectures struggle to capture long-term dependencies and multi-scale features.Although LSTM and Transformer models improve overall trend fitting, they still exhibit prediction lag and error accumulation under amplified market volatility, suggesting limited adaptability to complex and high-risk market regimes as illustrated in Fig. 5, 6.

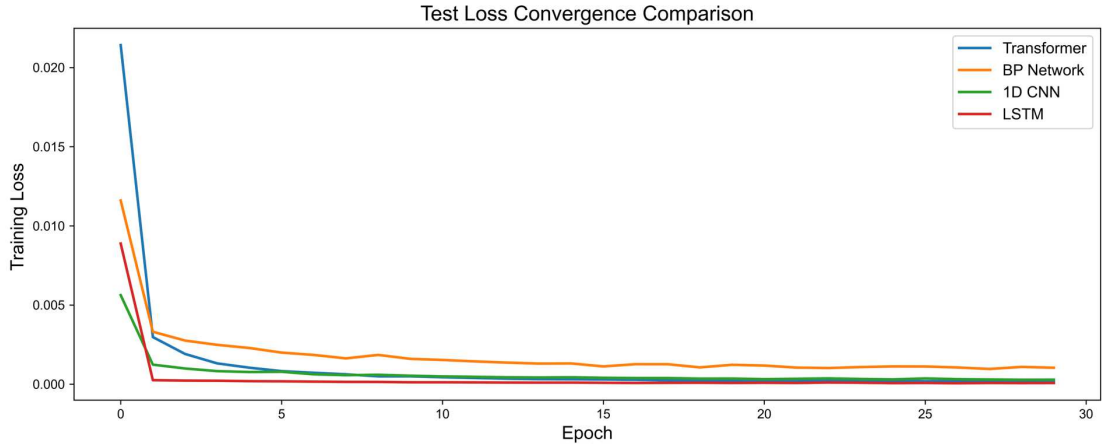


Fig. 5 Training Loss Convergence Curves of Different Models for A-Share Index

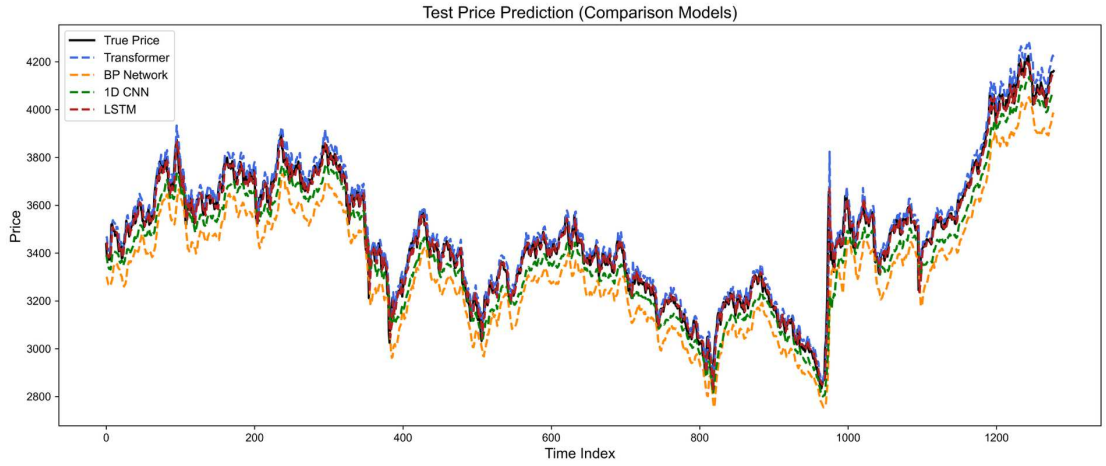


Fig. 6 Predictive fitting comparison on the A-share market index

LSTM and Transformer show clear improvements in fitting the overall trend on the A-share market index sample, but still exhibit noticeable lag in locally high-volatility intervals.

In individual stock samples (such as Fuyao Glass and Huajian Group), traditional models display more pronounced prediction deviations during high-volatility periods, while the Transformer becomes unstable under small sample size and high-noise conditions. These observations provide a baseline reference for subsequent structural modeling improvements as illustrated in Fig. 7.

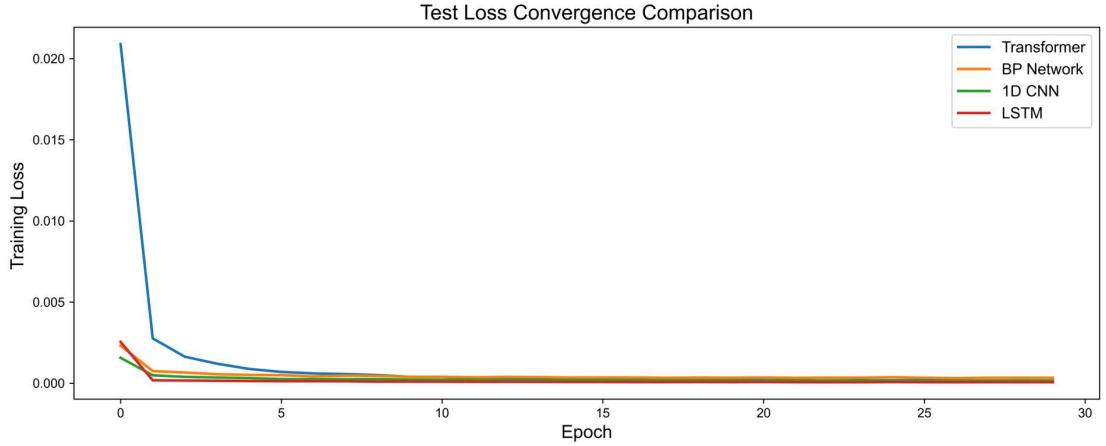


Fig. 7 Training Loss Convergence Curves of Different Models for Fuyao Glass Samples

Training loss curves of mainstream comparative models in the Fuyao Glass sample, where LSTM and Transformer exhibit more stable convergence than BP and CNN as illustrated in Fig. 8.

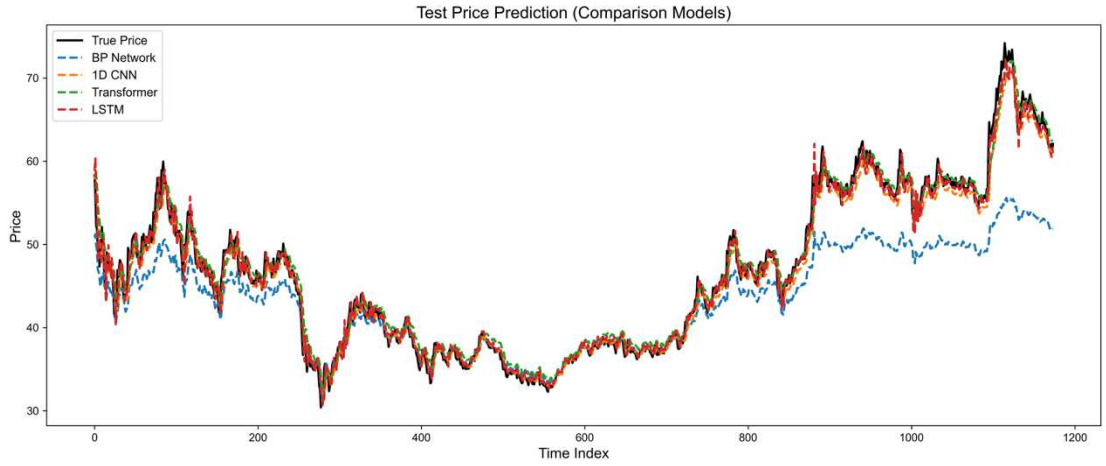


Fig. 8 Predictive fitting comparison on the Fuyao Glass

Predictive fitting curves of mainstream comparative models on the Fuyao Glass sample, where LSTM and Transformer capture mid- to short-term trends well, while traditional models exhibit large errors in high-volatility periods as illustrated in Fig. 9.

