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.

ST Stock Price Forecasting with the PSO-KAN-Transformer Hybrid Model

Shuaiqi Peng
Qinghai Minzu University
Yinkui Li
Qinghai Minzu University
Bin Zhang
binzhang0717@foxmail.com

Qinghai Minzu University

Research Article

Keywords: KAN, Transformer, PSO-KAN-Transformer, Stock price prediction

Posted Date: March 11th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8927196/v1

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

Additional Declarations: No competing interests reported.

ST Stock Price Forecasting with the PSO-KAN-Transformer Hybrid Model

Shuaiqi Peng, Yinkui Li†, Bin Zhang†

School of Mathematics and Statistics, Qinghai Minzu University,Xining, 810007, Qinghai, China.

Contributing authors: binzhang0717@foxmail.com; lyk463@163.com; pengshuaiqi21@163.com;
†These authors contributed equally to this work.

Abstract

The rapid advancement of big data and Internet of Things (IoT) technologies has expanded the application scope of high-dimensional time series data. To address the limited interpretability of Multilayer Perceptron models in ST stock price prediction, this study enhances the internal structure of the Transformer by replacing the MLP layer with a KAN layer, thereby improving the interpretability of the Transformer in stock price forecasting. For experimental analysis, empirical research is conducted on a dataset including the SSE 50 Index, SZSE 100 Index, and two ST stocks. Experimental results show that the PSO-KAN-Transformer hybrid model achieves superior performance in time series prediction, reducing the mean square error by an average of 18.85% compared with the original Transformer. This indicates that the proposed model has improved accuracy and interpretability.

Keywords: KAN;Transformer;PSO-KAN-Transformer;Stock price prediction

1 Introduction

The stock market serves as an indicator of economic fluctuations, and the prediction of stock closing prices is a matter of significant interest to both investors and academicians alike[1]. ST stock is a publicly traded equity in the securities market that entails significant investment risk. They have garnered recognition for providing appealing investment options and reversing investment returns to risk-aggressive

investors. The non-linearity, non-stationarity, and weak correlation among historical trade data series of ST stock render the precise prediction of its closing price exceedingly complex and dubious.

Over the past few years, a range of machine learning models like BP, CNN, LSTM, and BI-LSTM have found widespread application in predicting stock prices[2]. Technologies like dimensionality reduction, noise reduction, and hyperparameter optimization have significantly improved the prediction performance[3]. Yet, traditional machine learning models suffer from two limitations when applied to ST stocks: First, the nonlinear and event-driven characteristics of stock price sequences make it difficult for shallow BP and CNN models to capture long-range temporal dependencies[4]. Second, although LSTM-based models alleviate the problem of gradient vanishing through gating mechanisms[5], they are insufficient in depicting "cross-sequence" interactions and local sudden patterns, resulting in high volatility and weak robustness in the prediction of ST stock sequences.

Recently, Transformer models based on the self-attention mechanism have leveraged their advantages of "global dependency + parallel computing" to refresh the Mean Squared Error (MSE) records on various stock price datasets, thus becoming a new benchmark for ST stock prediction[6]. However, the "multi-head self-attention" of Transformers has a complex distribution and a high-dimensional parameter space, making it difficult for investors and regulators to intuitively understand the decision-making logic[7]. The lack of interpretability has become a core obstacle for Transformers to serve as prediction models[8]. Building on this, domestic and foreign researchers have carried out improvements on Transformer, such as, Jiayi Ren and Aiyin Wang[9]introduced a Stockformer model designed for stock price prediction utilizing the Transformer framework, which analyzes the temporal dependencies between stock prices and associated factors through a multiple causal attention mechanism, demonstrating superior predictive accuracy compared to a singular model. Shengqi Ren and Wei Song[10] employed a genetic algorithm to extract features from financial data and developed the GGInformer model for predicting financial product prices by integrating a GRU network with the Informer model. The results indicated a reduction in MSE on the dataset, thereby enhancing prediction accuracy. We observed that current Transformer-based stock price prediction models have strong generalization capabilities and prediction accuracy; yet, they possess certain limitations in modeling volatile stock data. For instance, challenges arise when handling non-linear, low signal-to-noise ratio stock sequence data and high-dimensional data, and its interpretability is very limited[11].
This research proposes the KAN-Transformer model, which enhances the internal structure of the Transformer model through the KAN network, and employs the Particle Swarm Optimization to optimize parameter extraction. The outcomes of the experiment indicate that the interpretability of the enhanced model has been augmented to a certain degree[12]. The experiments demonstrate that the interpretability of the enhanced model has been augmented to a certain degree. The PSO-KAN-Transformer model demonstrates superior accuracy and efficacy relative to

the Transformer model by utilizing the Index of SSE 50, SZSE 100, and two ST stocks from SSE and SZSE as datasets for experiment.

