Enhanced LFTSformer: A Novel Long-Term Financial Time Series Prediction Model Using Advanced Feature Engineering and the DS Encoder Informer Architecture

JIANAN ZHANG [1] **(Student Member, IEEE), HONGYI DUAN** [2,3] **(Student Member, IEEE).**

[1]School of Mathematica, Shanghai University of Finance and Economics, Yangpu District, Shanghai 200437 China
[2]Faculty of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an, Shannxi 710049 China
[3]School of Computing Information System, The University of Melbourne, Melbourne, VIC 3010 Australia

Corresponding author: Hongyi Duan (e-mail: Dann_Hiroaki@ieee.org).

JIANAN ZHANG and HONGYI DUAN Contributed to the work equally and should be considered co-first authors

ABSTRACT This study presents a groundbreaking model for forecasting long-term financial time series, termed the Enhanced LFTSformer. The model distinguishes itself through several significant innovations:

1) **VMD-MIC+FE Feature Engineering:** The incorporation of sophisticated feature engineering techniques, specifically through the integration of Variational Mode Decomposition (VMD), Maximal Information Coefficient (MIC), and feature engineering (FE) methods, enables comprehensive perception and extraction of deep-level features from complex and variable financial datasets.

2) **DS Encoder Informer:** The architecture of the original Informer has been modified by adopting a Stacked Informer structure in the encoder, and an innovative introduction of a multi-head decentralized sparse attention mechanism, referred to as the Distributed Informer. This modification has led to a reduction in the number of attention blocks, thereby enhancing both the training accuracy and speed.

3) **GC Enhanced Adam & Dynamic Loss Function:** The deployment of a Gradient Clipping-enhanced Adam optimization algorithm and a dynamic loss function represents a pioneering approach within the domain of financial time series prediction. This novel methodology optimizes model performance and adapts more dynamically to evolving data patterns.

Systematic experimentation on a range of benchmark stock market datasets demonstrates that the Enhanced LFTSformer outperforms traditional machine learning models and other Informer-based architectures in terms of prediction accuracy, adaptability, and generality. Furthermore, the paper identifies potential avenues for future enhancements, with a particular focus on the identification and quantification of pivotal impacting events and news. This is aimed at further refining the predictive efficacy of the model.

INDEX TERMS Feature Engineering, Financial Time Series, Integrated Models, Stacked Informants, Variational Mode Decomposition.

I. INTRODUCTION

A. RESEARCH BACKGROUND

The financial market constitutes a pivotal foundation for the sustenance of global economic endeavors, with its operations and dynamical shifts intricately intertwined with a confluence of complex determinants, encompassing economic structuration, seasonal variabilities, and the international milieu [1] [2]. Concomitant with economic progression and the burgeonment of financial markets, the adoption of time series analysis has emerged as an indispensable instrumentality within the finance domain. [3] This analytical paradigm has markedly enhanced the comprehension of market dynamics,

augmented the intelligence of decision-making processes, and facilitated the prognostication of investment returns. [2] [4] Consequently, time series analysis has garnered extensive scholarly interest, engendering a copious corpus of research findings and advancing the academic discourse on financial market analysis. [5] [6]

The domain of long-term time series forecasting, increasingly pivotal across an array of disciplines, encapsulates several quintessential applications including the temporal allocation mechanisms pivotal for wind energy integration [7], the granular dissection of protracted energy consumption patterns within architectural edifices [8], and the nuanced prognostication of load dynamics within thermal infrastructures [9], among others. In marked divergence from the traditional paradigms of time series forecasting, the arena of financial time series forecasting is imbued with pronounced complexity and a heightened degree of unpredictability. This complexity is particularly accentuated in scenarios where prognostications are extended towards long-term horizons or when the forecasts are subjected to unforeseen exogenous perturbations [10] [11].

Research on financial time series forecasting has continuously evolved since the 1950s, despite numerous challenges. Initially, the field predominantly employed pure statistical and mathematical methodologies. Over time, it incorporated machine learning and deep learning techniques. Contemporary research focuses on the refinement of these models through the integration of advanced attention mechanisms and sophisticated feature engineering strategies, representing the cutting edge in financial time series analysis.

B. FINANCIAL TIME-SERIES PREDICTION: A REVIEW
1) Classical Theories in Financial Time Series Analysis
Within the domains of finance and macroeconomics, the inaugural application of mathematical models, anchored in the principles of econometrics and statistical theory, has pioneered the analytical and prognostic assessment of time series data. Notably, Robert Engel's seminal introduction of the Autoregressive Conditional Heteroskedasticity (ACH) framework, while examining the volatility of UK inflation rates, has emerged as a significant landmark in the evolution of financial market volatility forecasting methodologies [12]. This foundational work was subsequently advanced by Tim Boleslef, who refined it into the Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model, thereby significantly broadening the analytical precision and applicability of this approach within the realm of financial time series analysis [13]. Complementing these developments, David A. Shea's exploration into the volatility of five major foreign exchange rates, through the employment of a sophisticated statistical model that accounts for time-variant mean and variance, has further augmented the repertoire of financial market forecasting methodologies [14]. Part of this era's research is distinguished by the strategic amalgamation of diverse financial time series datasets, fostering a multidimensional analytical perspective [15].