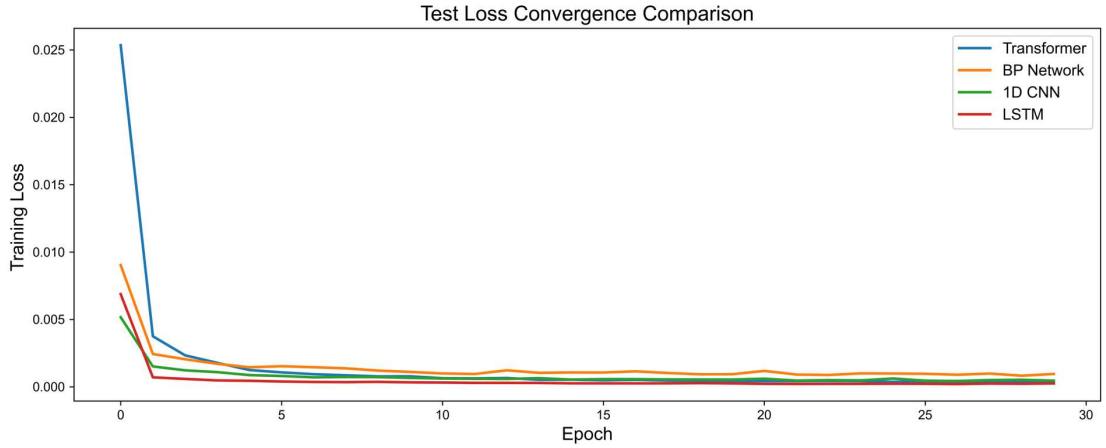


Fig. 9 Training Loss Convergence Curves of Different Models for Huajian Group

Training loss curves of mainstream comparative models on the Huajian Group sample, where LSTM and Transformer converge faster and exhibit better stability as illustrated in Fig. 10.

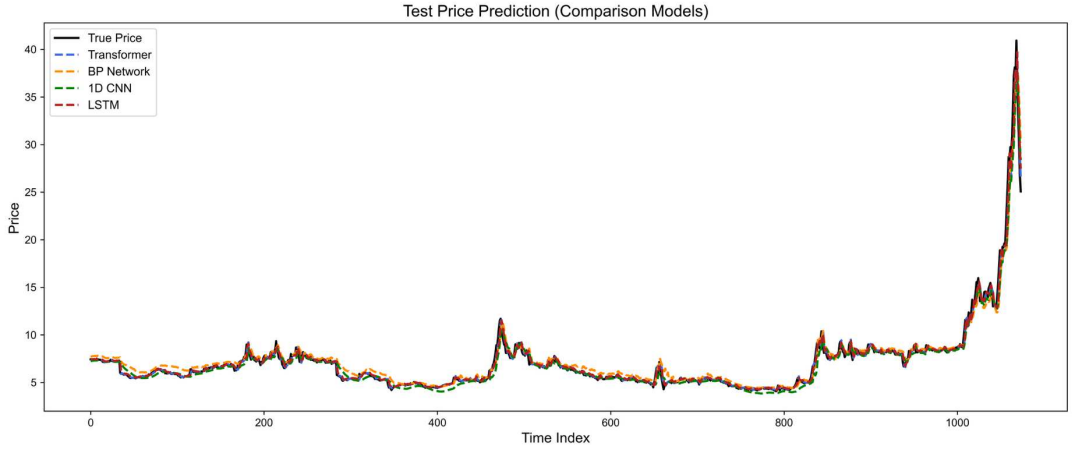


Fig. 10 Predictive fitting comparison on the Huajian Group

Predictive fitting curves of mainstream comparative models on the Huajian Group sample, where Transformer fits the overall trend better than traditional models but still lags in locally high-volatility intervals.

4.5.3 INFO–TCN–iTransformer Analysis of Model Prediction Results

Table 2 summarizes the predictive performance of the INFO–TCN–iTransformer and comparative models on the A-share index, Fuyao Glass, and Huajian Group samples. The results indicate that INFO–TCN–iTransformer consistently outperforms across different market levels and risk environments, leading in key metrics such as MAE, RMSE, and R^2. Taking the A-share index as an example, INFO–TCN–iTransformer accurately follows medium- to long-term trend changes and maintains high consistency during periods of amplified volatility, demonstrating excellent overall fitting and error-control capabilities as illustrated in Fig. 11.

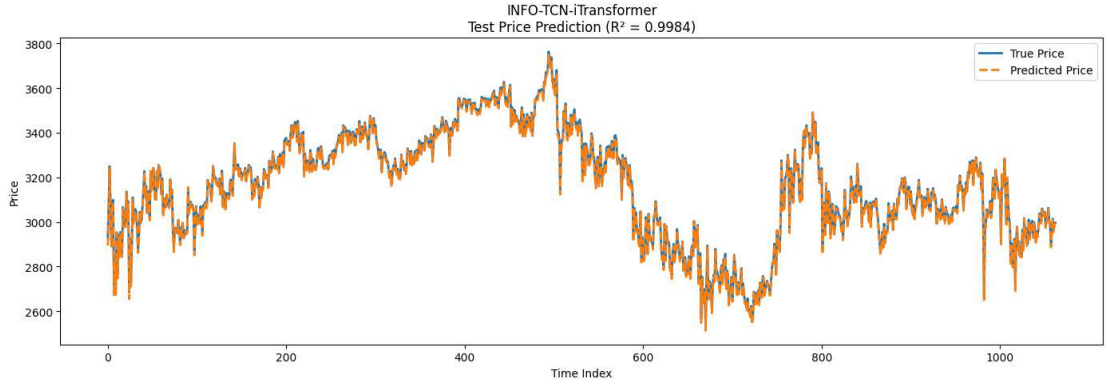


Fig. 11 INFO–TCN–iTransformer predictive fitting on the A-share index

Predictive fitting curves of INFO–TCN–iTransformer on the A-share index sample, accurately capturing medium- to long-term trends with smooth and stable handling of local fluctuations as illustrated in Fig. 12.

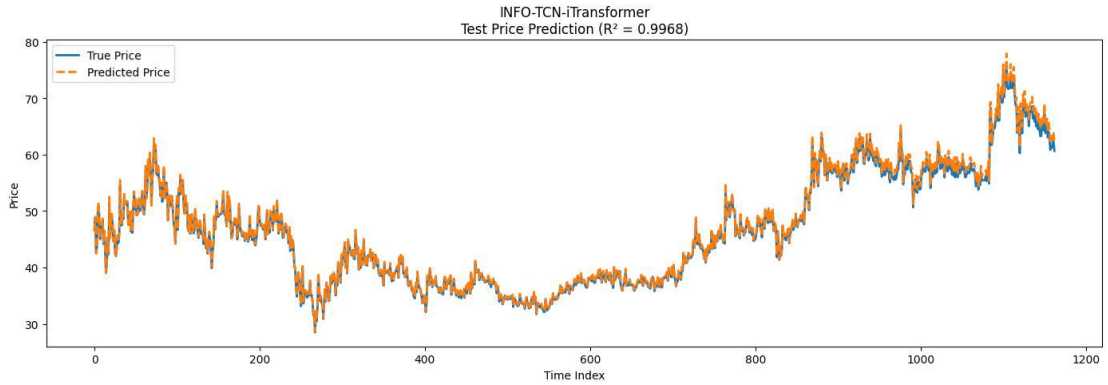


Fig. 12 INFO–TCN–iTransformer predictive fitting on the Fuyao Glass

Predictive fitting curves of INFO–TCN–iTransformer on the Fuyao Glass sample, with prediction errors significantly reduced during high-volatility periods as illustrated in Fig. 13.

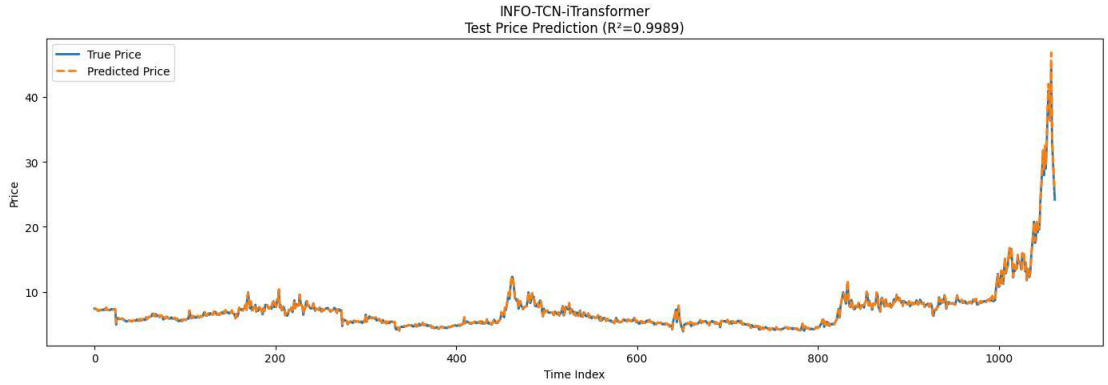


Fig. 13 INFO–TCN–iTransformer Prediction Fitting for Huajian Group

Predictive fitting curves of INFO–TCN–iTransformer on the Huajian Group sample, accurately capturing the overall trend with smooth handling of local fluctuations.。

4.5.4 Ablation Study and Structural Contribution Analysis

The ablation experiments evaluate the effects of different model structures and information-enhancement mechanisms on predictive performance, validating the contributions of each module in the INFO–TCN–iTransformer framework to error control, fitting accuracy, and stability.