2 Materials and Methods

2.1 KAN

KAN (Kolmogorov-Arnold Networks) exhibit superior accuracy and enhanced interpretability compared to MLP (Multilayer Perceptron) because to their unique activation functions and parameterization techniques[13]. This research presents the KAN-Transformer model, which employs B-splines as the activation function for the KAN network, demonstrating that the activation function is positioned on the "edge" and is subject to learning.

$$\begin{aligned} \phi(x) &= \sum_{j=0}^{N_1} c_j b_j(x) \\ \phi'(x) &= \sum_{j=0}^{N_2} c_j{}' b_j{}'(x) \end{aligned} \tag{1}$$

B-spline basis functions are constructed so that continuity of the function and its derivatives up to order $k-2$ is guaranteed across every knot span. This property is achieved through the recursive construction proposed by Cox and de Boor[14], summarized below:

$$\begin{aligned} b_{j,0}(x) &= \begin{cases} 1 & t_j \le x \le t_{j+1} \\ 0 & \text{otherwise} \end{cases} \\ b_{j,k}(x) &= \frac{x - t_j}{t_{j+k} - t_j} b_{j,k-1}(x) + \frac{t_{j+k+1} - x}{t_{j+k+1} - t_{j+1}} b_{j+1,k-1}(x) \end{aligned} \tag{2}$$

This research claims that the B-splines can be adaptively improved during training to improve predicted accuracy while employing fewer parameters.

2.2 Transformer

The Transformer model can analyze time series data with diverse and intricate properties using processes like multi-scale embedding and cross-channel dependency capture, thereby successfully extracting the features and information inherent in the series[15].The method of action is illustrated in fig:1.

The Transformer model employs a multi-head attention mechanism (Multi-Head Attention) to capture global temporal information in sequences during its operation[16]. The structure is illustrated in fig:2.

By introducing the channel dimensionality reduction method in the Multi-Head Attention, the overfitting noise of the Transformer model can be reduced to further enhance its performance in financial time series forecasting.

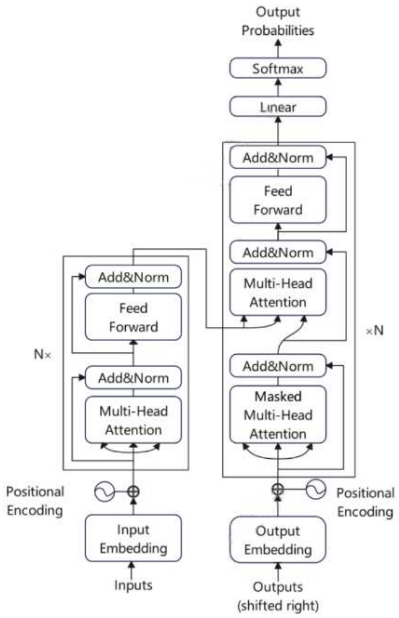


Fig. 1: Transformer

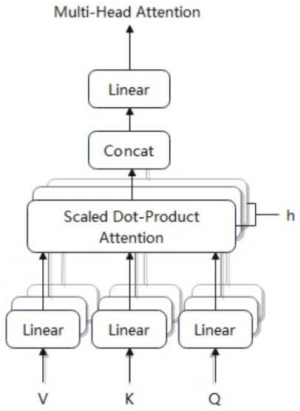


Fig. 2: Multi-Head Attention

2.3 Particle Swarm Optimization

The Particle Swarm Optimization (PSO) is a traditional optimization method that simulates collective behavior, dynamically modifying individual placements to identify the optimal solution through collaboration and learning among particles. Every particle possesses two primary characteristics: location and velocity[17]. Position denotes the existing solution, whereas velocity indicates the direction and magnitude

of the particle's motion. To update the particle's velocity, the following equation is employed[17]:

$$v_i^{t+1} = w \cdot v_i^t + c_1 \cdot r_1 \cdot \left(P_b - x_i^t\right) + c_2 \cdot r_2 \cdot \left(G_b - x_i^t\right) \tag{3}$$

P_b is the personal best position of particle i, and G_b is the global best position across all particles. Position update formula[17]:

$$p_i^{t+1} = p_i^t + v_i^{t+1} \tag{4}$$

ω is the inertia weight, c_1,c_2 is the learning facto, r_1,r_2 is the random number.
With the number of hidden layer units, learning rate, and Transformer hidden layer parameters in the hybrid model making up the particle population, this study sets the Mean Squared Error of the KAN-Transformer fusion model as the optimization goal.[18]. The technique aims to minimize the model's loss and determine the ideal configuration for the KAN-Transformer architecture by persistently updating both personal and global best positions.