In the 21st century's dynamic financial landscape, marked by burgeoning economic activities and market expansion, there has been a proliferation of statistical models for financial time series forecasting. The most remarkable is RG Brown's introduction of the ARIMA model in 2004 [16] has catalyzed its adoption for predicting stock and futures market trends, alongside its variants like AR, VAR, ARDL, ARCH, GARCH, and MIDAS [17]. These models excel in linear stock price scenarios, effectively assessing prediction risks [18] and forecasting returns [19]. Nonetheless, their predictive power for highly volatile stocks influenced by multifarious factors remains debated. To address this challenge, a large number of fundamental machine learning models and methods have been progressively incorporated into the forecasting of financial and stock time series under the rise of the artificial intelligence field.

2) Introduction of machine learning to financial time series prediction
The utilization of machine learning techniques in forecasting financial time series marks the advent of sophisticated analytical methodologies. The introduction of temporal difference learning by Sutton signals a pivotal shift, utilizing historical data to enhance predictive accuracy while maximizing computational resourcefulness [20]. The deployment of genetic algorithms by Mahfoud and Mani, along with Kim and Han, has refined the processes of network training and feature selection [21] [22]. Additionally, the integration of Support Vector Machines (SVM) by Trafalis and Ince, enriched by Lee, has notably augmented the precision of stock trend predictions, underscoring the utility of composite feature selection techniques [23] [24].

Deep learning, a significant evolution within machine learning, has revolutionized financial forecasting with marked improvements in precision and the ability to generate deep insights. Chen et al.'s employment of LSTM for the prediction of returns in China's stock market illustrates a considerable leap over traditional methodologies, adeptly navigating market unpredictability [25]. The CNN-GA model developed by Li et al. for short-term pork price prediction showcases deep learning's adaptability, yielding high precision and underscoring its value in the analysis of commodity markets [26]. Ding et al.'s model for event-driven forecasting, which leverages deep convolutional neural networks, offers an intricate analysis of how events influence stock prices, providing a profound augmentation over existing baselines [27].

3) Improvement of machine learning based financial time series Prediction
In the context of the rapidly growing volume of data in today's financial markets, machine learning and deep learning algorithms have demonstrated superior performance in time series forecasting compared to traditional forecasting methods. However, these advanced algorithms are not without limitations [28]. A particularly prominent challenge is

that financial time series data is too complex and variable, resulting in an inability to adequately extract salient features [29]. In addition, the fidelity of predictions tends to decrease over time, with RNNs and LSTMs being examples of decreasing predictive effectiveness [30]. To address these challenges, a large body of academic work has been devoted to the design of sophisticated feature engineering, as well as the introduction and optimisation of attentional mechanisms suitable for long time-series forecasting, which has yielded promising results in the field of financial forecasting.

Our research focuses primarily on attention mechanisms and the Informer model. Since Vaswani et al. established the fully attention-based Transformer model in 2016, which does not rely on recurrence and convolution [31], it has triggered a paradigm shift in the fields of machine translation and computer vision [32] [33], and has been widely recognized for its outstanding performance in areas such as time series prediction [34]. Following this, Zhou et al [35]. introduced Informer, aimed at addressing the complexities of long-sequence time series forecasting (LSTF) by incorporating an efficient attention mechanism that notably diminishes time complexity and memory consumption while preserving long-term dependencies with remarkable precision. This advancement, alongside further enhancements and innovations in the Transformer architecture, was promptly integrated into the domain of long-term financial time series forecasting.

As for Feature Engineering and other aspects, it merits attention that additional significant optimizations and inquiries into long-term time series forecasting have been explored, including stacking-based ensemble methodologies [36], and innovations in network structure fusion [37] [38], grounded in comprehensive parameter pruning strategies. Additionally, the decomposition-reconstruction approach, utilizing Variational Mode Decomposition (VMD) and Empirical Mode Decomposition (EMD) [39] [40] [41], has been identified as another pivotal research area. These optimization techniques, serving as integral components of our study, will also manifest in the context of feature engineering, thereby augmenting the efficacy and applicability of our research.

C. OUR RESEARCH CENTRE AND ARTICLE STRUCTURE

In summary, in the context of adequate feature extraction and accurate prediction for long time series, we combine the latest advances with the introduction of the pioneering Enhanced LFTS Informer, which combines the advantages of VMD-MIC/FE in the feature engineering stage, and through the mechanisms of Encoder Stacking/Decentralisation, coupled with the GC-enhanced Adam Optimisation as well as the Dynamic The introduction of Loss Function achieves excellent results in long term financial time series prediction (this time stocks are selected).

The rest of this article will be developed and discussed in the following order

1) **Section II**: A review of the progress and shortcomings of informer-based modules in financial time series and other time series forecasting in recent years.
2) **Section III**: Describes the overall architecture of the article's research, including data preprocessing before feature engineering, feature engineering methods, DS Encoder Informer architecture, etc.
3) **Section IV**: Demonstrates the process of feature engineering and results of ablation experiments / comparison experiments / stability experiments
4) **Section V**: Summarize the conclusions of this article and critically proposes potential improvement directions for this method.

II. RELATED WORKS

A. INFORMER-BASED MODEL IN FINANCIAL TIME SERIES PREDICTION

In the current research field of financial time series forecasting, studies based on Informer models are not only scarce in number, but also of varying quality. Despite these challenges, some representative studies demonstrate the potential and effectiveness of Informer models when dealing with complex financial data. By analysing these studies, we can gain a deeper understanding of the application of Informer models in stock price forecasting and market trend analysis.