Under identical datasets and parameter settings, multiple ablation models are constructed and compared. The results are illustrated through predictive fitting curves, revealing the relationship between model structural complexity and forecasting performance.

Taking the A-share market index as an example, the training loss decreases from 62.77 to 8.97, and the validation loss fluctuates within the range of 5.24 to 13.62, indicating that the model effectively captures volatility patterns, converges rapidly, and achieves stable fitting performance as illustrated in Fig. 14,15,16.

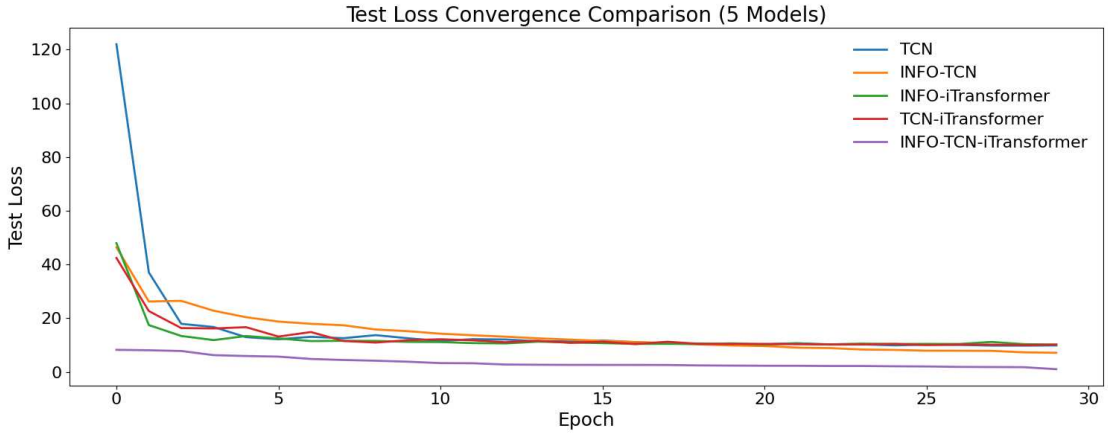


Fig. 14 Convergence curve of training loss on the A-share market index

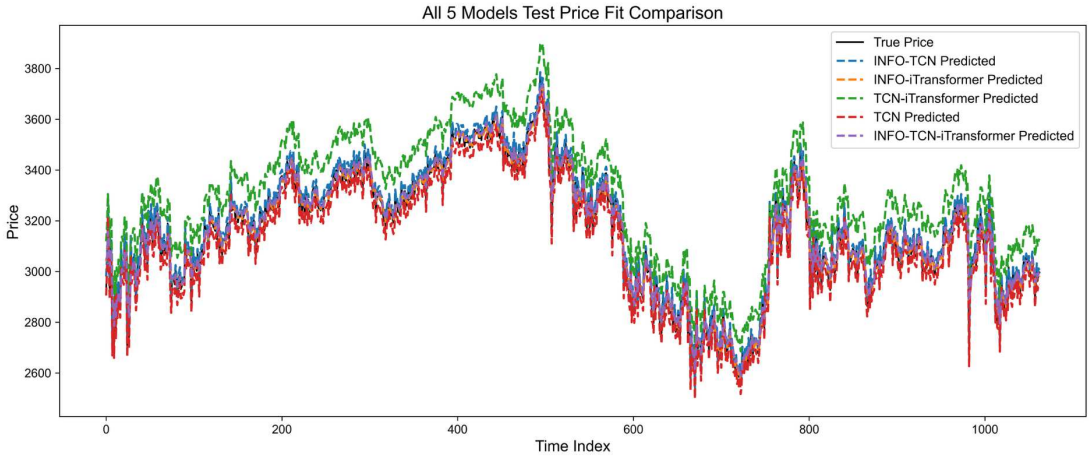
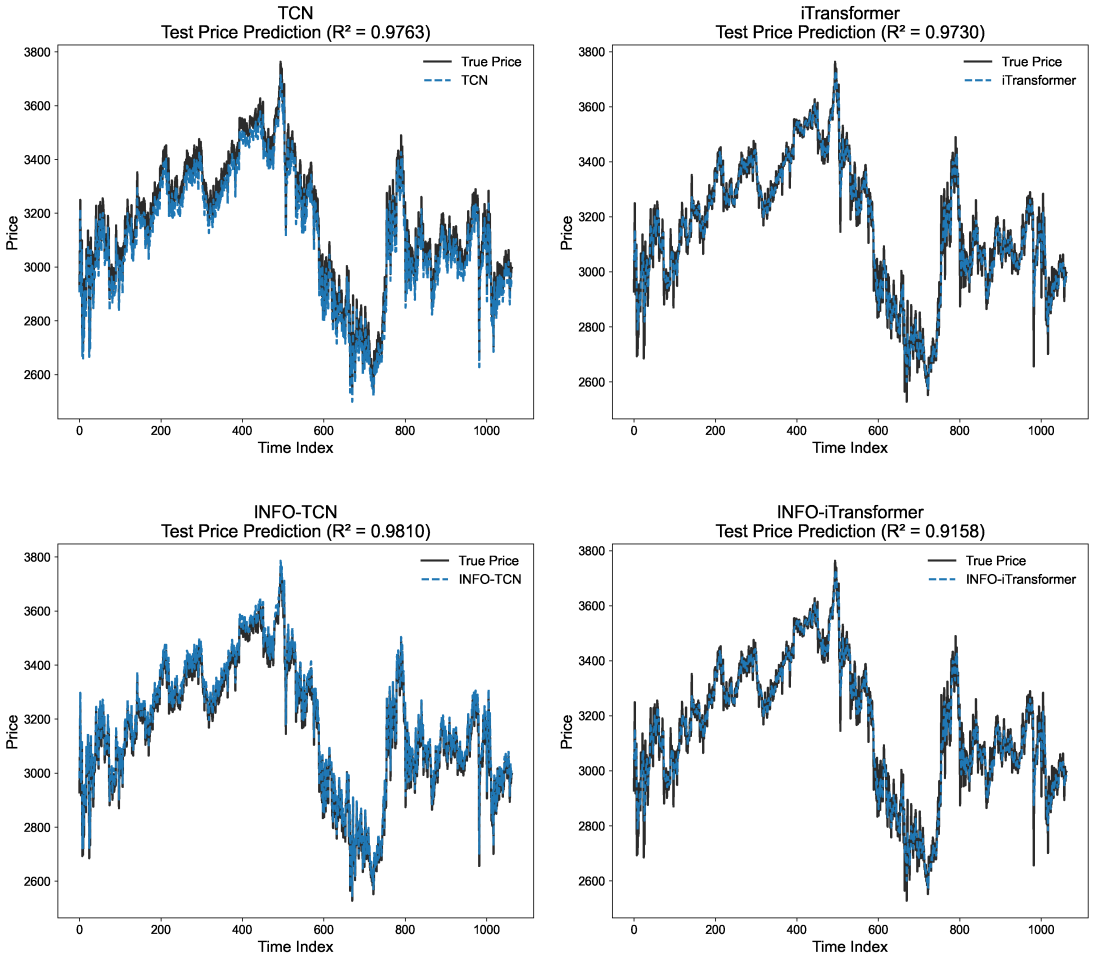


Fig. 15 Fitting curves of the A-share market index and ablation models



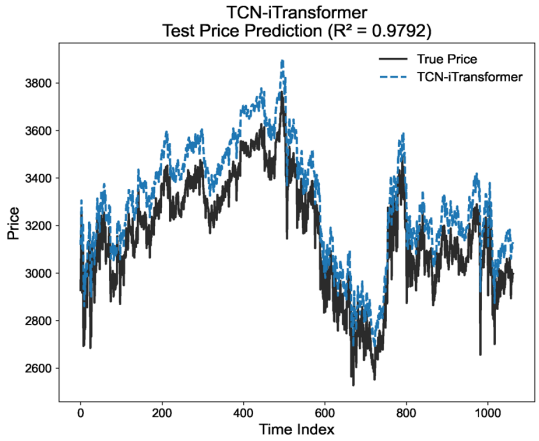


Fig. 16 Fitting curves of ablation models on the A-share market index

The ablation experiments evaluate the effects of different model structures and information-enhancement mechanisms on predictive performance, validating the contributions of each module in the INFO–TCN–iTransformer framework to error control, fitting accuracy, and stability.

