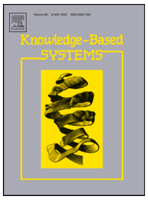
Knowledge-Based Systems 329 (2025) 114263

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys



BiMT-TCN: A cutting-edge hybrid model for enhanced stock price prediction

Guangyang Tian [a], Tingwen Huang [b], Chengyu Peng [c], Yin Yang [d], Shiping Wen [a,*]

[a] *Faculty of Engineering and Information Technology, Australian AI Institute, University of Technology Sydney, 2007, Sydney, Australia*
[b] *Faculty of Computer Science and Control Engineering, Shenzhen University of Advanced Technology, 518055, Shenzhen, China*
[c] *Laurentian University, 935 Ramsey Lake Rd, P3E 2C6, Sudbury, ON, Canada*
[d] *College of Science and Technology, Hamad Bin Khalifa University, 5855, Doha, Qatar*

ARTICLE INFO

Keywords:
BiMT-TCN model
Stock prediction
Time series analysis
BiLSTM
Modified transformer
TCN

ABSTRACT

In the face of the rapid evolution and escalating complexity of financial markets, precise stock price prediction has become a critical area of research for scholars and practitioners alike. Stock markets are subject to a vast array of influencing factors, both internal and external, which complicates prediction efforts. This study proposes BiMT-TCN, a novel model combining Bidirectional Long Short-Term Memory (BiLSTM), a modified Transformer, and Temporal Convolutional Network (TCN), aimed at enhancing the accuracy and stability in stock price prediction. BiLSTM facilitates the capture of bidirectional dependencies, which aids in decoding the intricate patterns within time-series data. The modified Transformer integrates global information, enhancing the model's capacity to manage long-range dependencies effectively. TCN, known for its parallel processing and proficiency in capturing deep historical patterns, further bolsters model stability and generalizability. Empirical evaluations on major indices such as SSE, HSI, and NASDAQ demonstrate that BiMT-TCN consistently outperforms state-of-the-art models, achieving R^2 scores of 0.9779, 0.9776, and 0.9969 respectively, along with significantly lower RMSE, MAE, and MAPE values. The implications of this work extend to practical investment decision-making, where improved forecast precision can enhance risk management, optimize trading strategies, and inform financial planning in volatile markets.

1. Introduction

Accurately forecasting stock prices remains one of the most fundamental yet challenging problems in financial research. The behavior of stock markets is shaped by a highly complex and dynamic interplay of economic indicators, geopolitical developments, investor sentiment, and various external shocks. These factors introduce significant nonlinearity, volatility, and uncertainty into financial time series, making it exceptionally difficult to model market dynamics with precision [1,2]. The need for reliable and precise prediction methods is driven by their potential to inform investment decisions, risk management, and financial planning, which are crucial for maximizing shareholder returns and stabilizing financial markets [3].

Over the past few decades, various prediction approaches have been developed and refined. Traditional methods, such as autoregressive integrated moving average (ARIMA), generalized autoregressive conditional heteroskedasticity (GARCH), and exponential smoothing, have been widely used in time-series prediction [4,5]. These statistical models excel in analyzing linear dependencies and capturing short-term trends within data. However, their performance can be limited when applied to the highly non-linear and non-stationary nature of stock price movements, which are common in real-world markets. The reliance on assumptions such as stationarity and normality constrains their predictive power and flexibility, especially in dynamic financial environments where sudden shifts and volatility are frequent [6].

To address the limitations of statistical models, machine learning techniques have gained traction. Methods like support vector machines (SVM), random forests, and gradient boosting machines provide improved modeling capabilities by accommodating non-linear patterns and complex relationships in the data [7]. However, these models often require significant manual feature engineering to extract informative attributes from the raw data. Additionally, traditional machine learning models may struggle with capturing intricate temporal dependencies inherent in financial time-series data, particularly in scenarios with long-term dependencies and context-specific shifts [8].

---

[*] Corresponding author.

E-mail addresses: guangyang.tian@student.uts.edu.au (G. Tian), huangtw2024@163.com (T. Huang), cpeng@laurentian.ca (C. Peng), yyang@hbku.edu.qa (Y. Yang), shiping.wen@uts.edu.au (S. Wen).

https://doi.org/10.1016/j.knosys.2025.114263
Received 25 April 2025; Received in revised form 12 July 2025; Accepted 11 August 2025

In recent years, the focus has shifted toward deep learning methods, which can automatically learn complex patterns from raw data, bypassing the need for extensive feature engineering. Deep learning architectures such as Long Short-Term Memory (LSTM) networks, Transformer models, and Temporal Convolutional Networks (TCNs) have been increasingly utilized in time-series prediction tasks [9]. Each model type has unique advantages. LSTM, particularly its bidirectional variant BiLSTM, is adept at capturing both past and future dependencies within sequences, which enhances context understanding. Transformers, initially developed for natural language processing, have demonstrated the ability to manage long-term dependencies and synthesize global context. However, they may face challenges in retaining precise sequence information. TCNs, known for their hierarchical dilated convolutional structure, excel at capturing deep sequential dependencies with stability and computational efficiency [10,11].

However, despite their individual advantages, these models also exhibit limitations when applied independently to stock prediction tasks. LSTM can be computationally inefficient over long sequences and may suffer from vanishing gradients. Transformers, while powerful in capturing global context, may lose local temporal precision. TCNs, though fast and stable, often require additional contextual enhancement to perform optimally in financial forecasting. These limitations motivate the need for a more robust architecture that can harness the complementary strengths of these models while mitigating their weaknesses [12].

To this end, we propose a novel hybrid model named BiMT-TCN, which integrates the strengths of BiLSTM, a modified Transformer, and TCN. The BiMT-TCN model leverages the bidirectional capabilities of BiLSTM for comprehensive sequence analysis, the enhanced long-term dependency handling of a modified Transformer for global context, and TCN's ability to learn deep sequential dependencies, which collectively improve the model's robustness and prediction accuracy. This approach is designed to outperform existing models in various market conditions and data sets, ensuring stable and accurate stock price forecasts.

1. Enhance the transformer model by integrating TCN to create a new architecture (MT-TCN) tailored for stock price prediction.
2. This study combines BiLSTM and MT-TCN, leveraging the strengths of BiLSTM, transformer, and TCN models.
3. Extensive experiments on major stock indices, including SSE, HSI, and NASDAQ, validate our model's effectiveness, achieving state-of-the-art results.
4. We conduct a thorough comparison between BiMT-TCN and existing baseline models, demonstrating its stability and accuracy across diverse time frames and market conditions.

2. Related works

Stock price prediction methods are generally classified into three categories: traditional statistical models, machine learning methods, and deep learning techniques. This section briefly reviews each category.

2.1. Traditional statistical models

Early approaches to prediction financial time series, such as stock prices, primarily relied on economic statistical models, which have since seen significant development. Common models include AR, MA, ARMA, and ARIMA [13].