3 construction of the KAN-Transformer

3.1 TKAT

The hybrid KAN-Transforemr model architecture constructed in this paper is TKAT(Temporal KAN-Transformer Model)[14]. fig:3displays the KAN-Transformer fusion model's structure.

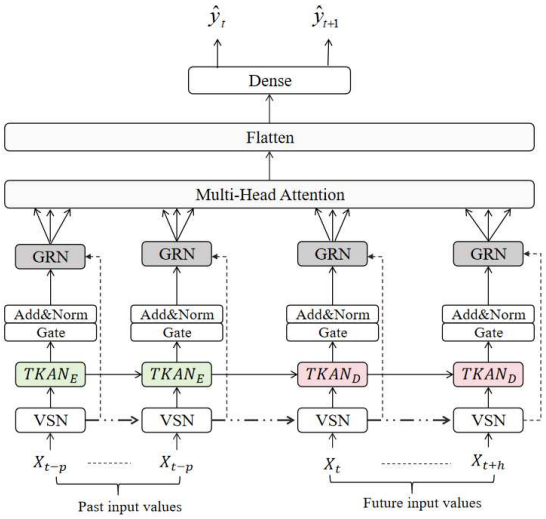


Fig. 3: TKAT

x_{t-p}: Past-observed Inputs; x_t, x_{t+h}: Apriori-known Future Inputs

$$\begin{aligned} x_{l+1,r}\left(t\right) &= \sum_{q=1}^{n} \tilde{x}_{l,r,q}\left(t\right) \quad = \sum_{q=1}^{n} \phi_{l,r,q,t}\left(x_{l,q}\left(t\right), h_{l,q}\left(t\right)\right) \\ r &= 1, \cdots, n_{l+1} \\ l &= 0, \cdots, L-1 \\ q &= 1, \cdots, n_l \end{aligned} \tag{5}$$

The term $\phi_{l,r,q}$ captures the temporal dependency, while $h_{l,q}(t)$ serves as a memory unit that retains historical information for node i in layer l.

$$h_{l,i}\left(t\right) = W_{hh} h_{l,i}\left(t-1\right) + W_{h,z} x_{l,i}\left(t\right) \tag{6}$$

W denotes the degree of correlation between different time series. Every layer of the RKAN (Recurrent KAN) layer incorporates "memory management":

$$KAN\left(x,t\right) = \left(\Phi_{L-1,t} \circ \Phi_{L-2,t} \circ \cdots \circ \Phi_{1,t} \circ \Phi_{0,t}\right)\left(x,t\right) \tag{7}$$

VSN: Variable Selection Network[14]; GRN: Gated Residual Network[14].

3.2 TKAN

In TKAT, TKAN (Temporal Kolmogorov-Arnold Networks) is the prevailing model[14]. TKAN is an enhanced iteration of LSTM that incorporates Kolmogorov Arnold Networks (KANs). The main parts in the TKAN structure are the RKAN layer and the gating mechanism layer. The RKAN layer uses a nonlinear activation function to update the hidden state after processing the current input and the hidden state from the previous time step. The gating mechanism layer mitigates the Vanishing Gradient problem in long sequence modeling by dynamically regulating the flow of information[15] . The TKAN structure is shown in fig:4.

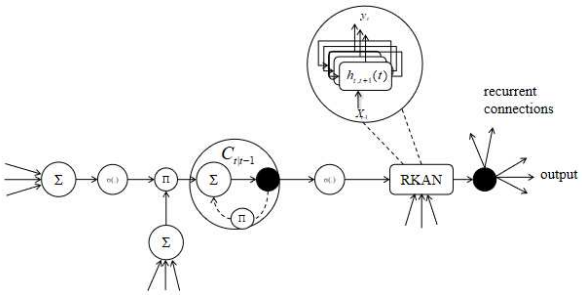


Fig. 4: TKAN

x_t is the d-dimensional input vector, b is the bia. The forgetting gate determines which past knowledge should be eliminated, and its activation vector function is represented as F_t[15]:

$$F_t = \sigma\left(W_F x_t + U_F h_{t-1} + b_F\right) \tag{8}$$

In a similar manner, the input gate chooses which fresh information should be kept in the current state.

$$I_t = \sigma\left(W_I x_t + U_I h_{t-1} + b_I\right) \tag{9}$$

The function for the output gate activation vector is articulated as O_t[15]:

$$O_t = \sigma\left(KAN\left(\vec{x}, t\right)\right) \tag{10}$$

The updated unit state $C_t$is represented as:

$$\begin{aligned} C_t &= F_t \odot C_{t-1} + I_t \odot \tilde{C}_t \\ \tilde{C}_t &= \sigma\left(W_C x_t + U_C h_{t-1} + b_C\right) \end{aligned} \tag{11}$$

Where W,U are parameter matrices,$\tilde{C}_t$ is the memory cell, and all these cell states have a dimension h. The output is represented as:

$$h_t = O_t \odot \tanh\left(C_t\right) \tag{12}$$

The Add and Norm layers in TKAT use the TKAN layer's output as input. "Add" signifies the Residual Connection, which mitigates the degradation of the neural network during training, while "Norm" refers to Layer Normalization, employed to standardize the activation levels of each layer[15].