Under identical datasets and parameter settings, multiple ablation models are constructed and compared. The results are illustrated through predictive fitting curves, revealing the relationship between model structural complexity and forecasting performance.

Taking the A-share market index as an example, the training loss decreases from 62.77 to 8.97, and the validation loss fluctuates within the range of 5.24 to 13.62, indicating that the model effectively captures volatility patterns, converges rapidly, and achieves stable fitting performance as illustrated in Fig. 17,18,19.

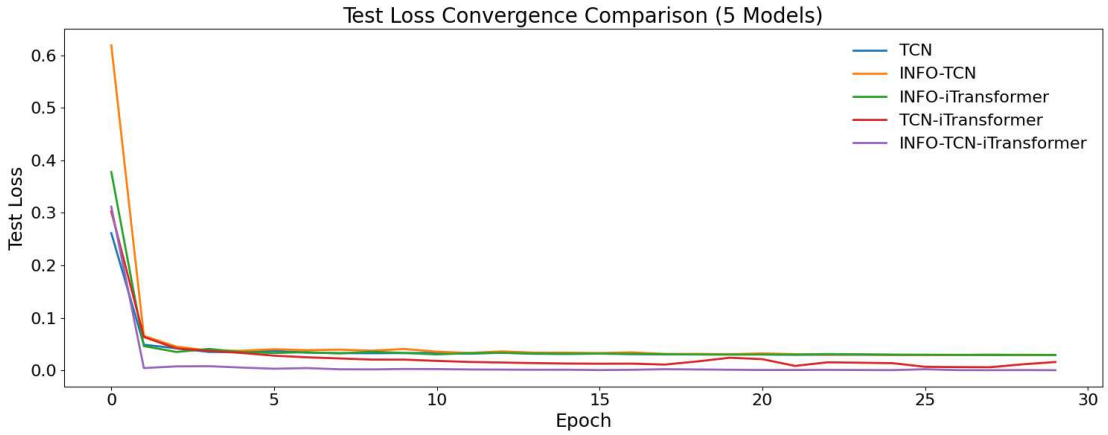


Fig. 17 Training loss convergence curve for the Fuyao Glass sample

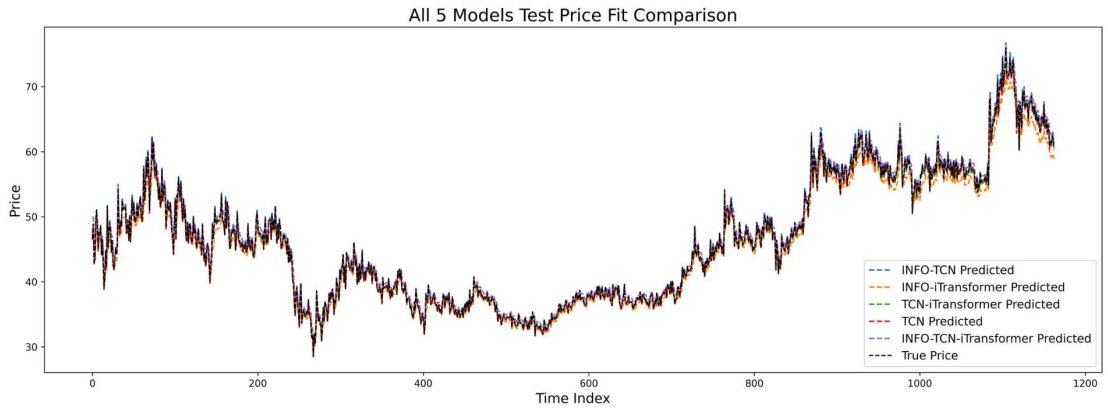


Fig. 18 Fitting curves of the Fuyao Glass sample and ablation models

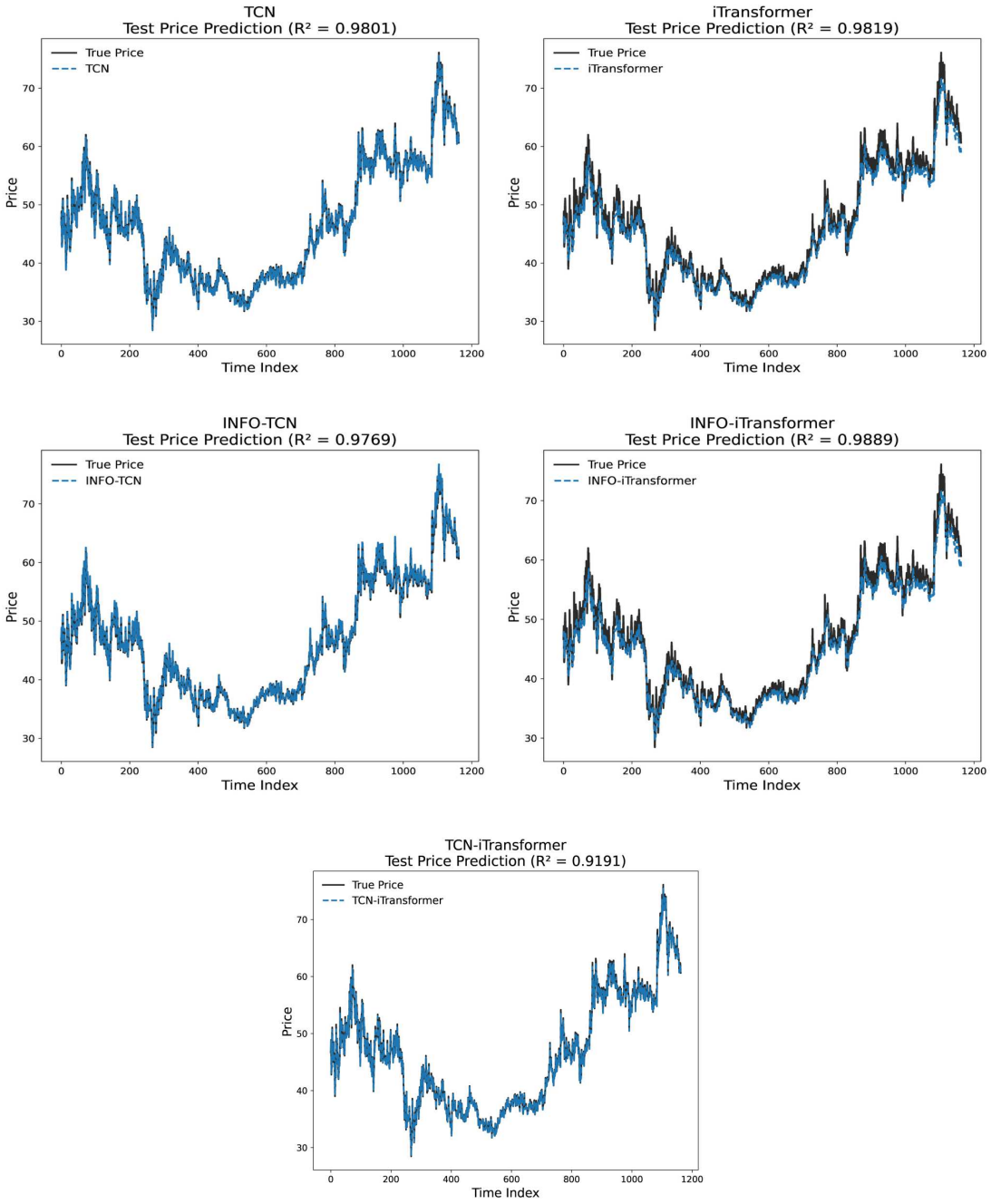


Fig. 19 Fitting curves of ablation models for the Fuyao Glass sample

Comparison of predictive fitting results of ablation models on the Fuyao Glass sample: single-structure models exhibit significant errors, while the fused architecture shows improved stability as illustrated in Fig. 20,21,22.