Challa et al. [14] applied the ARIMA model to predict daily returns for the S&P BSE Sensex and S&P BSE IT indices, showing that ARIMA performed reasonably well in forecasting stock data. Ansari et al. [13] compared ARIMA, SutteARIMA, and Holt-Winters for predicting closing price trends in BRIC nations, with results favoring SutteARIMA and Holt-Winters over ARIMA. Rather than using ARMA alone, Hossain and Nasser [15] combined ARMA with GARCH to predict the Nikkei 225 and S&P 500. The findings indicated that ARMA-GARCH outperformed support vector regression (SVR) and backpropagation (BP) models in terms of directional accuracy, although it was less effective in deviation performance. In addition, Bouri et al. [16] employed a Bayesian Graphical Structural VAR model to examine whether global investor sentiment, measured by the CBOE VIX, can predict fear contagion in BRICS stock markets. The study found strong interdependence between global and regional sentiment dynamics, suggesting that global fear is a significant exogenous predictor of emerging market volatility. Furthermore, Bouri et al. [17] examined how local, regional, and global business cycles contribute to predicting stock market volatility in emerging economies. Based on an extended HAR-RV framework, their findings indicate that domestic output gaps have the most substantial impact in many cases. However, business cycles at the global and emerging market levels also exhibit notable predictive power, particularly for countries such as China and India.

Despite their usefulness in financial prediction, statistical models encounter limitations with complex data, especially in today's context. These models rely on stationary assumptions and struggle with long-term predictions, which restricts their applicability for intricate time series analysis.

2.2. Machine learning methods

With the rise of machine learning, diverse methods are increasingly used for time series prediction. Qi [18] was among the first to demonstrate that financial and economic variables, such as interest rates, inflation, and money supply, could be used to predict stock returns more effectively through nonlinear models. By employing a multilayer feedforward neural network, the study revealed that machine learning techniques significantly outperform traditional linear models in both in-sample and out-of-sample settings. Guo et al. [19] introduced an adaptive SVR model to predict stock prices, utilizing a dynamic mechanism and particle swarm optimization for automated hyper-parameter tuning, which proved effective in experiments. Ren et al. [20] applied a SVM combined with investor sentiment analysis to forecast the SSE 50 index price movements, showing that sentiment data could enhance SVM's predictive accuracy. To further optimize SVR hyper-parameters, Liu et al. [21] developed a hybrid EGWO-SVR model by integrating the grey wolf optimizer, which outperformed the standard SVR in stock selection tasks. Zhang and Lou [22] used BP neural networks for stock price pattern prediction, with simulations demonstrating a reasonable prediction capability. Tackling non-linear components in stock indices, Yang and Lin [23] employed empirical mode decomposition (EMD) to break down indices into sub-series for SVR input, achieving superior prediction accuracy over other models.

Compared to traditional statistical methods, machine learning models generally deliver better predictive performance, especially when combined with optimization or decomposition techniques. However, they still face limitations in long-term prediction accuracy.

2.3. Deep learning techniques

With the success of convolutional neural networks (CNNs) in the ImageNet competition, computer vision entered a deep learning era. Deep learning has since expanded into fields like natural language processing, agriculture, and finance [24].

Gülmez [25] utilized an LSTM network optimized with an artificial rabbits algorithm for stock price prediction, achieving better performance than models like artificial neural networks, naive LSTM, and LSTM optimized by genetic algorithms. Qi et al. [26] improved prediction accuracy by combining Gated Recurrent Unit (GRU) with CEEMDAN, using a wavelet transform to minimize high-frequency noise, which enhanced model effectiveness. Yao et al. [27] developed a MEDM-TCN hybrid model using multivariate time series, applying TCN to decomposed subsequences, which yielded favorable results. Cui et al. [28] introduced a multi-channel model combining VMD, CBAM, and BiLSTM, which proved more reliable than alternative approaches. Wang et al.

[29] proposed a multivariate deep learning method based on XGBoost, demonstrating a significant performance advantage. Jiang et al. [30] developed a dual-CNN model, converting stock data into 2-D images for predictive analysis, showing superior accuracy. To capture both long-term and short-term dependencies, Liu et al. [31] combined VMD with a self-attention LSTM and TCN, resulting in robust generalization. Lu and Xu [32] created an efficient RNN for time series, integrating an additional feature extraction module to improve stock prediction accuracy. Sivadasan et al. [33] explored various GRU and LSTM configurations, finding that a GRU model with OHLC and technical indicators yielded the best performance.

Despite deep learning's advantages over traditional machine learning for financial prediction, the complexities of stock markets continue to challenge models' predictive accuracy and robustness.

3. Methodology

3.1. Transformer

The Transformer, a foundational NLP model introduced by Google in 2017 [34], serves as the basis for popular models like BERT. It leverages a self-attention mechanism without relying on the sequential RNN structure, enabling parallelization and the incorporation of global context. Unlike recurrent networks, the Transformer is free from gradient vanishing issues and can access any previous word, regardless of its distance.

The Transformer architecture comprises an encoder and a decoder. The encoder, which consists of stacked layers, processes input data to create encoded representations, while the decoder generates output based on this encoding, as illustrated in Fig. 1. The multi-head self-attention mechanism in the encoder is crucial for capturing dependencies across different time frames, as it assigns varying attention to different parts of the sequence, thereby enriching feature information. Self-attention output is given by the follows:

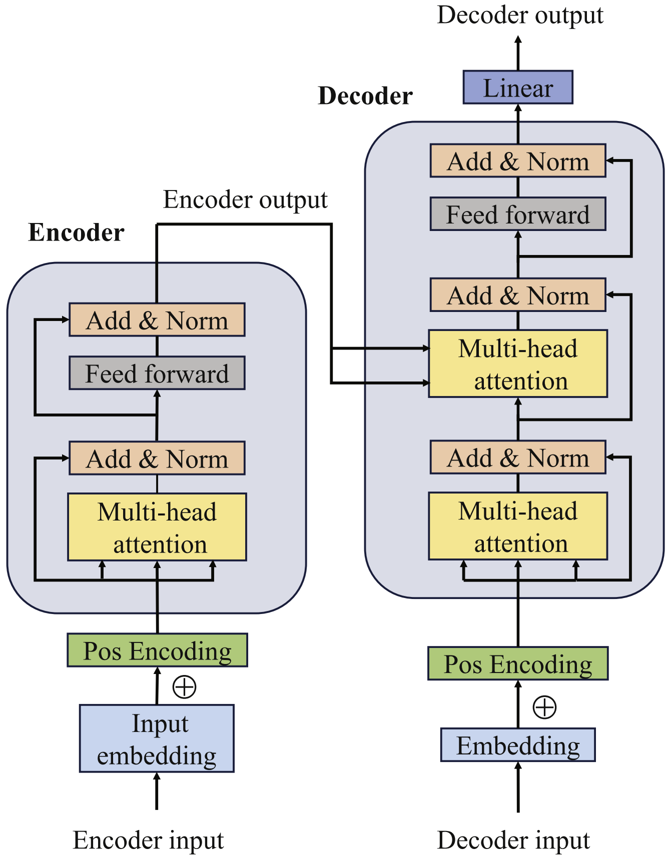


Fig. 1. The original framework of transformer.

$$\mathrm{Attention}(Q, K, V) = \mathrm{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{1}$$

Here, Q, K, and V represent the query, key, and value matrices, respectively. The term $\frac{QK^T}{\sqrt{d_k}}$ calculates the dot product of the queries and keys, scaled by $\sqrt{d_k}$, to manage gradient stability. Applying the softmax function normalizes the scores into a probability distribution, which is then used to weight the values V, enabling the model to focus on specific parts of the input sequence.

The Transformer's contextual understanding makes it particularly suited for time-series prediction. However, since it was initially designed for machine translation, it requires adaptation for time-series prediction tasks. To predict stock prices, we use the Transformer's encoder as the foundational model, as depicted in Fig. 3.