4 Empirical evaluations

4.1 Feature Selection and Preprocessing

Since ST stocks are generally accompanied by trading halts, accumulation of limit-down boards, and individual stocks with low liquidity often having zero transactions on extreme days, long gaps appear in the time series. Given that the time series of "high-liquidity" ST stocks can provide continuously observable samples, price signals brought by low bid-ask spreads, and abnormal fluctuations triggered by consecutive limit-up/limit-down movements, the prediction and interpretability of the KAN-Transformer model for high-risk targets can be verified under the condition of small samples. Therefore, this paper selects "high-liquidity" as the screening criterion and places the experiment within the data boundary that is verifiable and interpretable.
This paper selects two ST stocks with high liquidity and active trading from 2020 to 2024, namely ST Zhongrun (000506) and ST Huawang (603007), to verify the interpretable prediction performance of the model for high-risk individual stocks, and the other two representative indices for the experiments are the Index of SSE 50 (000016), Index of SZSE 100(399300). The four datasets are stock trading day data from January 2, 2020 to February 29, 2024, all sourced from the Flush website, and

the selected features comprise the opening price, closing price, highest price, lowest price , and trading volume.
Each stock in the experiment comprises 1007 data points. Due to the sequential nature of the dataset, random sampling is unfeasible. Consequently, the training and test sets are delineated chronologically in an 8:2 ratio. Specifically, data from the initial 807 trading days comprise the training set, while data from the final 200 trading days constitute the test set.
To prevent unknown information from influencing model training, the raw data is normalized using the following conversion formula:

$$\hat{x} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{13}$$

4.2 Model calibration and validation assessment

Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE) are used in this work as evaluation metrics for model prediction performance and as the foundation for model parameter tweaking. The comprehensive calculating formulas are as follows:

$$\begin{aligned} MSE &= \frac{1}{N}\sum_{i=1}^{N}(Y_i - \hat{Y}_i)^2 \\ RMSE &= \sqrt{\frac{1}{N}\sum_{i=1}^{N}(Y_i - \hat{Y}_i)^2} \\ MAE &= \frac{1}{N}\sum_{i=1}^{N} \mid Y_i - \hat{Y}_i \mid \end{aligned} \tag{14}$$

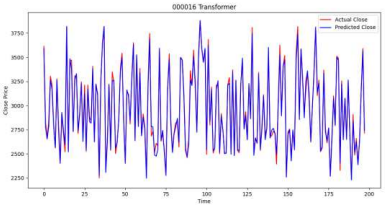
where Y_i denotes the actual value; $\hat{Y}_i$ denotes the anticipated value; N denotes the duration of the test set.

4.3 Stock Forecasting Analysis

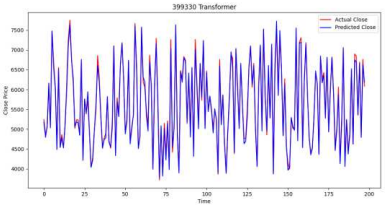
This study chooses the subsequent deep learning models for comparative analysis to evaluate the predictive efficacy of the KAN+Transformer fusion model in forecasting stock closing prices: Vanilla Transformer, Informer, Autoformer. The LSTM is selected as the basic model for training the Bi-LSTM-Attention model. The Bidirectional and LSTM classes are utilized to construct a bidirectional LSTM layer, configured with 50 units in the LSTM layer[19].The forecasts are presented in tab:1.

Table 1: Predictive results of former class models in the dataset.

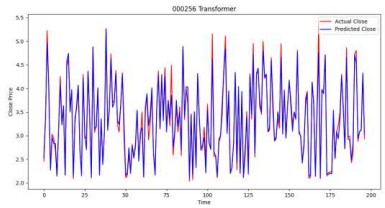
Data	Model	MAE	MSE	RMSE
Index of SSE 50	VanillaTransformer	0.1373	0.0689	0.2625
	Informer	0.3653	0.3483	0.5902
	Autoformer	0.3926	0.3623	0.6019
Index of SZSE 100	VanillaTransformer	0.1232	0.0543	0.2230
	Informer	0.2682	0.1863	0.4316
	Autoformer	0.2822	0.1949	0.4415
*ST Zhongrun	VanillaTransformer	0.1363	0.0874	0.2956
	Informer	0.2271	0.1153	0.3396
	Autoformer	0.3897	0.1663	0.4078
*ST Huawang	VanillaTransformer	0.1613	0.1225	0.3500
	Informer	0.3268	0.3070	0.5541
	Autoformer	0.4783	0.3854	0.6208