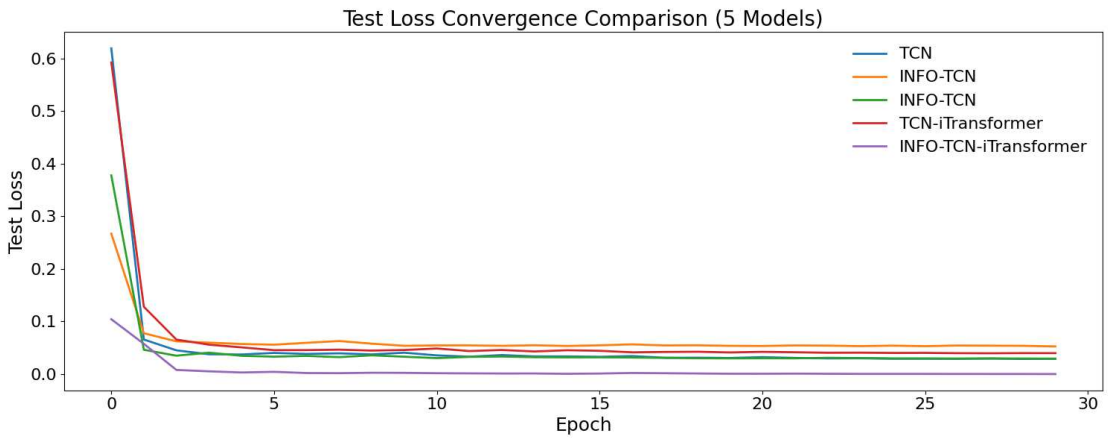


Fig. 20 Training loss convergence curve on the Huajian Group sample

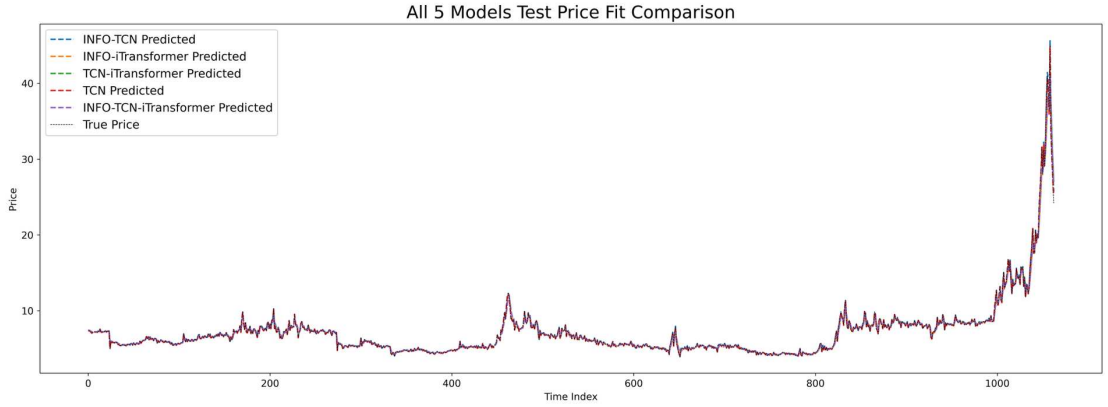
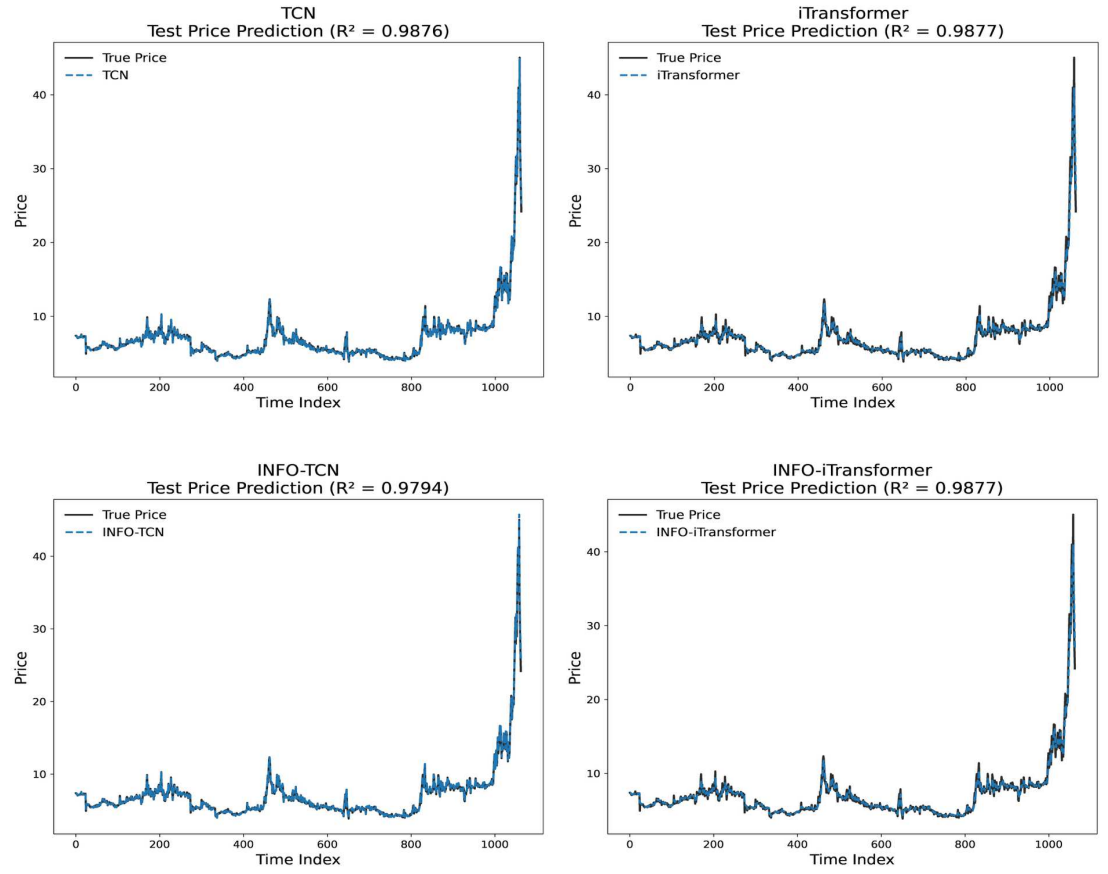


Fig. 21 Fitting curves of the Huajian Group sample and ablation models



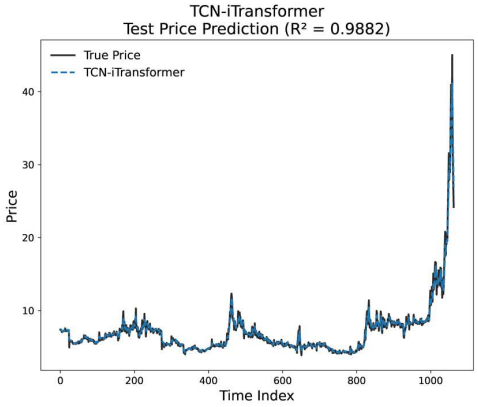


Fig. 22 Fitting curves of ablation models on the Huajian Group sample

Comparison of predictive fitting results of ablation models on the Huajian Group sample shows that structural fusion and information optimization significantly improve fitting accuracy and stability.

After introducing the INFO outer-layer information optimization mechanism, the model exhibits markedly improved predictive stability and out-of-sample performance across different samples, indicating that the information optimization module plays a critical role in mitigating training instability and reducing extreme errors. Overall, the modules form a synergistic gain within a unified framework, rather than a simple additive combination.

4.5.5 Error Distribution and Stability Analysis

To further characterize the model's robustness under extreme risk conditions, this study analyzes the distribution of prediction errors of different models on the test set.

By comparing the error distributions across the A-share market index, Fuyao Glass, and Huajian Group datasets, the INFO–TCN–iTransformer consistently exhibits highly concentrated and sharp error distributions, with most errors clustered near zero, rapidly decaying tails, and extremely low extreme-error probability. These results indicate the model's strong capability in capturing both price trends and short-term fluctuations, demonstrating superior robustness under high-risk conditions as illustrated in Fig. 23,24,25.