First, one category of research focuses on improving the accuracy of stock price forecasting. For example, Ren et al. developed a hybrid model combining Encoder Forest and Informer, which significantly improves the forecasting ability through the decomposition and reconstruction approach [42].Ding et al. then optimised intraday stock price forecasting using the Informer model, demonstrating its superiority in capturing long-range dependencies [43]. These studies highlight the ability of Informer models in accurately capturing market dynamics.

Secondly, another class of studies focuses on the improvement of long-term forecasting.Liu et al. employed the Informer model optimised by the PSO algorithm to provide more accurate forecasts for long-term stock price trends, overcoming the shortcomings of the traditional model in long series forecasting [44]. In addition, Sababipour Asl used the Informer model to improve the prediction accuracy of stock market volatility, demonstrating its potential application in financial risk management [45].

Finally, there are also studies focusing on the versatility and integrated application of models. For example, Abdulsahib and Ghaderi demonstrated the advantages of the Informer encoder in capturing global and local market dynamics, proposing a new financial market prediction model [46]. In addition, Fei Xiong and Yuhao Feng's master's thesis investigated quantitative trading strategies based on the Informer mechanism, verifying the applicability and effectiveness of the model in real financial trading [47] [48].

B. INFORMER-BASED MODEL IN OTHER TIME SERIES PREDICTION

Informer-based models have not been widely studied in the field of financial time series. Therefore, for analysis

and reference, several representative Informer-based models from other time-series prediction domains are selected for this review. These models mainly focus on three aspects, namely structural improvement, performance optimisation and application-specific customisation, in order to improve prediction accuracy, processing power and application adaptability. Thus, they provide possible insights and directions for financial time series prediction

In terms of the underlying structural optimisation of the model structure, the MSRN-Informer model enhances the ability to extract features from time-series data by integrating the multi-scale structure and residual network, reduces information loss, and effectively improves the limitations of the traditional deep network in processing time-series data [49]. In addition, the Enformer model employs a simplified Transformer structure that uses only the encoder and omits the decoder, which reduces the computational complexity by introducing a sparse periodic attention mechanism, and at the same time enhances the model's ability to capture the periodic dependence of the sequence [56]

In terms of performance optimisation, the MTS-Informer model enhances the model's ability to handle complex multivariate data by introducing causal convolution and a sparse self-attention mechanism, and improves the handling of unstable sequences [51]. Literature [52] demonstrates the application of Informer in the prediction of motor bearing vibration, which optimises the temporal and spatial complexity and improves the prediction accuracy and reliability of the model by employing ProbSparse self-attention and self-attention distillation techniques.

In terms of customised tuning for specific application scenarios, the improved stacked Informer network is designed for power line fault prediction, optimising the network's generalisation capability and training speed by incorporating gradient concentration techniques [53]. Whereas the model in literature [54] combines the integration of Empirical Modal Decomposition (EEMD) with Particle Swarm Optimisation algorithm (PSO) tuning of Informer, specifically designed for multi-step prediction of building energy consumption to cope with non-linear and non-smooth energy consumption data.

III. METHODOLOGY
A. DATA PREPROCESSING
1) Raw Data

In this study, we adopt stocks, the most common and widespread financial time series, as the object of study. Based on this, we collect the stocks of 22 large enterprises listed on the two major stock exchanges in China (Shanghai Stock Exchange and Shenzhen Stock Exchange) as the research object, and the basic information is shown in Table 1.

For the raw data collected, the basic paradigm and Attributes are shown in Table 2. As for the complete initial data, please refer to this link. One thing to note is that all data related to transaction volume are in RMB.

TABLE 1. Data Collected and Used in This Study

Stock Code	Volume	Interval	Start Date	End Date
SHA: 601398	190225	30min	2006/10/30	2023/03/17
SHE: 002415	146833	30min	2010/05/28	2023/03/21
SHA: 600007	269429	30min	1999/10/28	2023/03/21
SHA: 601628	188545	30min	2007/01/10	2023/03/21
SHA: 601857	179041	30min	2007/11/06	2023/03/21
SHA: 600654	244903	30min	1999/10/28	2023/03/21
SHA: 600004	229681	30min	2003/04/29	2023/03/21
SHE: 300059	148609	30min	2010/03/19	2023/03/21
SHA: 600006	266793	30min	1999/10/28	2023/03/21
SHE: 000004	255569	30min	1999/10/28	2023/03/21
SHA: 600009	269384	30min	1999/10/28	2023/03/21
SHE: 000938	248423	30min	1999/11/11	2023/03/21
SHE: 600653	258001	30min	1999/10/28	2023/03/21
SHA: 600601	268896	30min	1999/10/28	2023/03/21
SHA: 601800	128689	30min	2012/03/09	2023/03/21
SHA: 600000	265010	30min	1999/12/07	2023/03/17
SHE: 000001	262608	30min	1999/12/21	2023/03/21
SHA: 600602	255553	30min	1999/10/28	2023/03/21
SHE: 000089	268902	30min	1999/10/27	2023/03/21
SHA: 600028	249217	30min	2001/08/08	2023/03/21
SHE: 000063	264673	30min	1999/10/27	2023/03/21
SHA: 600065	251770	30min	1999/10/28	2023/03/21

TABLE 2. Stock Raw Data Format Example

Time	Open	High	Low	Close	Volume	Amount
199910280950	6.3	6.33	6.3	6.32	62700	395321
199910280955	6.32	6.32	6.3	6.32	84400	532593
199910281000	6.31	6.33	6.31	6.31	37600	237544
199910281005	6.3	6.31	6.3	6.31	37700	237674
....	...	...	...	....	.....	......