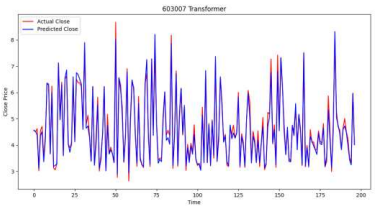
(a) 000016



(b) 399330



(c) 000256



(d) 603007

Fig. 5: VanillaTransformer prediction effect

The table in tab:1 demonstrates that the predictive performance of Vanilla Transformer markedly surpasses that of the other former models. The average values of the prediction performance assessment measures indicate that the Transformer decreases the MSE by 17.95% relative to the previous models.
Consequently, the Transformer is more beneficial for managing high dimensional, nonlinear stock data characterized by a poor signal-to-noise ratio.

To assess the efficacy of the self-attention mechanism in time series processing, it is incorporated into the Bi-LSTM model for experimentation[20].he Bi-LSTM-Attention model employs LSTM as the foundational model, utilizing the Bidirectional and LSTM classes to construct a bidirectional LSTM layer with 50 units. The prediction results are illustrated in tab:2.

Table 2: Comparison of Bi-LSTM-Attention and Bi-LSTM prediction results.

Data	Model	MAE	MSE	RMSE
Index of SSE 50	Bi-LSTM	0.0711	0.0098	0.0990
	Bi-LSTM-Attention	0.0717	0.0090	0.0949
Index of SZSE 100	Bi-LSTM	0.0707	0.0094	0.0970
	Bi-LSTM-Attention	0.0672	0.0089	0.0943
*ST Zhongrun	Bi-LSTM	0.0826	0.0124	0.1114
	Bi-LSTM-Attention	0.0760	0.0110	0.1049
*ST Huawang	Bi-LSTM	0.0633	0.0100	0.1000
	Bi-LSTM-Attention	0.0612	0.0086	0.0927

The experiments demonstrate that the prediction accuracy of Bi-LSTM is enhanced following the implementation of the Attention mechanism, thereby confirming the capability of the Transformer to capture dependencies within time series data.
The relevant parameter settings and termination criteria for finding the optimal parameters of the KAN-Transformer hybrid model using *PSO* are shown in tab:3.

Table 3: PSO Parameter Setting.

Parameter	Specification	Value
Swarm size	Sufficient empirical coverage of the 12-dimensional KAN structural space	30
Maximum iterations	Early-stopping triggered after 5 successive generations without improvement	200
c_1, c_2	Cognitive and social acceleration coefficient	1.5, 1.5
Inertia weight (w)	Linearly decreasing schedule	$0.9 \rightarrow 0.4$
Search bounds	Normalised B-spline scaling coefficients	$[0, 1]$

To rigorously assess the optimization stability of the *PSO*-KAN-Transformer, we maintain a fixed training–validation split and replicate the entire training procedure under ten independent random seeds (2024+i, i = 1,..., 10). During each iteration, the Mean Squared Error (MSE) on the validation set, attained by the globally optimal particle, is documented, and both the mean and standard deviation across 10 trials

are calculated to characterize the convergence behavior. Empirically, within the maximum allowance of 200 generations, the mean validation MSE declines monotonically from 0.026 to 0.014, and the relative fluctuation (σ/μ) over the final ten generations remains below 1.5%, signifying that the PSO has converged and is statistically stable. To rigorously substantiate the superior interpretability of KAN over MLP, we computed the cumulative contribution of each input feature to the model output on the training set using Integrated Gradients, subsequently averaging these attributions along the feature dimensions.

Table 4: Characteristic attribution ratio.

Features	**KAN**	**MLP**
Close	49.8%	21.7%
Volume	23.7%	15.4%
Open	10.2%	12.1%
High	8.4%	9.8%
Low	5.9%	7.5%

For the stock price prediction task, Close serves as the most information-rich state variable, as it fully reflects all publicly available market information. Volume acts as an amplifier for the credibility of price changes, reflects short-term reversal signals. As shown in tab:4, the KAN-Transformer concentrates 73.5% of the total attribution on two key features—Close and Volume—and its distribution is highly consistent with financial prior knowledge. In contrast, the contribution of the traditional MLP-Transformer is scattered across 5 features, with the highest individual feature attribution being less than 22%. This scattered distribution results in a lack of interpretability.