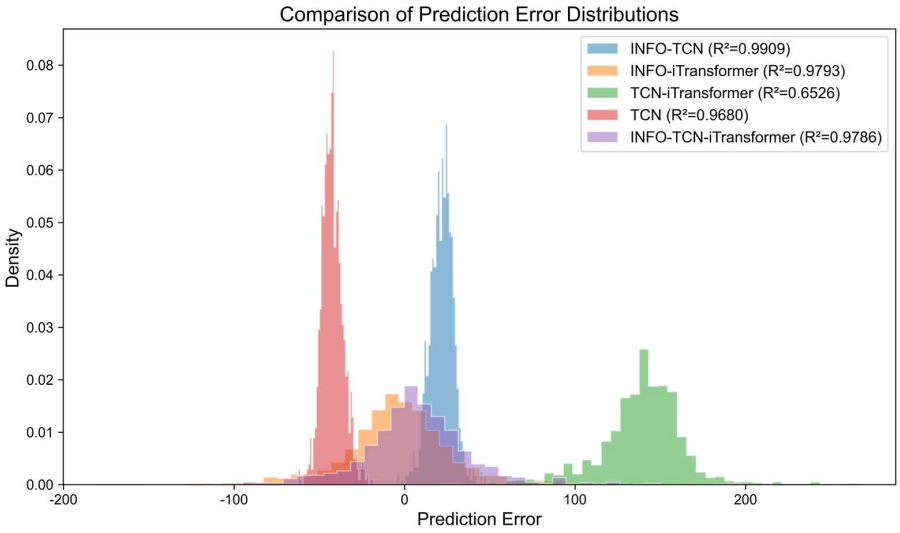


Fig. 23 Comparison of Prediction Error Distributions of Different Models on the A-share Market Index

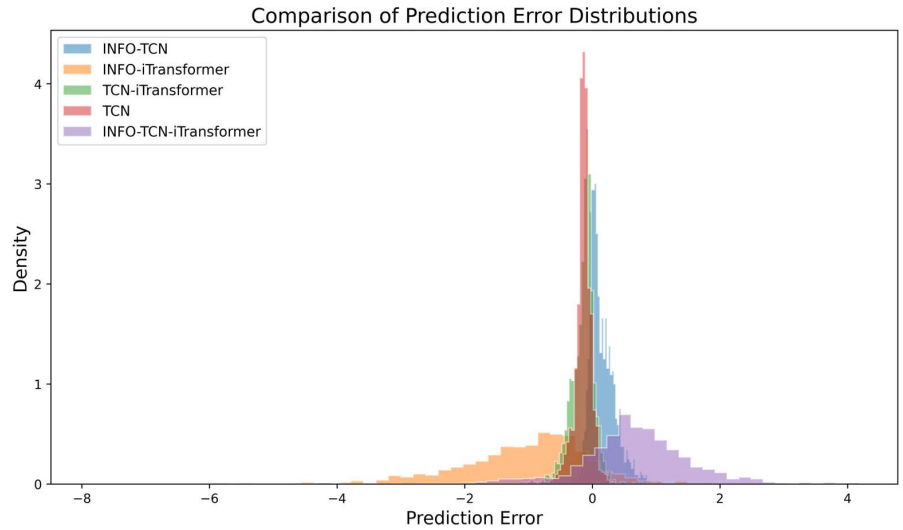


Fig. 24 Comparison of Prediction Error Distributions of Different Models on the Fuyao Glass Sample

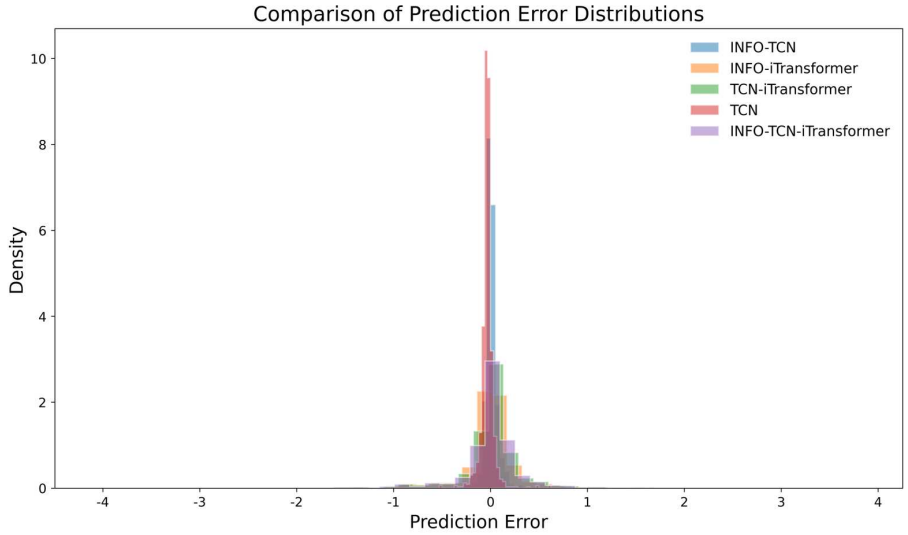


Fig. 25 Comparison of Prediction Error Distributions of Different Models on the Huajian Group Sample

The ablation experiments evaluate the impact of different model structures and information-enhancement mechanisms on predictive performance, verifying the contributions of each module in the INFO–TCN–iTransformer to error control, fitting accuracy, and stability.Under identical datasets and parameter settings, multiple ablation variants are constructed and compared, with results illustrated through predictive fitting curves, revealing the relationship between model structural complexity and forecasting performance.Taking the A-share market index as an example, the training loss decreases from 62.77 to 8.97, while the validation loss fluctuates between 5.24 and 13.62, indicating that the model effectively captures volatility patterns, converges rapidly, and achieves stable fitting.

4.5.6 Model Interpretability Analysis

To further elucidate the prediction mechanism of the INFO–TCN–iTransformer model, this study conducts an interpretability analysis of the model's decision behavior from the temporal dimension. The results show a clear non-uniform distribution of temporal attention weights: recent historical periods receive significantly higher weights than distant ones, indicating that the model effectively

identifies short-term dependencies and treats recent market states as key drivers of future trend prediction.At the same time, although the weights decay with increasing temporal lag, the model retains a certain degree of long-term memory, demonstrating sensitivity to long-term trends while capturing short-term dynamics. This adaptive temporal weighting mechanism does not rely on manually imposed decay assumptions, is consistent with the real operating logic of financial markets, and provides a clear explanation for the model's robustness and generalization performance in complex environments as illustrated in **Fig. 26**.

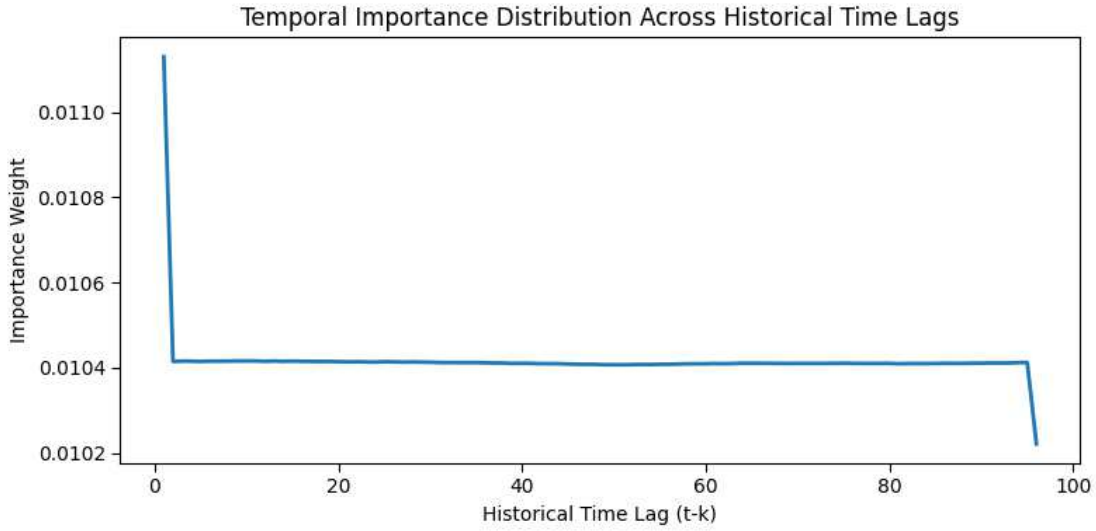


Fig. 26 Comparison of error distributions of different models on the A-share market index sample

5 Conclusion

To enhance the accuracy and stability of stock price forecasting under complex market environments, this study proposes a hybrid prediction framework, termed INFO–TCN–iTransformer, from a unified perspective of structural synergy and information enhancement. By integrating the local multi-scale temporal modeling capability of TCN with the global dependency modeling strength of iTransformer, the model achieves coordinated characterization of short-term fluctuations and long-term trends at the structural level. Meanwhile, the introduced information enhancement mechanism effectively reconstructs and filters both input features and intermediate representations, thereby improving the effectiveness and robustness of feature representation without significantly increasing model complexity.