3.2. TCN

The TCN, introduced by Lea et al. [35] in 2016, is a recent approach for time series prediction. Its design centers around causal convolution, dilated convolution, and residual connections. Each layer in TCN employs a causal convolution with a unidirectional flow, illustrated in Fig. 2, to maintain temporal consistency by ensuring that the output at any time T is computed solely from elements occurring before T. This property eliminates information leakage from future time steps, which is critical for accurate time series modeling. Furthermore, TCN accommodates input sequences of arbitrary length, providing an output sequence of equivalent length, as represented in Eq. (2). This characteristic makes TCN flexible for a wide range of time-dependent applications, allowing it to effectively capture dependencies over extended historical sequences.

$$F(s) = \sum_{i=0}^{k-1} f(i) \cdot x_{s-di} \tag{2}$$

Dilated convolution, on the other hand, is employed to capture long-term dependencies without needing additional layers or pooling, thus expanding the receptive field effectively.

3.3. BiMT-TCN model

Modified Transformer-TCN (MT-TCN), an enhanced transformer model, is designed to improve stock price prediction. This section first explains the origin of the MT-TCN model architecture, followed by a detailed description of the entire BiMT-TCN framework.

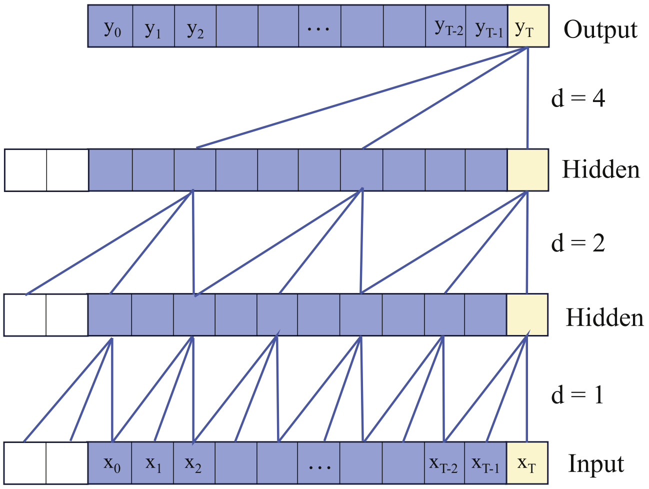


Fig. 2. The architecture of dilated casual convolution.

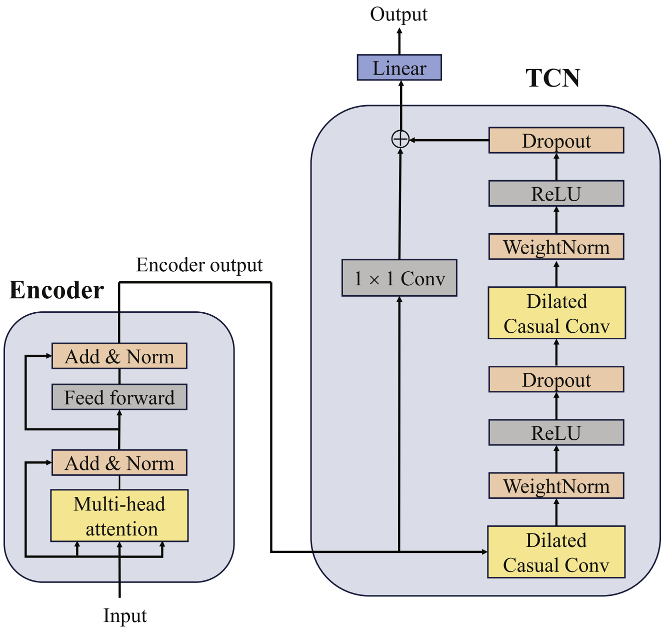


Fig. 3. A diagram combining transformer and TCN.

3.3.1. *The architecture of MT-TCN*

The transformer model is modified primarily by restructuring the decoder, as illustrated in Fig. 3.

- Remove the Input Embedding module. This module is needed for vectorizing language in NLP tasks but is unnecessary for stock price prediction.
- Move the Position Encoding module from MT-TCN to precede the BiLSTM layer.
- Replace the transformer decoder with a TCN layer followed by a fully connected layer.
- Limit the decoder's input to only the encoder's output, discarding other inputs.

Recent studies underscore TCN's efficacy in sequence data prediction. In 2023, Yang et al. [36] proposed an LSTM-TCN with self-attention for stock market prediction. The work affirms TCN's utility in sequence data analysis, supporting its inclusion in the enhanced transformer-based model.

The final MT-TCN architecture, shown in Fig. 3, includes a transformer encoder on the left and a TCN layer with a fully connected output on the right. The encoder comprises stacked layers, each with two sub-layers: the first features multi-head attention, normalization, and residual connections, while the second includes a feed-forward layer, normalization, and residual connections. The TCN component contains multiple residual blocks, each with two layers of dilated causal convolution, weight normalization, and dropout. The encoder output serves as the input to the initial residual block of the TCN.

3.3.2. *Overall framework*

This study builds on the transformer model by incorporating TCN to better adapt it for stock series prediction. While the transformer model excels in parallel processing and capturing global patterns, it has limitations in sequential information capture, affecting its predictive accuracy for stock data. TCN complements this by effectively capturing multi-level features with stable gradients, enabling parallel handling of time series data, which enhances both accuracy and training efficiency. By combining TCN and the transformer, the model leverages both components' strengths for more robust sequence prediction. Additionally, BiLSTM is integrated to reinforce sequential dependency capture, forming the hybrid network BiMT-TCN, as shown in Fig. 4.

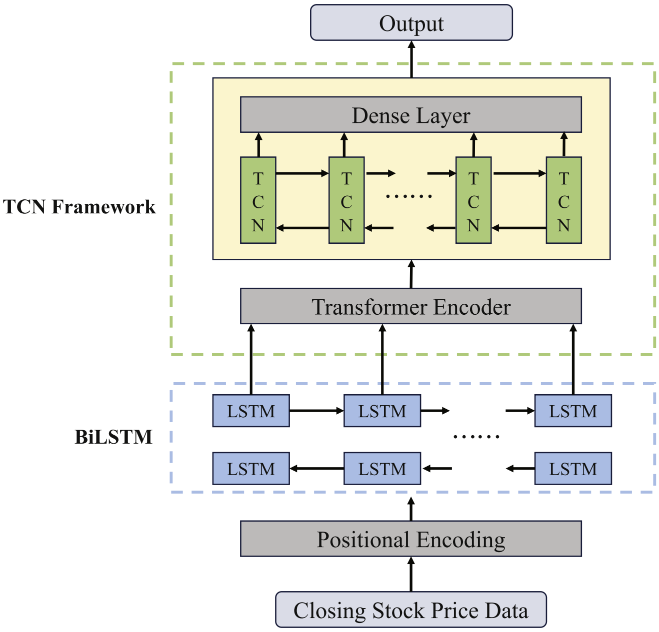


Fig. 4. Overview of the BiMT-TCN architecture. The model processes closing stock price data through positional encoding, BiLSTM, a Transformer encoder, and a TCN framework, followed by a dense layer to generate the final output.

BiLSTM's output at each step reflects both current inputs and prior context, making it effective for sequence information retention. However, due to the extensive time range in stock data, BiLSTM may experience gradient decay over long sequences, potentially losing significant features. The MT-TCN model's multi-head self-attention mechanism emphasizes critical features, filtering out less relevant information to improve prediction. Nonetheless, MT-TCN's sequential information capture is limited, as positional encoding only partially conveys order. To address this, BiLSTM precedes MT-TCN's self-attention to strengthen order dependency modeling, boosting prediction efficiency.