The PSO is utilized to ascertain the ideal configurations for the model-related parameters: hidden_dim=128, num_heads=4, and dropout=0.1. The training settings are defined as follows: num_epochs=50, batch_size=32, and learning_rate=0.001.

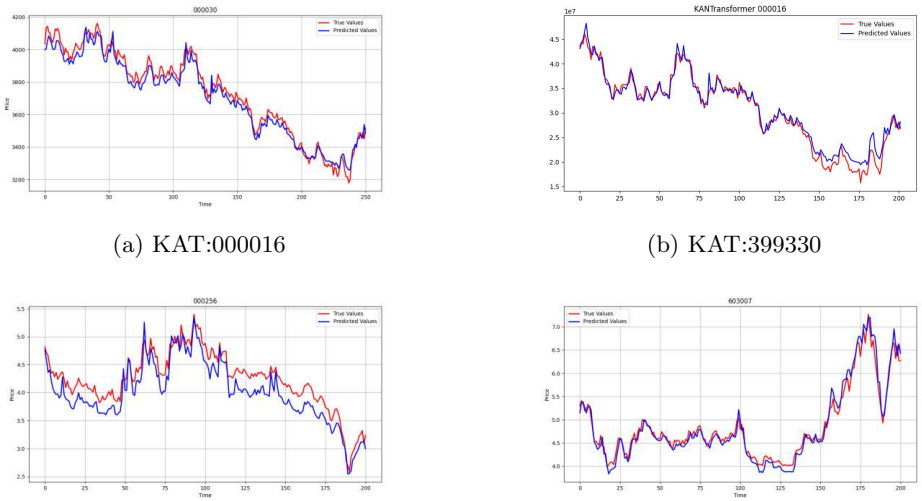
To assess the function and efficacy of the KAN-Transformer model introduced in this paper for analyzing high-dimensional stock data, Open, High, Low, Volume are incorporated into the model for training. Additionally, the prediction effect diagrams of the KAN-Transformer and KAN-LSTM models are compared to explore the ability of these two hybrid models to understand and predict long-term time series.

Table 5: KAN-Transformer Parameter Setting.

Parameter	Value	Reason
B-spline order	5	Liu et al. (2024)[22] demonstrated that the 5th order (quadratic) B-spline is the best compromise between smoothness and computational efficiency in a similar financial time-series experiment.
Grid_size	128	Each B-spline function is made to correspond exactly to 128 learnable control points, without the need for additional mapping, while maintaining alignment with the number of multi-head attention heads (which is 4).
hidden_width	64	The hidden width of KAN is set to half of the Transformer's hidden_dim (128/2 = 64). Experiments show that this configuration achieves the optimal trade-off between prediction error and interpretability.

Table 6: Prediction results of ST stocks under different models.

Data	Model	MAE	MSE	RMSE
Index of SSE 50	KAN-Transformer	0.0050	0.0064	0.0800
	KAN-LSTM	0.0070	0.0090	0.0949
Index of SZSE 100	KAN-Transformer	0.0049	0.0064	0.0800
	KAN-LSTM	0.0150	0.0187	0.1367
*ST Zhongrun	KAN-Transformer	0.0134	0.0173	0.1315
	KAN-LSTM	0.0137	0.0183	0.1353
*ST Huawang	KAN-Transformer	0.0116	0.0154	0.1241
	KAN-LSTM	0.0114	0.0159	0.1261



(a) KAT:000016

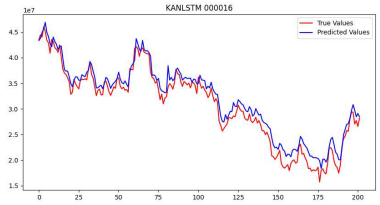
(b) KAT:399330

(c) KAT:000256

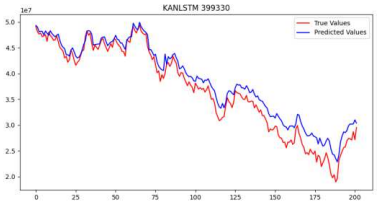
(d) KAT:603007

Fig. 6: KAN-Transformer prediction effect

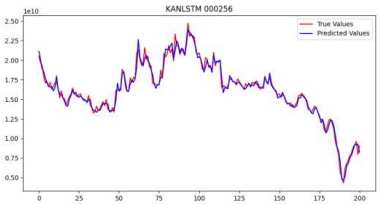
The predictive data for each evaluation metric in tab:6 demonstrates that the KAN-Transformer model displays the minimal MSE and RMSE values. This model exhibits exceptional prediction accuracy, especially for the comparatively stable large-cap indices, such as Index of SSE and SZSE. The KAN-Transformer model exhibits an average decrease in mean square error of 18.85% relative to the Transformer model.