Extensive experiments on the A-share market index and multiple individual stocks across different industries demonstrate that the proposed model consistently outperforms BP, LSTM, CNN, Transformer, and various ablation variants. In particular, INFO–TCN–iTransformer exhibits stable and superior performance on MAE, RMSE, and R^2, validating its strong generalization capability across different market structures and volatility regimes.

Further ablation studies systematically analyze the contributions of individual structural components and information mechanisms. The results indicate that single-structure models, while effective in either local or global modeling, struggle to simultaneously achieve trend representation and error control. Although simple structural combinations improve fitting capacity, they fail to achieve systematic error reduction. In contrast, the INFO–TCN–iTransformer model, through the synergistic integration of structural fusion and information optimization, achieves a superior balance

among stability, accuracy, and robustness, demonstrating that neither structural design nor information enhancement alone is sufficient to fully unlock model potential—rather, their organic integration is essential for improving financial time-series forecasting performance.

Despite the strong empirical performance observed across multiple datasets, there remains room for further improvement. On the one hand, the current framework primarily relies on historical prices and technical indicators, without explicitly incorporating exogenous factors such as macroeconomic policies, industry cycles, or market sentiment. On the other hand, the model's dynamic adaptability under extreme market volatility and event-driven non-stationary conditions still requires enhancement. Future research may focus on the deep integration of multi-source heterogeneous data—including news text, corporate announcements, and sentiment factors—within the existing framework, as well as the exploration of more adaptive structural and information interaction mechanisms, to further strengthen the model's practical applicability and decision-support value in real financial markets.

Author contributions

W.H. contributed to research conception, overall framework design, data collection, data preprocessing, and preliminary experimental analysis. (email: wuhuiyong@syuct.edu.cn)

L.L. was responsible for model construction and algorithm implementation, experimental design and execution, result analysis, and drafting of the original manuscript. (email: 13504685643@163.com)

W.R. contributed to literature review, experimental result validation, and technical discussion. (email: WangRong@syuct.edu.cn)

D.Y. contributed to manuscript revision, language editing, and formatting standardization. (email: yanliding@yeah.net)

All authors reviewed and approved the final version of the manuscript. Correspondence Address:College of science ,Shenyang University of Chemical Technology, CEP 110142, LN,CHN

Funding

This work was supported by the Basic Scientific Research Project of the Department of Education of Liaoning Province (Project title: Generative AI in Multimodal Risk Scenario Simulation and Dynamic Emergency DecisionMaking, Grant No. LJ212510149011 to HW and ZW). The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests

Te authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to J.K.

Data availability

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

References

1. Abilio R, Coelho P G, Silva D A E A. Evaluating Named Entity Recognition: A comparative analysis of mono- and multilingual transformer models on a novel Brazilian corporate earnings call transcripts dataset[J]. Applied Soft Computing, 2024, 166:112158–112158. DOI: 10.1016 /J.ASOC.2024.112158.
2. Ali R A, Kamal H. Hybrid HHO–WHO Optimized Transformer-GRU Model for Advanced Failure Prediction in Industrial Machinery and Engines[J]. Sensors, 2026, 26(2):534–534. DOI:10.3390/S26020534.
3. Alsamraee A S, Khanna S. A hybrid transformer–TCN–GRU based model for thermal load forecasting of a large university campus[J]. Energy, 2026, 344:140114–140114. DOI:10.1016/J.ENERGY.2026.140114.
4. Bao W, Cao Y, Yang Y, et al. Long Short-Term Financial Time Series Forecasting Based on Residual Multiscale TCN Sparse Expert Network and Informer[J]. IEEE Transactions on Neural Networks and Learning Systems, 2025. DOI:10.1109/TNNLS.2025.3584369.
5. Chen Y, Min P, Chen H, et al. Parameter Identification of Photovoltaic Models Using an Enhanced INFO Algorithm[J]. CAAI Transactions on Intelligence Technology, 2025, 10(6):1844–1866. DOI:10.1049/CIT2.70065.
6. Chen Y, Ye N, Zhang W, et al. Meta-LSTR: Meta-Learning with Long Short-Term Transformer for futures volatility prediction[J]. Expert Systems With Applications, 2025, 265:125926–125926. DOI:10.1016/J.ESWA.2024.125926.
7. Ding J, Gao L, Shi S, et al. Intelligent optimization of finite element model parameters for large-span bridges based on MA-INFO algorithm[J]. Structures, 2024, 70:107617–107617. DOI:10.1016/J.ISTRUC.2024.107617.
8. Du H, Lv L, Wang H, et al. TGNS: A transformer-based graph neural network for stock trend forecasting[J]. Information Sciences, 2025, 720:122555–122555. DOI:10.1016/ J.INS.2025.122555.
9. Edeh E, Liang X, Cao C. Correction: Probing beyond: The impact of model size and prior informativeness on Bayesian SEM fit indices[J]. Behavior Research Methods, 2025, 57(5):148–148. DOI:10.3758/S13428-025-02672-9.
10. Fan J, Yu A G, Zhao M, et al. MCPT-CAF-BiGRU: A multi-scale CNN and ProbSparse-Masked Transformer model with cross-attention fusion and BiGRU for hourly wind speed forecasting[J]. Expert Systems With Applications, 2026, 307:131081–131081. DOI:10.1016 /J.ESWA.2025.131081.
11. Feng L, Sinchai A. Time Series Forecasting Using Transfer Learning with an Attention-Revamped Transformer: A Case Study of Financial Instruments[J]. Arabian Journal for Science and Engineering, 2025, 50(23):1–27. DOI:10.1007/S13369-025-10242-6.
12. Feng L, Sinchai A. Deep context-attentive transformer transfer learning for financial forecasting[J]. PeerJ Computer Science, 2025, 11:e2983. DOI:10.7717/PEERJ-CS.2983.
13. Fischer T, Sterling M, Lessmann S. Fx-spot predictions with state-of-the-art transformer and time embeddings[J]. Expert Systems With Applications, 2024, 249:123538–123538. DOI:10.1016/J.ESWA.2024.123538.
14. Gao S, Wang Y, Liu L, et al. Ultra-short-term wind power prediction method combining financial technology feature engineering and XGBoost algorithm[J]. Heliyon, 2023,

9(6):e16938–e16938. DOI:10.1016/J.HELIYON.2023.E16938.