2) Data Indexing

To fully reflect the inherent patterns in the data, we extracted 29 features from the original dataset. This includes 8 basic indicators as shown in Table 3. Building on the studies by Achelis [55], Kara et al. [56], and Guresen et al. [57], we identified an additional 21 advanced features for this research, as listed in Table 4. While the calculation methods for all advanced features are comprehensively discussed in Achelis [55].

TABLE 3. Basic Features

Feature name	Window size	Number of features
Open, Close, High, Low	1	4
Weighted Price	1	1
Volume(Stock), Volume(RMB)	1	2
High - Low	1	1

Due to space constraints, we only explain the following specific indicators here.

TABLE 4. Advanced Features

Feature name	Window size	Number of features
Return	$1, \ldots, 6$	6
Correlation MA5 and MA30	1	1
Sum3, Sum5	5,3	2
Sum5 - Sum3	1	1
|Sum5 - Sum3|	1	1
$\sqrt{(Sum5)^2 - (Sum3)^2}$	1	1
RSI(6, 14)	6,14	2
Rate of change	9,14	2
Williams R	1	1
ATR	5,10	2
CCI	1	1
DEMA	1	1

1) **Return:** Refers to returns at 6 different lag levels (from 1 to 6 lag length)
2) **Average True Range (ATR):** Calculated by the following formula:

$$ART = MA(TR,N) \tag{1}$$

Where $MA(TR, N)$ represents the moving average of TR over N time units, i.e., the average price over the past N time periods. The true range TR_t for the tth time period is computed as:

$$TR_t = \max \left\{ \begin{bmatrix} High_t \\ High_t \\ Close_{t-1} \end{bmatrix} - \begin{bmatrix} Low_t \\ Close_{t-1} \\ Low_t \end{bmatrix} \right\} \tag{2}$$

where High, Low, and Close are the highest, lowest, and closing prices respectively.

3) **Relative Strength Index (RSI):** Calculated as:

$$RSI = 100 - \frac{100}{1 + \dfrac{MA(U, N)}{MA(D, N)}} \tag{3}$$

When the price increases, we have:

$$U_t = Close_t - Close_{t-1}, \quad D = 0 \tag{4}$$

and when it decreases, we have:

$$D_t = Close_{t-1} - Close_t, \quad U = 0 \tag{5}$$

4) **Price Return:** Calculated with the following formula:

$$Return_{i_t} = Close_t - Close_{t-i} \tag{6}$$

5) **Rate of Change (ROC):** Calculated as:

$$ROC_{i_t} = 100 \cdot \left(\frac{Close_t}{Close_{t-i}} - 1 \right) \tag{7}$$

6) **Stochastic Oscillator:** Reflects the momentum effect in the Bitcoin price, and its calculation formula is as follows:

$$K = 100 \cdot \frac{Close - Low}{High - Low} \tag{8}$$

$$\downarrow$$
$$Stochastic\ oscillator = MA(K, N) \tag{9}$$

7) **Commodity Channel Index (CCI):** Reflects the average price deviation, calculated as:

$$\begin{cases} TP = \dfrac{High + Low + Close}{3} \\[2mm] MC = MA(Close, N) \\[2mm] MD = MA(MC - Close, N) \end{cases} \tag{10}$$

$$\downarrow$$

$$CCI = \frac{TP - MC}{0.015MD} \tag{11}$$

3) Data Manipulation

Given the heightened complexity inherent to stock prices, the natural logarithm of the terminal prices was extracted, followed by first-order differencing. The corresponding formula is presented as:

$$new\ close_i = \log(\frac{close_i}{close_{i-1}}) \tag{12}$$

To enhance the efficiency of our neural network model, we standardized the 29 indicators. The calculation formula is:

$$new\ value_i = \frac{value_i - E(value)}{\sqrt{D(value)}} \tag{13}$$

Where E(value) and D(value) represent the expected value and variance of the indicator, respectively.

B. FEATURE ENGINEERING

1) VMD Model

Variance Mode Decomposition (VMD) is a sophisticated data decomposition technique, facilitates the transformation of time-domain financial sequences into the frequency domain [58]. This transformation yields a set of K intrinsic mode functions (IMFs). The foundational step in this procedure entails formulating the variational constraint equation:

$$\begin{cases} \min \left\{ \sum_{K=1}^{K} \left\| \partial_t \left[\left(\delta(t) + \dfrac{i}{\pi t} \right) * u_K(t) \right] e^{-iw_K t} \right\|_2^2 \right\} \\[4mm] s.t. \sum_{K=1}^{K} u_K = x(t) \end{cases} \tag{14}$$

Here, $x(t)$ denotes the financial data sequence. The terms w_K and u_K respectively represent the central frequency and band components of the k-th IMF. The impulse function is represented by $\delta(t)$, while ∂t indicates the derivative with respect to time. The convolution operation (*) signifies the convolution computation in this context.