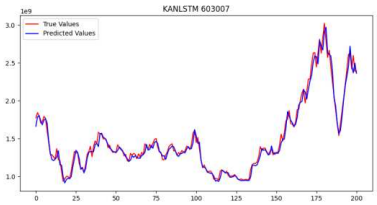
(a) KAL:000016



(b) KAL:399330



(c) KAL:000256



(d) KAL:603007

Fig. 7: KAN-LSTM prediction effect

5 Conclusion

This study introduces an approach that combines the KAN with the Transformer. The methodology entails replacing the MLP layer in the Transformer with the KAN layer, utilizing Multi-Head attention to examine the temporal dependencies between stock closing prices and feature variables, and applying the particle swarm optimization algorithm to identify the optimal parameters for the model. The experimental findings indicate that the KAN-Transformer model outperforms the existing deep learning models, including KAN-LSTM, Transformer, Autoformer,and Bi-LSTM-Attention, when compared to the rest of the benchmark prediction methods on four stock datasets. This model has superior feature extraction capabilities, parallel processing efficiency, and operational effectiveness in handling multi-dimensional long time series, and offers greater interpretability than the Transformer. Due to the influence of political, economic, social, and psychological factors on stock values, properly forecasting their fluctuations is challenging. Subsequently, future research will focus on incorporating sentiment analysis theories, such as financial news, public opinion,

and investor sentiment, into the stock price prediction model to improve the predictive accuracy of KAN-Transformer, thus providing investors with a more precise and practical foundation for decision-making.

Declarations

Some journals require declarations to be submitted in a standardised format. Please check the Instructions for Authors of the journal to which you are submitting to see if you need to complete this section. If yes, your manuscript must contain the following sections under the heading 'Declarations':

- Funding: This paper is supported by Qinghai Minzu University campus level project (23GCC10).

- Conflict of interest: The authors have no conflicts of interest to declare that are relevant to the content of this article.
- Ethics approval and consent to participate: Not applicable.
- Consent for publication: Not applicable.
- Data availability: The datasets supporting the conclusions of this article is available in the Tonghuashun Data Center, https://data.10jqka.com.cn/.
- Materials availability: Not applicable.
- Code availability: Not applicable.
- Author contribution: Conceptualization, Y.L.; methodology, Y.L.; software, S.P.; validation, B.Z.; formal analysis, S.P.; writing—original draft preparation, S.P.; writing—review and editing, S.P.; visualization, S.P.; supervision, B.Z.; project administration, B.Z.; funding acquisition, Y.K. All authors have read and agreed to the published version of the manuscript.

References

[1] Zhang, Y., & Liu, Z. (2022). Comparison of KAN and MLP in hybrid deep models for stock closing price forecasting. *Modern Electronics Technique*, 45(23), 121–126.

[2] Liu, M., & Shan, Y. (2021). Prediction of closing price of stock index based on EMD-LSTM model. *Journal of Chongqing University of Technology (Natural Science)*, 35(12), 269–276. doi:10.3969/j.issn.1674-8425(z).2021.12.032

[3] Shetty, D. K., & Ismail, B. (2021). Forecasting stock prices using hybrid non-stationary time series model with ERNN. *Communication in Statistics-Simulation and Computation*, 2021, 1–13.

[4] Chen, W., Li, S., Yu, X., Wang, H., Chen, W., & Wang, T. (2024). Automatic de-biased temporal-relational modeling for stock investment recommendation. In *Proc. of the Thirty-Third Int. Joint Conf. on Artificial Intelligence*, pp. 1999–2008.

[5] Li, W., & Law, K. L. E. (2024). Deep learning models for time series forecasting: A review. *IEEE Access*, 12, 22. doi:10.1109/ACCESS.2024.3422528

[6] Wu, S., Zhao, Y., et al. (2024). A study of dynamic stock relationship modeling and S&P500 price forecasting based on differential graph transformer. *Expert Systems with Applications*, 238, 122–134. doi:10.1016/j.eswa.2024.122134

[7] Zhou, H., Zhang, S., Peng, J., et al. (2020). Informer: Beyond efficient transformer for long sequence time-series forecasting. *arXiv preprint arXiv:2012.07436*. doi:10.48550/arXiv.2012.07436

[8] Zhu, X., Lin, P., Sun, M., et al. (2024). Feature selection and forecasting of financial data based on the BSFinformer model. *Journal of Nanjing University (Natural Sciences)*, 60(3), 442–450.

[9] Ren, J., & Wang, A. (2023). Stock price prediction using causal attention transformer model. *Journal of Computer Engineering & Applications*, 59(13).

[10] Ren, S., & Song, W. (2024). Financial data feature extraction and price prediction based on GGInformer model. *Journal of Zhengzhou University (Science Edition)*, 56(5), 62–70.