With deeper layers, rapid weight shifts can hinder performance on new data. TCN mitigates this by efficiently handling variable-length time series through convolutional layers that capture dependencies, while residual connections reduce network depth and parameter count, improving generalization and training efficiency. Thus, incorporating TCN to enhance the transformer and combining it with BiLSTM significantly elevates the model's predictive and computational performance.

To summarize, the integration of BiLSTM, the modified Transformer, and TCN within the BiMT-TCN is not a simple aggregation of models but a deliberate design to address the multi-scale nature of financial time series. Each component targets a specific modeling weakness of the others: BiLSTM captures immediate dependencies, the Transformer models long-range global trends, and TCN contributes temporal consistency and robustness through hierarchical convolution. Their combination enables the model to simultaneously handle short-term volatility and long-term structural patterns, making it highly applicable to real-world stock prediction scenarios where both localized signals and broader trends play important roles. This design choice enhances not only predictive accuracy but also adaptability across diverse market environments.

4. Experimental study

4.1. *Dataset introduction and preprocessing*

To evaluate the performance of the proposed BiMT-TCN model, we conducted experiments using three major stock indices: the Shanghai Stock Exchange Composite Index (SSE), the Hang Seng Index (HSI), and the NASDAQ Composite Index. The SSE, HSI, and NASDAQ datasets can all be obtained from Yahoo Finance. These indices were selected due to their representation of different financial markets, which allows us to assess the model's adaptability and robustness across diverse market conditions. The datasets span from January 2017 to August 2024, covering multiple distinct phases in global financial markets. These include sustained growth periods (2017–2018), trade tension-related corrections (2019), extreme volatility during the COVID-19 pandemic (2020–2021), and the post-pandemic recovery phase with rising inflation and interest rate shifts (2022–2024). This broad temporal coverage ensures that the model is exposed to a wide range of market conditions, including bullish and bearish trends, crisis-induced shocks, and macroeconomic uncertainty. By evaluating model performance across such conditions, we provide a more rigorous test of its generalization capability in real-world financial environments.

- SSE: The SSE represents the stock market performance in China and is comprised of all stocks that are listed on the Shanghai Stock Exchange. This dataset provides insight into one of the world's most dynamic and fast-evolving markets.
- HSI: The HSI is the primary index used to record and monitor the daily changes of the largest companies listed on the Hong Kong Stock Exchange. It serves as a significant indicator of the economic health of Hong Kong and is closely followed as a measure of global financial trends due to its international exposure.
- NASDAQ: The NASDAQ Composite includes over 3000 stocks listed on the NASDAQ stock exchange, with a heavy weighting in technology and growth-oriented companies. This index is a key indicator for understanding the performance of tech stocks and innovative sectors of the economy.

To standardize and prepare the data for model training, we applied the Z-score normalization method. This preprocessing step transforms the data such that the resulting distribution has a mean of 0 and a standard deviation of 1, facilitating better convergence during model training and improving model performance by handling different scales across datasets. The use of Z-score normalization helps in stabilizing training and preventing issues related to feature scaling.

4.2. Experiment setup and evaluation metrics

The experiments were conducted on a system equipped with a 12th Gen Intel(R) Core(TM) i9-12900H processor (base clock speed 2.50 GHz), 32 GB of RAM, and an Nvidia 3070TI GPU. This hardware configuration provided sufficient computational resources to support data processing, model training, and evaluation for the BiMT-TCN framework. The detailed training parameters of BiMT-TCN are shown in Table 1.

A fixed time window of 15 trading days was utilized to structure the input data. This time window size was chosen to capture sufficient sequential information that includes short-term trends while maintaining computational efficiency. The sliding window approach allows the model to leverage past 15-day information to predict future stock price movements effectively.

Table 1
Detailed training parameters of the BiMT-TCN model.

Parameters	Value
Time window length	15
Batch size	16
Learning Rate	0.0001
Hidden size of BiLSTM	128
BiLSTM layer number	2
Encoder head number	4
Encoder layer number	6
Dropout Rate of TCN	0.3
TCN layer neurons number	64
TCN hidden layer number	4
Kernel size of TCN layer	7
Activation function of TCN	ReLU

To objectively measure the performance of the proposed BiMT-TCN model, we utilize four commonly used evaluation metrics: Coefficient of Determination (R^2), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE). These metrics provide comprehensive insights into the prediction accuracy and error distribution.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{3}$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{4}$$

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{5}$$

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100\,\% \tag{6}$$

where y_i is the actual value, $\hat{y}_i$ is the predicted value, and $\bar{y}$ is the mean of the actual values.

The R^2 metric assesses how well predicted values align with actual values, indicating the proportion of variance explained by the model. An R^2 value closer to 1 reflects better predictive performance. RMSE measures the average magnitude of prediction errors, penalizing larger errors more heavily, making it useful for assessing error variance. A lower RMSE indicates higher predictive accuracy and consistency. MAE represents the mean absolute error, showing the average prediction error magnitude without overemphasizing larger errors. It is straightforward and indicates the average closeness of predictions to actual values. MAPE expresses prediction error as a percentage of actual values, making it suitable for comparing errors across different scales. Lower MAPE values reflect more accurate predictions.

4.3. Performance comparison

The experimental results, presented in Table 2, show that our proposed BiMT-TCN model consistently outperforms existing deep learning models, including LSTM [37], CNN-LSTM [38], BiLSTM [24], GRU [37], and Transformer [39], across all three stock indices: SSE, HSI, and NASDAQ. The analysis is based on four key metrics: R^2, RMSE, MAE, and MAPE. And qualitative visualizations of the prediction results for SSE, HSI, and NASDAQ are shown in Figs. 5–7, respectively.

Table 2
Performances on SSE, HSI, and NASDAQ datasets.

Stock Index	Models	R^2	RMSE	MAE	MAPE (%)
SSE	LSTM	0.9361	37.9622	29.7924	0.9637
	CNN-LSTM	0.9680	24.9706	19.7079	0.6403
	BiLSTM	0.9605	27.7605	22.2272	0.7234
	GRU	0.9498	31.3015	24.2603	0.7844
	Transformer	0.9702	22.4204	17.2996	0.5617
	Ours	**0.9779**	**20.7672**	**16.1948**	**0.5262**
HSI	LSTM	0.9407	266.9839	210.8292	1.1748
	CNN-LSTM	0.9596	220.7775	182.0191	1.0147
	BiLSTM	0.9507	230.6107	188.7139	1.0566
	GRU	0.9463	247.0029	196.4868	1.0904
	Transformer	0.9659	209.0594	170.5293	0.9503
	Ours	**0.9776**	**201.5163**	**161.7342**	**0.9019**
NASDAQ	LSTM	0.9773	189.3892	147.0618	1.0225
	CNN-LSTM	0.9857	139.7472	111.0046	0.7703
	BiLSTM	0.9837	145.4637	117.3653	0.8141
	GRU	0.9819	158.4135	125.8831	0.8794
	Transformer	0.9938	124.3545	99.7428	0.6924
	Ours	**0.9969**	**116.8035**	**94.2259**	**0.6555**

For the SSE dataset, BiMT-TCN achieves the highest R^2 value at 0.9779, indicating it captures the variance in the data better than other models. The next closest model is the Transformer with an R^2 of 0.9702. In terms of RMSE, BiMT-TCN records 20.7672, outperforming the Transformer at 22.4204, while LSTM has a much higher RMSE of 37.9622, highlighting its limitations in this dataset. The MAE for BiMT-TCN is also the lowest at 16.1948, compared to 17.2996 for the Transformer and 19.7079 for CNN-LSTM. BiMT-TCN also achieves the lowest MAPE at 0.5262 %, while LSTM has a much higher MAPE of 0.9637 %. These results suggest that BiMT-TCN provides substantial improvements in accuracy and stability over all other models on the SSE dataset.