For a more tractable representation, the variational constraint equation is restructured using the Lagrange function $\lambda(t)$ and the penalty factor α:

$$\begin{cases} \text{TDR} &= \alpha \sum_{K=1}^{K} \left\| \partial_t \left[\left(\delta(t) + \dfrac{i}{\pi t} \right) * u_K(t) \right] e^{-i w_K t} \right\|_2^2 \\\\ \text{DAE} &= \left\| x(t) - \sum_{K=1}^{K} u_K(t) \right\|_2^2 \\\\ \text{LM} &= \left\langle \lambda(t), x(t) - \sum_{K=1}^{K} u_K(t) \right\rangle \end{cases} \tag{15}$$

$$\downarrow$$

$$L = (\{u_K\}, \{\omega_K\}, \lambda) = \text{TDR} + \text{DAE} + \text{LM} \tag{16}$$

This reformulation alleviates the constraints, rendering the problem more manageable. The optimal solution is then determined using the Alternating Direction Method of Multipliers (ADMM), a renowned optimization technique. The iterative steps are as follows:

$$\hat{u}_K^{n+1}(w) = \frac{\hat{x}(w) - \sum_{i \neq K} \hat{u}_j(w) + \dfrac{\hat{\lambda}(w)}{2}}{1 + 2\alpha (w - w_K)^2} \tag{17}$$

$$\downarrow$$

$$w_K^{n+1} = \frac{\displaystyle\int_0^\infty w \left| \hat{u}_K^{n+1}(w) \right|^2 dw}{\displaystyle\int_0^\infty \left| \hat{u}_K^{n+1}(w) \right|^2 dw} \tag{18}$$

Following this sequence, the primary financial data sequence is decomposed into K distinct sub-series.

2) VMD-MIC Framework

While VMD is underpinned by a robust mathematical framework and adept at disentangling complex signal components, it's constrained by the need for predetermined decomposition parameters. This requirement can compromise the fidelity of the decomposition. To address these limitations, this study proposes an amalgamation of VMD with the Mutual Information Coefficient (MIC) [59]. This integration determines the optimal number of decompositions, K, enhancing the decomposition's adaptability.

The Mutual Information Criterion (MIC) is a statistical measure used to quantify the dependency between two variables, indicating the amount of information shared between them. As for the Calculation of the Mutual Information Criterion (MIC), when given two variables, X and Y, their mutual information $I(X;Y)$ is defined as:

$$I(X;Y) = \sum_{x \in X} \sum_{y \in Y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)} \tag{19}$$

Here, $p(x,y)$ represents the joint probability distribution of X and Y, while $p(x)$ and $p(y)$ are the marginal probability distributions of X and Y, respectively.

The Mutual Information Criterion (MIC) is a normalized form of mutual information, calculated as follows:

$$\text{MIC}(X;Y) = \frac{I(X;Y)}{\sqrt{H(X)H(Y)}} \tag{20}$$

In this equation, $H(X)$ and $H(Y)$ represent the entropy of X and Y respectively, used to normalize the mutual information such that the values of MIC range from 0 to 1.

In the context of Variational Mode Decomposition (VMD), the Mutual Information Criterion (MIC) serves as a robust measure for quantifying the accuracy and effectiveness of the decomposition process. This criterion is instrumental in determining the optimal number of modal components k, which is critical for achieving high fidelity in the reconstruction of the original signal. The methodology can be outlined in several systematic steps:

1) **Decomposition:** Initially, the original signal y is decomposed into various modal components using VMD for different potential values of k. This step is essential to explore the range of possible decompositions and to prepare for subsequent analysis.

2) **Reconstruction:** Following the decomposition, these modal components are reassembled to form the reconstructed signal y_0. This step tests the completeness of the original decomposition by attempting to recreate the initial signal using the modal components derived.

3) **Evaluate MIC:** The Mutual Information Criterion (MIC) is then calculated to assess the degree of similarity between the original signal y and the reconstructed signal y_0. The MIC, denoted as $\text{MIC}yy_0$, evaluates the mutual information, providing a statistical measure of how well the reconstructed signal represents the original.

4) **Optimization:** The final step involves selecting the value of k that yields the maximum MIC value. An MIC value approaching 1 signifies minimal information loss during the reconstruction process, indicating that the decomposition has been optimally performed with that particular number of modal components.

By utilizing the MIC in this manner, VMD can be fine-tuned to ensure minimal information loss and optimal reconstruction quality. This approach not only enhances the analytical capabilities of VMD but also ensures that the signal processing is both precise and efficient, leading to more accurate interpretations and outcomes in practical applications.

3) Fuzzy Entropy Approach

Fuzzy Entropy (FE) serves as a prominent technique tailored for assessing the complexity inherent in temporal series, offering a profound insight into the intricacies of the data structure [60]. A salient feature of FE is its adaptive nature, which allows for a seamless evolution in accordance with changes in its prescribed parameters. This adaptability confers upon FE a robust defense mechanism against noise-induced perturbations, ensuring its integrity even in the presence of external disturbances.