[11] Chen, D., & Mao, L. (2024). Stock price prediction using incremental learning and transformer model. *Journal of Computer Science and Exploration*.

[12] Zhang, H., Liu, Z., & Wang, J. (2024). Kansformer: Integrating KAN and transformer for stock return forecasting. In *Proc. of the 3rd Int. Conf. on Artificial Intelligence for Finance (ICAIF 2024)*, pp. 183–190. doi:10.1109/ICAIF59757.2024.00035

[13] Somvanshi, S., Javed, S. A., Islam, M. M., Pandit, D., & Das, S. (2024). A survey on Kolmogorov-Arnold network. *arXiv preprint arXiv:2411.06078*.

[14] Yang, X., & Wang, X. (2024). Kolmogorov-Arnold transformer. In *The Thirteenth Int. Conf. on Learning Representations*.

[15] Genet, R., & Inzirillo, H. (2024). A temporal Kolmogorov-Arnold transformer for time series forecasting. *ArXiv*.

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

[17] Song, G., Zhang, Y., Bao, F., et al. (2019). Stock prediction model based on particle swarm optimization LSTM. *Journal of Beijing University of Aeronautics and Astronautics*, 45(12), 2489–2496. doi:10.13700/j.bh.1001-5965.2019.0388

[18] Shao, B., Li, M., Zhao, Y., et al. (2019). Nickel price forecast based on the LSTM neural network optimized by the improved PSO algorithm. *Mathematical Problems in Engineering*, 2019, 1–15. doi:10.1155/2019/4180790

[19] Livieris, I. E., Pintelas, E., & Pintelas, P. (2020). A CNN-LSTM model for gold price time-series forecasting. *Neural Computing and Applications*, 32, 17351–17360.

[20] Sunny, M. A. I., Maswood, M. M. S., & Alharbi, A. G. (2020). Deep learning-based stock price prediction using LSTM and bi-directional LSTM model. In *2020 2nd Novel Intelligent and Leading Emerging Sciences Conference (NILES)*, pp. 87–92.

[21] Kong, Y., Qin, Y., & Li, Y. (2025). Stock return prediction based on threshold distance-weighted loss. *Journal of Econometrics (China)*, 5(1), 35–51.

[22] Liu, Z., Wang, Y., Vaidya, S., Ruehle, F., Halverson, J., Soljačić, M., Hou, T. Y., & Tegmark, M. (2024). KAN: Kolmogorov–Arnold Networks. *arXiv preprint arXiv:2404.19756*.

[23] Zhao, S., Lin, X., & Weng, X. (2023). Attention-BiLSTM stock price trend forecasting model based on empirical mode decomposition and investor sentiment. *Journal of Computer Applications*, 43(S1), 112–118.

[24] Zhang, Y., & Wang, H. (2023). Stock price forecasting via a transformer model with causal attention. *Computer Engineering and Applications*, 59(14), 235–242. doi:10.3778/j.issn.1002-8331.2211-0243

[25] Zhu, X., Lin, P., Sun, M., et al. (2025). KAN-BiLSTM: A hybrid model for stock index prediction based on Kolmogorov-Arnold networks and bidirectional LSTM. *Journal of Chongqing University of Science and Technology*, 27(2), 45–52.

[26] Muhammad, A. U., Yahaya, A. S., Kamal, S. M., Adam, J. M., Muhammad, W. I., & Elsafi, A. (2020). A hybrid deep stacked LSTM and GRU for water price prediction. In *2020 2nd Int. Conf. on Computer and Information Sciences (ICCIS)*, pp. 1–6.

[27] Li, T., Liu, Z., Shen, Y., Wang, X., Chen, H., & Huang, S. (2024). MASTER: Market-guided stock transformer for stock price forecasting. In *Proc. of the AAAI Conf. on Artificial Intelligence*, Vol. 38, pp. 162–170.

[28] Zhou, Q., Wang, X., & Zhou, C. (2025). FS-CNN-BGRU: Stock closing price prediction with feature selection and bidirectional GRU enhanced by Kolmogorov-Arnold network. *PLOS ONE*, 20(3), e0262501. doi:10.1371/journal.pone.0262501

[29] Lu, W., & Wu, S. (2024). CNN-Attention-GRU-Attention: A novel model for stock closing price prediction. In *Proc. of the IEEE 4th Int. Conf. on Power, Electronics and Computer Engineering (ICPECE 2024)*, pp. 183–190. doi:10.1109/ICPECE60733.2024.00035

[30] Wen, X. (2022). High-frequency financial time-series forecasting via empirical mode decomposition and linear transformer. *Modern Electronics Technique*, 45(23), 121–126.