15. Hadizadeh A, Tarokh J M, Ghazani M M. A novel transformer-based dual attention architecture for the prediction of financial time series[J]. Journal of King Saud University Computer and Information Sciences, 2025, 37(5):72–72. DOI:10.1007/S44443-025-00045-Y.
16. He L, Li H, Wang S, et al. Decoding Gender in Cough Sounds: A Transformer-Based Analysis[J]. The Laryngoscope, 2026. DOI:10.1002/LARY.70393.
17. He Y, Ning P, Song Y. Jump Volatility Forecasting for Crude Oil Futures Based on Complex Network and Hybrid CNN–Transformer Model[J]. Mathematics, 2026, 14(2):258–258. DOI:10.3390/MATH14020258.
18. Hiçyılmaz M. Bayesian optimization-enhanced vision transformer for damage detection in steel truss structures using continuous wavelet transform analysis[J]. Structures, 2026, 85:111059–111059. DOI:10.1016/J.ISTRUC.2026.111059.
19. Hu L, Yu Z, Liang M, et al. Optimization of cable connection layout of unattended AC/DC joint construction station based on INFO algorithm[J]. Discover Applied Sciences, 2025, 7(6):501–501. DOI:10.1007/S42452-025-07005-1.
20. Islam S, Haque R, Khan A M, et al. Ensemble transformer with post-hoc explanations for depression emotion and severity detection[J]. iScience, 2026, 29(2):114605–114605. DOI:10.1016/J.ISCI.2025.114605.
21. Izadi A M, Hajizadeh E. Time series prediction for cryptocurrency markets with transformer and parallel convolutional neural networks[J]. Applied Soft Computing, 2025, 177:113229–113229. DOI:10.1016/J.ASOC.2025.113229.
22. Jiang S, Zhang H, Chen Z, et al. Physics-based feature enhancement method and physics constraint Transformer model for multi-step tool wear and RUL prediction[J]. CIRP Journal of Manufacturing Science and Technology, 2026, 65:179–206. DOI:10.1016/J.CIRPJ.2026.01.001.
23. Jing J, Yongqing Z. Innovation in Financial Enterprise Risk Prediction Model: A Hybrid Deep Learning Technique Based on CNN-Transformer-WT[J]. Journal of Organizational and End User Computing, 2024, 36(1):1–26. DOI:10.4018/JOEUC.361650.
24. Junbo X, Junxing Z, Lin G, et al. Automated Three-Dimensional Sand Particle Shape Classification by PointConv-Transformer[J]. Journal of Computing in Civil Engineering, 2026, 40(3). DOI:10.1061/JCCEE5.CPENG-6928.
25. Li J, Wen X, Jia L, et al. A Transformer-based model fusing temporal dependence and variable correlation for short and medium-term electricity price forecasting[J]. Energy, 2025, 338:138799–138799. DOI:10.1016/J.ENERGY.2025.138799.
26. Li Z, Zhao K, Sun H, et al. Thermal errors in high-speed motorized spindle: An experimental study and INFO-GRU modeling predictions[J]. Case Studies in Thermal Engineering, 2025, 70:106085–106085. DOI:10.1016/J.CSITE.2025.106085.
27. Li X, Chen S, Qiao X, et al. Multi-perspective Learning Based on Transformer for Stock Price Trend[J]. International Journal of Computational Intelligence Systems, 2025, 18(1):44–44. DOI:10.1007/S44196-025-00768-W.
28. Li S, Xu S. Enhancing stock price prediction using GANs and transformer-based attention mechanisms[J]. Empirical Economics, 2024, 68(1):1–31. DOI:10.1007/S00181-024-02644-6.
29. Liu Z, Hu Y, Li Z. Transformer-based surrogate models for predicting flow velocities in downstream approach channels of multi-line ship locks[J]. Journal of Hydrology, 2026,

668:134994–134994. DOI:10.1016/J.JHYDROL.2026.134994.
30. Ma X, Wang L, Wu C, et al. Study on Prediction of Particle Migration at Interburden Boundaries in Ore-Drawing Process Based on Improved Transformer Model[J]. Processes, 2026, 14(2):366–366. DOI:10.3390/PR14020366.
31. Mao X, Yang S, Huang Z, et al. A novel INFO-SVR-based strategy integrated with terahertz metamaterial sensing for detecting trace thiabendazole residues[J]. Measurement, 2026, 258:119092–119092. DOI:10.1016/J.MEASUREMENT.2025.119092.
32. Mishra K A, Renganathan J, Gupta A. Volatility forecasting and assessing risk of financial markets using multi-transformer neural network based architecture[J]. Engineering Applications of Artificial Intelligence, 2024, 133:108223–108223. DOI:10.1016/J.ENGAPPAI.2024.108223.
33. Pang S, Chen J, Li D, et al. BP-Transformer hybrid model for time-series drag prediction of nanoparticle agglomerates[J]. Chemical Engineering Science, 2026, 323:123239–123239. DOI:10.1016/J.CES.2025.123239.
34. Ren X, Sun R, Jiang Z, et al. Time series is not enough: Financial Transformer Reinforcement Learning for portfolio management[J]. Neurocomputing, 2025, 647:130451–130451. DOI:10.1016/J.NEUCOM.2025.130451.
35. Shen C, Wu J. Research on credit risk of listed companies: a hybrid model based on TCN and DilateFormer[J]. Scientific Reports, 2025, 15(1):2599–2599. DOI:10.1038/S41598-025-86371-7.
36. Shelly C M, Hochstein D, Barda N, et al. Deep learning model design for ECG-based prediction of sex and ventricular dysfunction: A technical exploration of CNN and transformer architectures[J]. Heliyon, 2026, 12(2):e44527–e44527. DOI:10.1016/J.HELIYON.2026.E44527.
37. Shi T, Shide K. A comparative analysis of LSTM, GRU, and Transformer models for construction cost prediction with multidimensional feature integration[J]. Journal of Asian Architecture and Building Engineering, 2026, 25(1):634–649. DOI:10.1080/13467581.2025.2455034.
38. Singh N, Dubey K A, Khanna N N, et al. Attention/Transformer-based artificial intelligence models for carotid segmentation and intima media Thickness/Plaque area measurements in Japanese ultrasound scans[J]. Measurement, 2026, 267:120311–120311. DOI:10.1016/J.MEASUREMENT.2026.120311.
39. Song Y, Zhang Y, Huang J, et al. Volatility and value-at-risk forecasting using BERT and transformer models incorporating investors' textual sentiments[J]. Finance Research Letters, 2025, 85:108210–108210. DOI:10.1016/J.FRL.2025.108210.
40. Tian G, Huang T, Peng C, et al. BiMT-TCN: A cutting-edge hybrid model for enhanced stock price prediction[J]. Knowledge-Based Systems, 2025, 329:114263–114263. DOI:10.1016/J.KNOSYS.2025.114263.
41. Wang J, Fang X, He D, et al. Research on prediction method of rock uniaxial compressive strength based on interpretable INFO-Stacking model[J]. Engineering Analysis with Boundary Elements, 2025, 179:106390–106390. DOI:10.1016/J.ENGANABOUND.2025.106390.
42. Wang Z, Ni A, Tian Z, et al. Research on blockchain abnormal transaction detection technology combining CNN and transformer structure[J]. Computers and Electrical Engineering, 2024, 116:109194–109194. DOI:10.1016/J.COMPELECENG.2024.109194.

43. Xie Y, Liu Z J, Zeng L D, et al. A dynamic hybrid model of supercritical once-through boiler-turbine unit including recirculation mode and once-through mode[J]. Energy, 2024, 309:132941–132941. DOI:10.1016/J.ENERGY.2024.132941.
44. Xue C, Zhu J, Cao H, et al. INFO-RF-based fault diagnosis and analysis method for busbars[J]. Scientific Reports, 2025, 15(1):23502–23502. DOI:10.1038/S41598-025-07402-X.
45. Yan D, Li W, Xing E, et al. Integrative Transcriptomic and Bioinformatics Approaches Combined With Transformer Models Identify Key Gene Networks in Atherosclerosis[J]. FASEB Journal, 2026, 40(2):e71422. DOI:10.1096/FJ.202503988R.
46. Yousri D, V. S, Farag E H, et al. Cost-effective one-time configuration for bridge-linked thermoelectric generator array using INFO optimizer[J]. Journal of Cleaner Production, 2024, 467:142816–142816. DOI:10.1016/J.JCLEPRO.2024.142816.
47. Zhao C, Hu P, Yao X. Exploiting experience accumulation in stock price prediction with continual learning[J]. Applied Soft Computing, 2025, 170:112658–112658. DOI:10.1016/J.ASOC.2024.112658.
48. Zhao Z, Huang J, Xue Y. Comparative Analysis of ARIMA and Transformer-Based Models for Gold Price Forecasting[J]. Academic Journal of Management and Social Sciences, 2025, 13(3):764–773. DOI:10.54097/6CJKDF11.
49. Zhong Z, Wang J, Shen J, et al. High-precision bedrock motion inversion from surface records using decoder-only convolutional attention Transformer[J]. Soil Dynamics and Earthquake Engineering, 2026, 203:110113–110113. DOI:10.1016/J.SOILDYN.2026.110113.
50. Zhu M, Qi H, Ni S, et al. MSGformer: A Hybrid Multi-Scale Graph–Transformer Architecture for Unified Short- and Long-Term Financial Time Series Forecasting[J]. Electronics, 2025, 14(12):2457–2457. DOI:10.3390/ELECTRONICS14122457.