On the HSI dataset, BiMT-TCN once again outperforms other models with an R^2 of 0.9776, compared to 0.9659 for the Transformer. The RMSE of BiMT-TCN is 201.5163, lower than the Transformer at 209.0594 and CNN-LSTM at 220.7775, while LSTM shows a much higher RMSE of 266.9839, indicating poor predictive performance. The MAE of BiMT-TCN is the lowest at 161.7342, compared to 170.5293 for the Transformer and 182.0191 for CNN-LSTM. In terms of MAPE, BiMT-TCN achieves 0.9019 %, better than the Transformer's 0.9503 %,

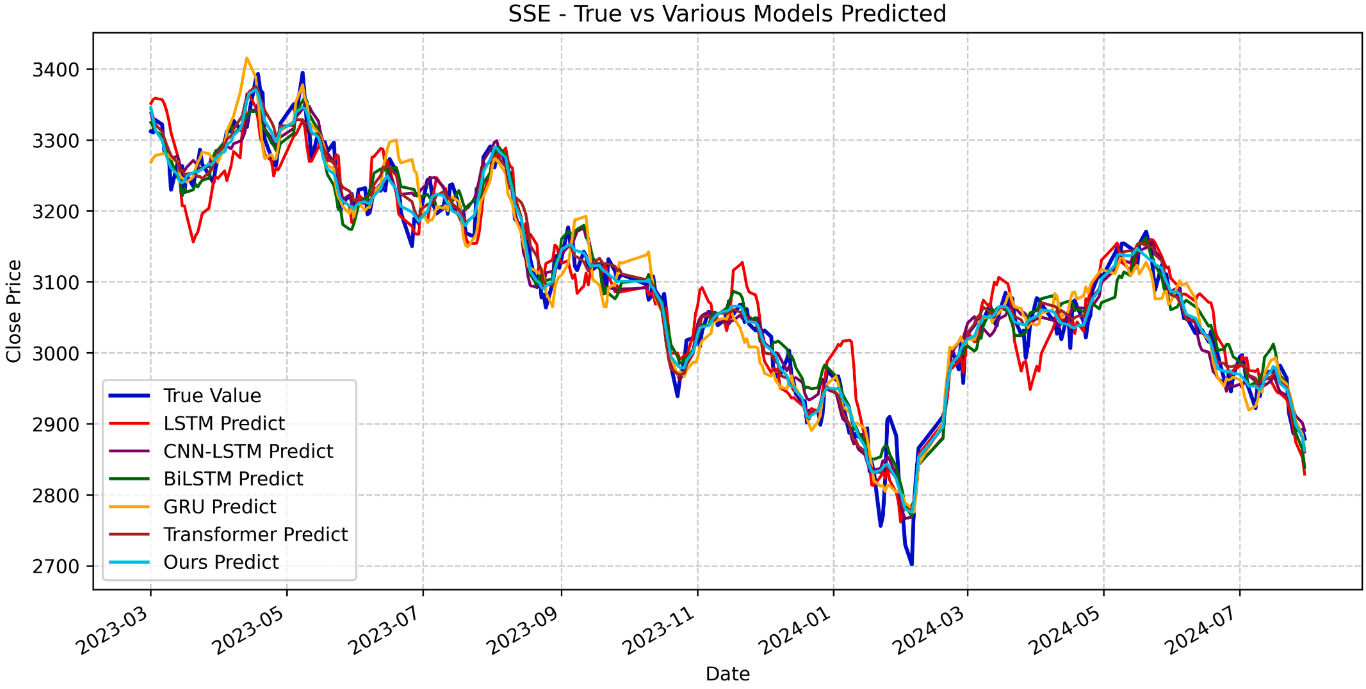


Fig. 5. Comparison of prediction results of different models in SSE dataset.

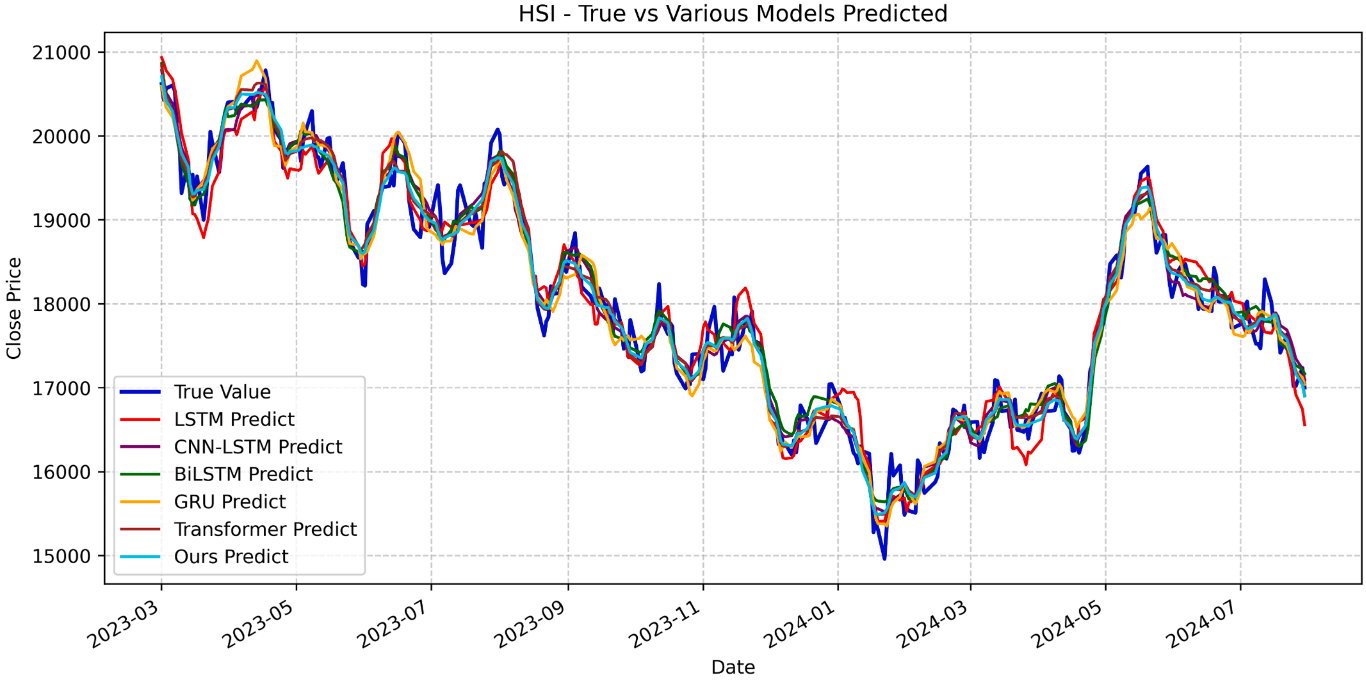


Fig. 6. Comparison of prediction results of different models in HSI dataset.