For a time series of length n, the FE algorithm is seamlessly integrated within the framework governed by fuzzy membership functions. This methodological integration results in the subsequent formulation:

$$\ln \mathrm{D}(x) = -\ln(2) \left(\frac{x}{r} \right)^2 \qquad (21)$$

Here, r represents the similarity tolerance, and x is equated to d_{ij}^m, the distance between vectors that reconstruct the time series in an m-dimensional phase space. For $i, j = 1, 2, \ldots, n - m + 1$, taking an average over each i in D_{ij}^m produces the average similarity function:

$$\phi^m(r) = \frac{1}{\mathrm{N} - m + 1} \int\limits_{i=1}^{\mathrm{N}-m+1} \left(\frac{1}{\mathrm{N} - m} \int\limits_{j=1, j\neq i}^{\mathrm{N}-m+1} \mathrm{D}_{ij}^m \right) \qquad (22)$$

$$\downarrow$$

$$\mathrm{FE}(m, r, n) = \ln \left(\frac{\Phi^m(r)}{\Phi^{m+1}(r)} \right) \qquad (23)$$

C. DS ENCODER INFORMER: ENCODER-BASED OPTIMIZED INFORMER

In our research on long-term financial time series prediction, we've established a rolling prediction setting with fixed window sizes, linking input at time t with its corresponding sequence prediction output. This model supports output lengths that exceed the length of previous inputs and is not limited to univariate feature dimensions. At the core of many popular models, including our proposed DS Encoder Informer Architecture, is the encoder-decoder architecture designed to encode the input representation X_t into a hidden state H_t, and then decode this state to produce the output Y_t. This involves a process known as "dynamic decoding," where the decoder iteratively infers by calculating a new hidden state $h_{t(k+1)}$ from the previous state h_{tk} and the required outputs at step k, and predicts the next sequence $y_{t(k+1)}$. For an overview of this architecture and more details, please refer to our framework diagram.

1) Temporal Feature Embedding Mechanism

In the context of financial time series forecasting, the integration of hierarchical timestamps (like weeks, months, and years, as well as holidays and significant events) is vital for improving predictive accuracy. Traditional self-attention mechanisms often struggle with effectively processing this rich temporal information, which can degrade the interaction between query and key in encoder-decoder architectures.

To address these challenges, our study introduces a sophisticated input representation that blends both local and global temporal information. Here's how we've refined the methodology:

1) **Local Context Encoding:** Each point in the time series, denoted as $\mathcal{X}^t$, retains its local temporal context through fixed positional embeddings. These embeddings are formulated as follows for any position 'pos' within the sequence:

$$\mathrm{Scale}_{j,x} = (2L_x)^{\frac{2j}{d_{\mathrm{model}}}} \qquad (24)$$

$$\downarrow$$

$$\begin{cases} \mathrm{PE}_{(\mathrm{pos},2j)} = \sin \left[\dfrac{\mathrm{pos}}{\mathrm{Scale}_{j,x}} \right] \\[3mm] \mathrm{PE}_{(\mathrm{pos},2j+1)} = \cos \left[\dfrac{\mathrm{pos}}{\mathrm{Scale}_{j,x}} \right] \end{cases} \qquad (25)$$

where j is an index from 0 to $\lfloor d_{\mathrm{model}}/2 \rfloor$.

2) **Global Timestamp Embedding:** We utilize a set of learnable embeddings for global timestamps. These embeddings account for a variety of significant temporal markers, with a vocabulary designed to recognize 60 unique timestamps up to minute-level precision $\mathrm{SE}_{(\mathrm{pos})}$ tailored to capture broader temporal dynamics.

3) **Feature Vector Transformation and Integration:** Each feature $\mathbf{x}_i^t$ at time t and position i is transformed into a higher dimensional space using a one-dimensional convolutional operation, producing $\mathbf{u}_i^t$. The complete input vector for the model is constructed by combining these transformed feature vectors with both local and global embeddings:

$$\mathcal{X}_{\mathrm{feed}[i]}^t = \alpha \mathbf{u}_i^t + \mathrm{PE}_{(i)} + \sum_p \mathrm{SE}_{(i,p)}, \qquad (26)$$

where α is a scaling factor set to 1 to maintain balance between the original features and the embedded representations.

The visualisation of this process is visible mainly in Fig 1, This new approach not only ensures effective encoding of both local and global temporal details but also enhances the model's capability to forecast with higher accuracy and computational efficiency.

2) Multi-head Distributed Sparse Attention Mechanism (Distributed Informer)

In this study, which explores new dimensions beyond traditional time series forecasting methods such as wind or power forecasting, financial time series forecasting leverages a dual-feature framework. The primary types of features include
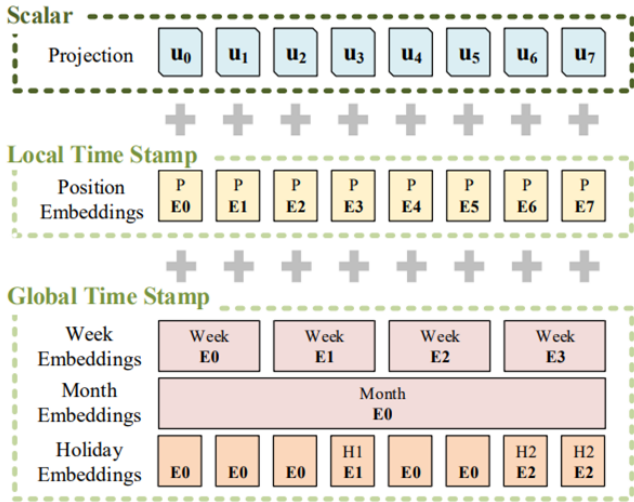
FIGURE 1. The input embedding for the Informer model consists of three distinct components: scalar projection, local timestamp embeddings (positional), and global timestamp embeddings.