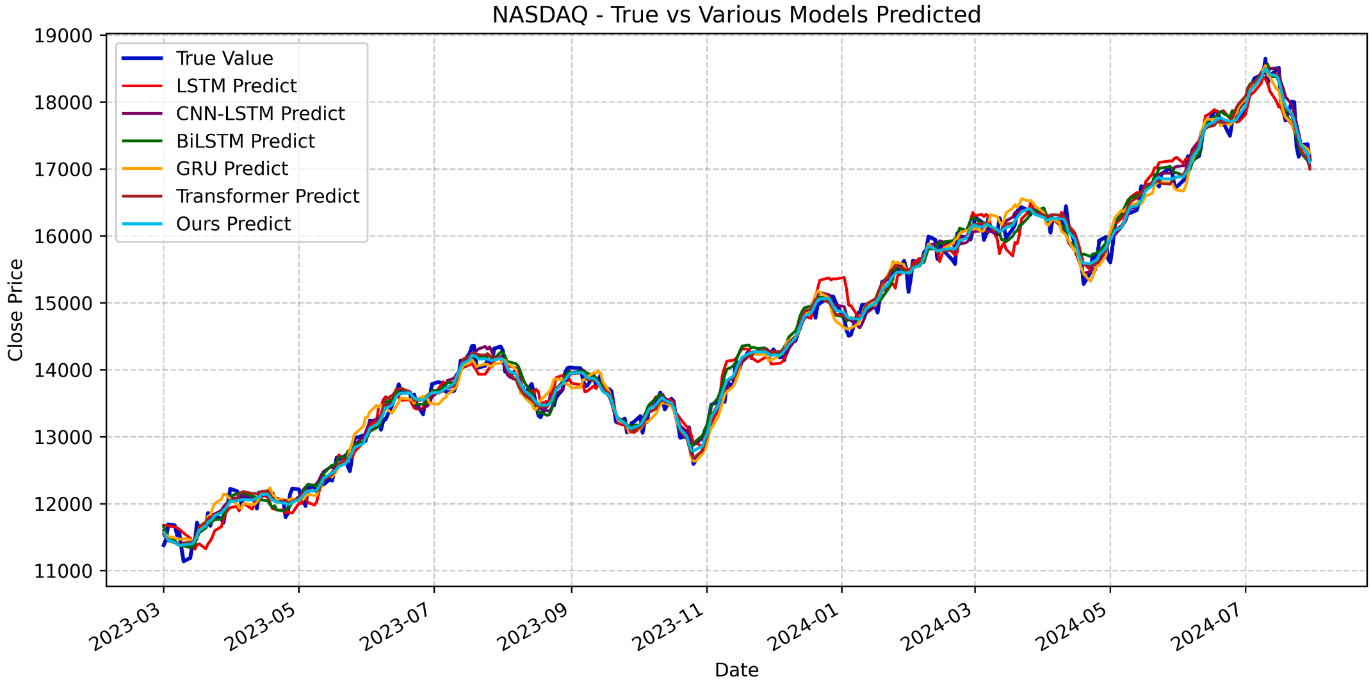


Fig. 7. Comparison of prediction results of different models in NASDAQ dataset.

demonstrating its accuracy. These results confirm that BiMT-TCN achieves superior accuracy and error minimization on the HSI dataset.

For the NASDAQ dataset, BiMT-TCN achieves an outstanding R^2 of 0.9969, higher than the Transformer's 0.9938, while LSTM lags behind with 0.9773. BiMT-TCN records the lowest RMSE at 116.8035, compared to the Transformer's 124.3545, while LSTM has a much higher RMSE of 189.3892. In terms of MAE, BiMT-TCN achieves the lowest value at 94.2259, followed by the Transformer at 99.7428, with CNN-LSTM at 111.0046. BiMT-TCN's MAPE is the lowest at 0.6555 %, with the Transformer close behind at 0.6924 %, while LSTM's MAPE is higher at 1.0225 %. This confirms the high predictive power and robustness of BiMT-TCN on the NASDAQ dataset.

These empirical findings also offer important implications from an economic perspective. The fact that BiMT-TCN consistently outperforms benchmark models across different indices suggests the existence of exploitable patterns in historical price data, which challenges the strict weak-form Efficient Market Hypothesis (EMH). In particular, the model's superior performance in more volatile and information-sensitive markets such as NASDAQ and HSI supports the view that these markets may exhibit short-term inefficiencies due to delayed information absorption or behavioral anomalies. This aligns with prior studies suggesting that technical signals and time-based patterns retain predictive value, especially during periods of heightened uncertainty.

4.4. Ablation study and analysis

To thoroughly evaluate the contribution of each component in the proposed BiMT-TCN model, we designed an ablation study. The goal of this study is to isolate the impact of each module—BiLSTM, the modified Transformer, and TCN on the overall performance of the model. This analysis allows us to determine the necessity and effectiveness of these components in enhancing predictive accuracy and robustness. We conduct the ablation study by systematically removing each component from the BiMT-TCN model and observing the changes in performance on the dataset SSE. The definitions of each model are shown in Table 3. The results of the ablation study are shown in Table 4.

The model using TCN only demonstrates the lowest R^2 (0.9405), with relatively high RMSE (35.5632), MAE (28.1274), and MAPE (0.8850 %), indicating that TCN alone is insufficient to achieve high accuracy in the stock prediction task. The model using only BiLSTM performs slightly better, with all evaluation metrics surpassing those of the TCN only model, suggesting that BiLSTM offers a slight advantage over TCN alone, though it remains inadequate for this prediction task. Additionally, the model using only the modified Transformer shows a significant performance improvement, reaching an R^2 of 0.9570, which indicates the strong capability of the modified Transformer in handling stock prediction data. However, this model's performance still falls short of the models that combine multiple components.

The MT-TCN (without BiLSTM) and BiLSTM-Transformer (without TCN) configurations both exhibit significant performance improvements, achieving R^2 values of 0.9767 and 0.9735, respectively. Comparatively, the MT-TCN model has lower RMSE (21.1234) and MAE (17.0567), suggesting that the combination of Transformer and TCN is better at capturing key features than the combination of BiLSTM and Transformer. Meanwhile, the BiLSTM-TCN (without Transformer)

Table 3
Ablation study configuration.

Configuration	Description
TCN Only	A standalone TCN model without BiLSTM and the modified Transformer to assess its individual contribution.
Modified Transformer Only	A standalone modified Transformer model without BiLSTM and TCN to analyze its capability in isolation.
BiLSTM Only	A standalone BiLSTM model without the modified Transformer and TCN to evaluate baseline performance.
MT-TCN (without BiLSTM)	A model that includes only the modified Transformer and TCN, without BiLSTM.
BiLSTM-Transformer (without TCN)	A model that retains only BiLSTM and modified Transformer, with TCN removed.
BiLSTM-TCN (without Transformer)	A reduced model using only BiLSTM and TCN, excluding the modified Transformer.
BiMT-TCN Model	The complete model incorporating BiLSTM, modified Transformer, and TCN.

Table 4
Ablation study performance on SSE.

Stock Index	Models	R^2	RMSE	MAE	MAPE (%)
SSE	TCN Only	0.9405	35.5632	28.1274	0.8850
	Modified Transformer Only	0.9570	29.4185	23.7593	0.7361
	BiLSTM Only	0.9423	33.2156	26.1048	0.8125
	MT-TCN (without BiLSTM)	0.9767	21.1234	17.0567	0.5421
	BiLSTM-Transformer (without TCN)	0.9735	23.1743	18.1984	0.5890
	BiLSTM-TCN (without Transformer)	0.9668	25.8497	20.3127	0.6542
	BiMT-TCN Model	**0.9779**	**20.7672**	**16.1948**	**0.5262**

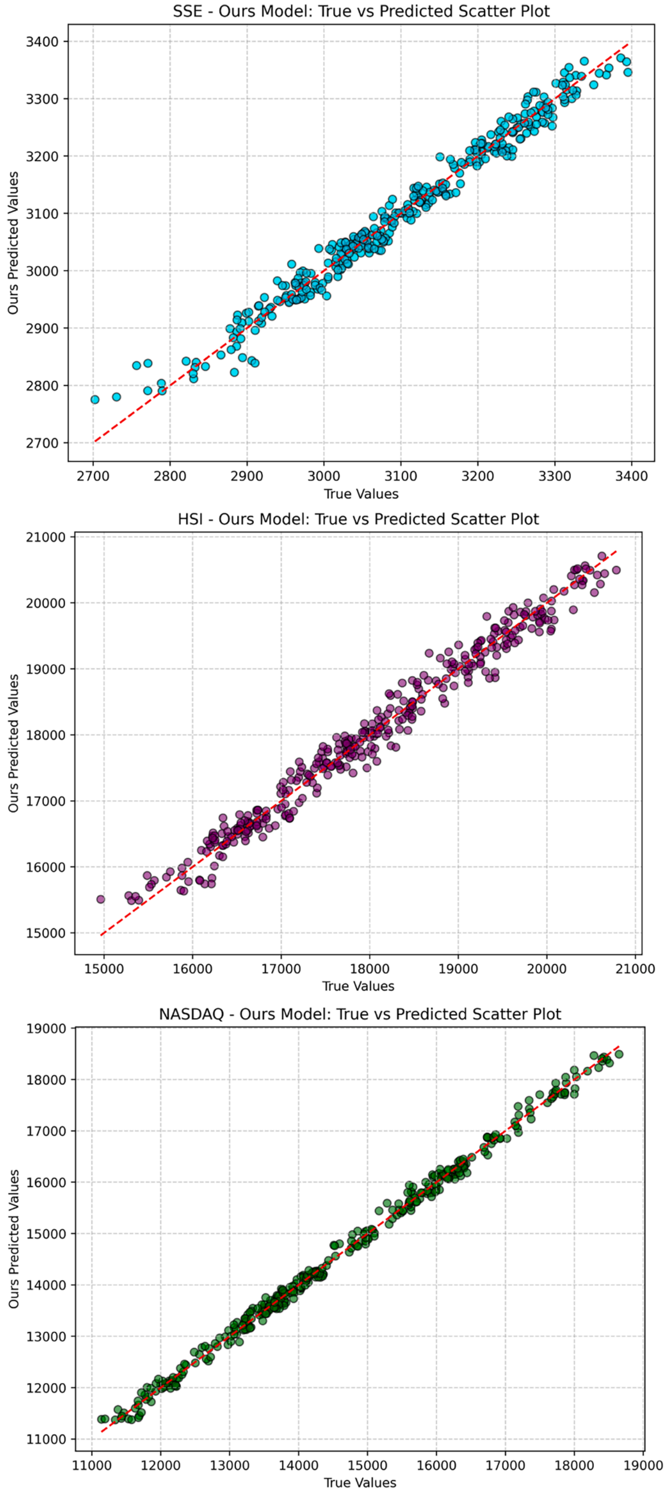


Fig. 8. True vs predicted scatter plots for different datesets.

model, while reaching an R^2 of 0.9668, has relatively higher RMSE, MAE, and MAPE, highlighting the importance of the Transformer component.

Finally, the BiMT-TCN model, which integrates BiLSTM, the modified Transformer, and TCN, achieves the best overall performance, with an R^2 of 0.9779, RMSE of 20.7672, MAE of 16.1948, and MAPE of 0.5262 %. These results clearly indicate that the comprehensive integration of BiLSTM, the modified Transformer, and TCN components collaboratively enhances predictive accuracy and reduces error, significantly outperforming all other model variants.

4.5. *Qualitative analysis*

Fig. 8 provides an overall view of our model's predictive performance on the SSE, HSI, and NASDAQ datasets. From the scatter plots, it is evident that the predicted values closely align with the true values, with data points densely distributed around the ideal red dashed line. This alignment demonstrates the model's high accuracy and stability across different stock index datasets, showing its ability to capture the complex trends in price movements. Although there are minor deviations and a few outliers in the SSE and HSI datasets, the model maintains a robust performance, indicating good generalization and reliability on these datasets. Overall, this model not only exhibits high adaptability under varying market conditions but also provides reliable predictions in volatile financial data, further confirming its potential as a valuable tool for financial market prediction.

5. Conclusion

In this study, we introduce a novel hybrid model, BiMT-TCN, which integrates BiLSTM, a modified Transformer, and TCN, aiming to improve the accuracy and stability of stock price predictions across different datasets. Through extensive experiments on the SSE, HSI, and NASDAQ indices, our model consistently outperformed other baseline models, demonstrating its superior predictive capability. The results from both quantitative metrics and scatter plot visualizations confirm that BiMT-TCN effectively captures complex temporal dependencies, showing high adaptability and robustness in different market environments. The ablation study further highlights the unique contributions of each component—BiLSTM's ability to capture bidirectional dependencies, the modified Transformer's strength in modeling long-term relationships, and TCN's role in enhancing stability and generalization. The complete BiMT-TCN model achieves optimal performance by leveraging the complementary strengths of these components, surpassing all partial configurations.

In practice, BiMT-TCN can assist investors in making better-informed trading decisions and timing strategies under changing market conditions. It may also provide regulators with tools for monitoring irregular market behavior and price volatility. While our current model focuses on daily closing prices, future work could explore additional input features or apply the model to alternative financial assets. One promising direction is the application of BiMT-TCN in cryptocurrency markets, where high volatility and trading frequency present unique modeling challenges. Prior research, including studies by Gerritsen et al. [40] and Alonso-Monsalve et al. [41] , has shown that deep learning and technical indicators can be effective in crypto trend forecasting. Extending our model to this domain may yield valuable insights for both trading and academic research.

CRediT authorship contribution statement

Guangyang Tian: Writing – original draft, Visualization, Validation, Software, Methodology; **Tingwen Huang:** Writing – review & editing; **Chengyu Peng:** Visualization, Software; **Yin Yang:** Investigation; **Shiping Wen:** Writing – review & editing, Supervision.

Data availability

No data was used for the research described in the article.

Declaration of competing interest

There is no conflict of interest in this paper.

References

[1] B.P. Ghosh, M.S. Bhuiyan, D. Das, T.N. Nguyen, R.M. Jewel, M.T. Mia, D.M. Cao, R. Shahid, Deep learning in stock market forecasting: comparative analysis of neural network architectures across NSE and NYSE, J. Comput. Sci. Technol. Stud. 6 (1) (2024) 68–75.

[2] G. Sonkavde, D.S. Dharrao, A.M. Bongale, S.T. Deokate, D. Doreswamy, S.K. Bhat, Forecasting stock market prices using machine learning and deep learning models: a systematic review, performance analysis and discussion of implications, Int. J. Financ. Stud. 11 (3) (2023) 94.

[3] U.R. Saxena, P. Sharma, G. Gupta, Comprehensive study of machine learning algorithms for stock market prediction during COVID-19, J. Comput., Mech. Manage. 2 (1) (2023) 01–07.

[4] A. Pagliaro, Forecasting significant stock market price changes using machine learning: extra trees classifier leads, Electronics 12 (21) (2023) 4551.

[5] M. Ali, D.M. Khan, H.M. Alshanbari, A.A.-A.H. El-Bagoury, Prediction of complex stock market data using an improved hybrid EMD-LSTM model, Appl. Sci. 13 (3) (2023) 1429.

[6] C.M. Liapis, A. Karanikola, S. Kotsiantis, Investigating deep stock market forecasting with sentiment analysis, Entropy 25 (2) (2023) 219.