Note: The figure is quoted from the literature [35]

economically significant indicators, such as trend indices and oscillation indicators. These are formally represented as follows:

$$\begin{cases} \mathcal{X}^t = \left\{ \mathbf{x}_1^t, \ldots, \mathbf{x}_{L_x}^t \mid \mathbf{x}_i^t \in \mathbb{R}^{d_x} \times 1 \right\} \\ \mathcal{Y}^t = \left\{ \mathbf{y}_1^t, \ldots, \mathbf{y}_{L_y}^t \mid \mathbf{y}_i^t \in \mathbb{R}^{d_y \times 1} \right\} \end{cases} \quad (27)$$

The secondary type involves features extracted directly from the time series through advanced computational techniques such as Fourier transformations. This approach diverges from prior models that typically merged these features without considering their inter-relationships. Our findings highlight that understanding the intricate correlations between these types of indicators is pivotal for enhancing the accuracy of financial forecasts.

During the initial stages of our modeling process, we incorporate three core components: the 'key', 'query', and 'value'. These elements are integrated as follows:

$$[k^{L \times d_k}, q^{L \times d_q}, v^{L \times d_v}] \quad (28)$$

where L represents the sequence length, and d_K, d_Q, and d_V denote the dimensions of the key, query, and value features, respectively. The alignment and integration with the learnable weight matrices are detailed in:

$$[\omega_k^{d_k \times d}, \omega_a^{d_q \times d}, \omega_v^{d_v \times d}] \rightarrow \begin{cases} K^{L \times d} = k w_k \\ Q^{L \times d} = q w_a \\ V^{L \times d} = v w_v \end{cases} \quad (29)$$

This integration fosters a dense and structured data representation, essential for the efficiency of the Distributed In-

former, our advanced adaptation of the traditional multi-head attention mechanism. This model excels in processing sparse datasets and enhances predictive performance significantly. The attention dynamics are quantitatively defined as:

$$\mathcal{A}(\mathbf{q}_i, \mathbf{K}, \mathbf{V}) = \sum_j \frac{k(\mathbf{q}_i, \mathbf{k}_j)}{\sum_l k(\mathbf{q}_i, \mathbf{k}_l)} \mathbf{v}_j = \mathbb{E}_{p(\mathbf{k}_j \mid \mathbf{q}_i)} [\mathbf{v}_j] \quad (30)$$

where

$$p(\mathbf{k}_j \mid \mathbf{q}_i) = \frac{k(\mathbf{q}_i, \mathbf{k}_j)}{\sum_l k(\mathbf{q}_i, \mathbf{k}_l)} \quad (31)$$

utilizes an asymmetric exponential kernel, enhancing the model's sensitivity to critical features. Furthermore, the computational complexity is also addressed:

$$KL(q\|p) = \ln \sum_{l=1}^{L} \frac{e^{\mathbf{q}_i \mathbf{k}_l^\top}}{\sqrt{d}} - \frac{1}{L} \sum_{j=1}^{L_K} \frac{\mathbf{q}_i \mathbf{k}_j^\top}{\sqrt{d}} - \ln L_K \quad (32)$$

The self-attention mechanism refines the model's predictions by evaluating output probabilities, thereby enhancing forecasting accuracy through an optimized attention distribution in sparsely populated data environments.

3) Stacked Informer Structure

The Stacked Informer architecture, as depicted in Fig 2, is an advanced, tiered variant of the multi-layer Informer network specifically engineered for long-term time series forecasting. This structure begins with the original long time-series data, denoted as L. Initially, L is halved to $L/2$, and then quartered to $L/4$. The sequence L is processed through three attention blocks and two convolution layers, while the $L/2$ sequence goes through two attention blocks and one convolution layer. The $L/4$ sequence is handled by a single attention block. Subsequently, features extracted from all sequences at the $L/8$ scale are consolidated into a unified feature map, which is then transmitted to the decoding module.

This Stacked Informer design effectively addresses several critical aspects of long-term forecasting. First, it focuses on efficiently distilling intrinsic temporal features from extensive time-series data. Second, it enhances the robustness and resilience of the forecasts produced. Additionally, the architecture supports parallel processing, thereby improving prediction efficiency. The design utilizes a multi-scale input representation, wherein different temporal scales are processed through distinct pathways. This strategy ensures that the model captures a broad spectrum of features across varying time scales, thereby enriching the informational content essential for long-term forecasting tasks. Moreover, by integrating features from diverse scales into a coherent feature map, the model effectively accounts for interactions and dependencies across these scales.