[7] Z. Li, H. Yu, J. Xu, J. Liu, Y. Mo, Stock market analysis and prediction using LSTM: a case study on technology stocks, Innovations Appl. Eng. Technol. 2 (1) (2023) 1–6.

[8] S. Pourroostaei Ardakani, N. Du, C. Lin, J.-C. Yang, Z. Bi, L. Chen, A federated learning-enabled predictive analysis to forecast stock market trends, J. Ambient Intell. Humaniz. Comput. 14 (4) (2023) 4529–4535.

[9] A. Kurani, P. Doshi, A. Vakharia, M. Shah, A comprehensive comparative study of artificial neural network (ANN) and support vector machines (SVM) on stock forecasting, Anal. Data Sci. 10 (1) (2023) 183–208.

[10] H. Oukhouya, K. El Himdi, Comparing machine learning methods—SVR, XGBoost, LSTM, and MLP—for forecasting the Moroccan stock market, in: Computer Sciences & Mathematics Forum, volume 7, MDPI, 2023, p. 39.

[11] D. Sheth, M. Shah, Predicting stock market using machine learning: best and accurate way to know future stock prices, Int. J. Syst. Assur. Eng. Manage. 14 (1) (2023) 1–18.

[12] H. Song, H. Choi, Forecasting stock market indices using the recurrent neural network based hybrid models: CNN-LSTM, GRU-CNN, and ensemble models, Appl. Sci. 13 (7) (2023) 4644.

[13] A. Ansari Saleh, S. Pawan Kumar, T. Nguyen Van, T. Nguyen Viet, H. Vo Minh, Prediction of BRIC stock price using ARIMA, sutteARIMA, and holt-winters, Comput. Mater. Continua 70 (1) (2022) 523–534.

[14] M.L. Challa, V. Malepati, S.N.R. Kolusu, S&P BSE sensex and S&P BSE it return forecasting using ARIMA, Financ. Innovation 6 (1) (2020) 47.

[15] A. Hossain, M. Nasser, Comparison of the finite mixture of ARMA-GARCH, back propagation neural networks and support-vector machines in forecasting financial returns, J. Appl. Stat. 38 (3) (2011) 533–551.

[16] E. Bouri, R. Gupta, S. Hosseini, C.K.M. Lau, Does global fear predict fear in BRICS stock markets? Evidence from a Bayesian graphical structural VAR model, Emerging Mark. Rev. 34 (2018) 124–142.

[17] E. Bouri, R. Demirer, R. Gupta, X. Sun, The predictability of stock market volatility in emerging economies: relative roles of local, regional, and global business cycles, J. Forecast. 39 (6) (2020) 957–965.

[18] M. Qi, Nonlinear predictability of stock returns using financial and economic variables, J. Bus. Econ. Stat. 17 (4) (1999) 419–429.

[19] Y. Guo, S. Han, C. Shen, Y. Li, X. Yin, Y. Bai, An adaptive SVR for high-frequency stock price forecasting, IEEE Access 6 (2018) 11397–11404.

[20] R. Ren, D.D. Wu, T. Liu, Forecasting stock market movement direction using sentiment analysis and support vector machine, IEEE Syst. J. 13 (1) (2018) 760–770.

[21] M. Liu, K. Luo, J. Zhang, S. Chen, A stock selection algorithm hybridizing grey wolf optimizer and support vector regression, Expert Syst. Appl. 179 (2021) 115078.

[22] D. Zhang, S. Lou, The application research of neural network and BP algorithm in stock price pattern classification and prediction, Future Gener. Comput. Syst. 115 (2021) 872–879.

[23] H.-L. Yang, H.-C. Lin, An integrated model combined ARIMA, EMD with SVR for stock indices forecasting, Int. J. Artif. Intell. Tools 25 (02) (2016) 1650005.

[24] A. Kanwal, M.F. Lau, S.P.H. Ng, K.Y. Sim, S. Chandrasekaran, BicuDNNLSTM-1dCNN—a hybrid deep learning-based predictive model for stock price prediction, Expert Syst. Appl. 202 (2022) 117123.

[25] B. Gülmez, Stock price prediction with optimized deep LSTM network with artificial rabbits optimization algorithm, Expert Syst. Appl. 227 (2023) 120346.

[26] C. Qi, J. Ren, J. Su, GRU neural network based on ceemdan–wavelet for stock price prediction, Appl. Sci. 13 (12) (2023) 7104.

[27] Y. Yao, Z.-y. Zhang, Y. Zhao, Stock index forecasting based on multivariate empirical mode decomposition and temporal convolutional networks, Appl. Soft Comput. 142 (2023) 110356.

[28] C. Cui, P. Wang, Y. Li, Y. Zhang, McVCsB: a new hybrid deep learning network for stock index prediction, Expert Syst. Appl. (2023) 120902.

[29] J. Wang, Q. Cheng, Y. Dong, An XGBoost-based multivariate deep learning framework for stock index futures price forecasting, Kybernetes 52 (10) (2023) 4158–4177.

[30] M. Jiang, W. Chen, H. Xu, Y. Liu, A novel interval dual convolutional neural network method for interval-valued stock price prediction, Pattern Recognit. 145 (2024) 109920.

[31] Y. Liu, S. Huang, X. Tian, F. Zhang, F. Zhao, C. Zhang, A stock series prediction model based on variational mode decomposition and dual-channel attention network, Expert Syst. Appl. 238 (2024) 121708.

[32] M. Lu, X. Xu, TRNN: an efficient time-series recurrent neural network for stock price prediction, Inf. Sci. 657 (2024) 119951.

[33] E.T. Sivadasan, N. Mohana Sundaram, R. Santhosh, Stock market forecasting using deep learning with long short-term memory and gated recurrent unit, Soft comput. 28 (4) (2024) 3267–3282.

[34] A. Vaswani, Attention is all you need, Adv. Neural Inf. Process. Syst. (2017) 30, 5998–6008.

[35] C. Lea, R. Vidal, A. Reiter, G.D. Hager, Temporal convolutional networks: a unified approach to action segmentation, in: Computer Vision–ECCV 2016 Workshops: Amsterdam, The Netherlands, October 8–10 and 15-16, 2016, Proceedings, Part III 14, Springer, 2016, pp. 47–54.

[36] Y. Zhiyong, Y. Yuxi, Z. Yu, Application of BiLSTM-SA-TCN time series model in stock forecasting, J. Nanjing Inf. Eng. Univ. 6 (2023) 1–12.

[37] Y. Gao, R. Wang, E. Zhou, Stock prediction based on optimized LSTM and GRU models, Sci. Program. 2021 (1) (2021) 4055281.

[38] W. Lu, J. Li, Y. Li, A. Sun, J. Wang, A CNN-LSTM-based model to forecast stock prices, Complexity 2020 (1) (2020) 6622927.

[39] T. Muhammad, A.B. Aftab, M. Ibrahim, M.M. Ahsan, M.M. Muhu, S.I. Khan, M.S. Alam, Transformer-based deep learning model for stock price prediction: a case study on bangladesh stock market, Int. J. Comput. Intell. Appl. 22 (03) (2023) 2350013.

[40] D.F. Gerritsen, E. Bouri, E. Ramezanifar, D. Roubaud, The profitability of technical trading rules in the bitcoin market, Finance Res. Lett. 34 (2020) 101263.

[41] S. Alonso-Monsalve, A.L. Suárez-Cetrulo, A. Cervantes, D. Quintana, Convolution on neural networks for high-frequency trend prediction of cryptocurrency exchange rates using technical indicators, Expert Syst. Appl. 149 (2020) 113250.