Each scale within the Stacked Informer employs attention blocks, based on the Transformer architecture, to capture long-distance dependencies within the input sequences.
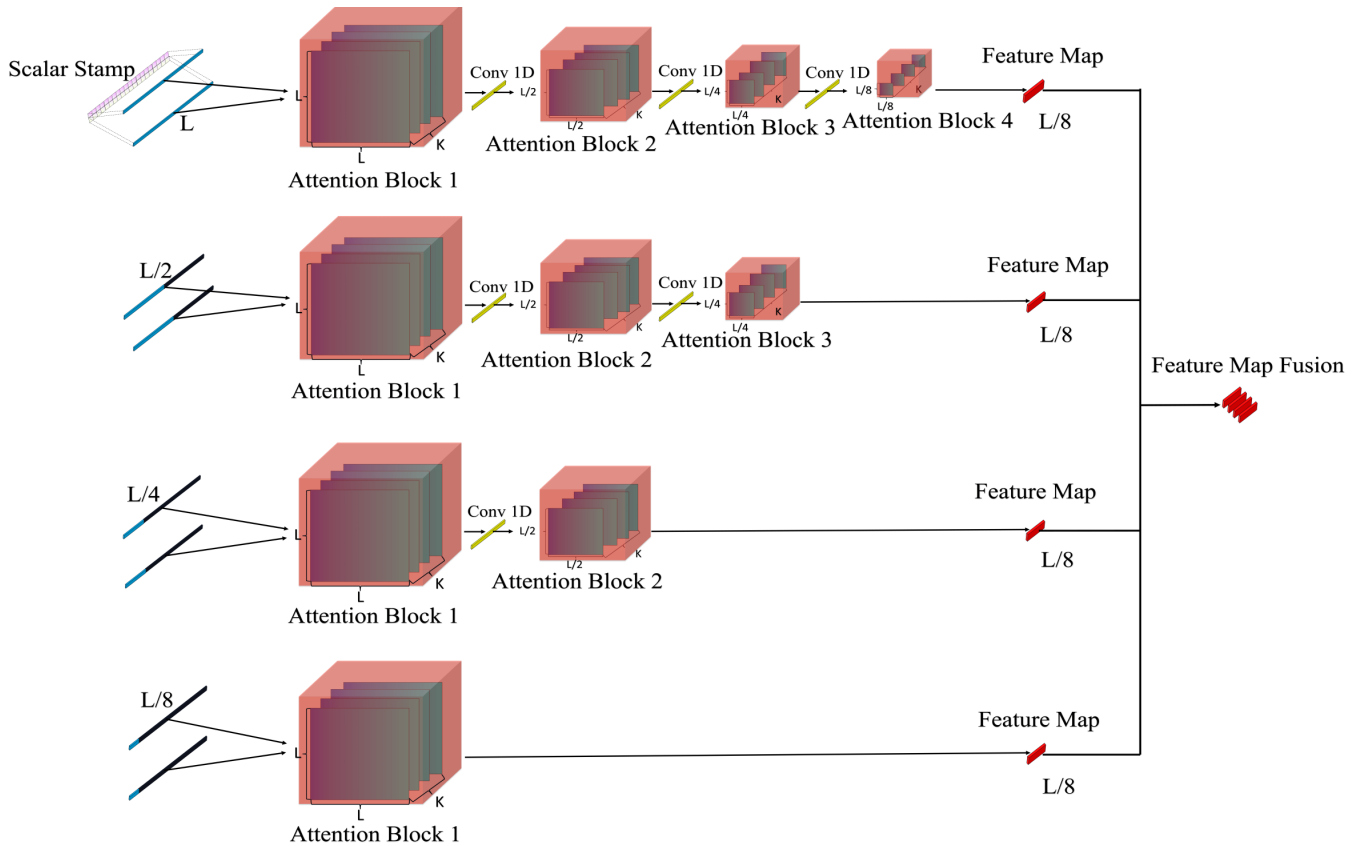
FIGURE 2. Stacked Informer Structure with 4 Layers.

Furthermore, convolutional layers are used to extract local features, which are then integrated with the outputs from the attention blocks to enhance the predictive capabilities of the model.

Overall, the Stacked Informer architecture, with its layered and multi-scale approach, is adept at capturing complex patterns and dependencies within long time series, leading to more precise predictions. This structured design not only reduces the number of attention blocks needed—halving them from four to two for reducing L to $1/32L$—but also shifts the computation from serial to parallel, optimizing for GPU computations. This enhances computational efficiency, integrates global attention mechanisms with local information processing, and minimizes the introduction of noise, thus reducing model complexity and the risk of overfitting. These enhancements allow the Stacked Informer to deliver superior performance in long-term time series forecasting.

D. OTHER METHODS AND OPTIMISATION

1) GC-Enhanced Adam Optimizer

Optimization strategies are of paramount importance in the effective and efficient training of Deep Neural Networks (DNNs). While traditional optimizers such as Stochastic Gradient Descent (SGD) and Adam have been widely adopted in DNNs, this work incorporates the novel Gradient Centralization (GC) technique into the Adam optimizer, aiming to further enhance its efficacy.

The GC technique was pioneered by Yong et al. in 2020 [61]. Unlike other methods that modify the learning rates or adaptively adjust weights, GC directly acts upon the gradients by centralizing them, effectively making their mean values zero. One can draw parallels between GC and projected gradient descent, though the former operates within the constraints of a specific loss function. This centralization is believed to foster a more stable and efficient training process.

Given a weight vector $\mathbf{w}$, with corresponding gradients denoted as $\nabla w_i L$ (where i spans from 1 to N), the GC operator, ϕ_{GC}, is formally defined as:

$$\mu \nabla_{w_i} L = \frac{1}{M} \sum_{j=1}^{M} W_{i,j} L \qquad (33)$$

$$\downarrow$$

$$\phi_{\text{GC}}\left(\nabla_{w_i} L\right) = \nabla_{w_i} L - \mu \nabla_{w_i} L \qquad (34)$$

Where M denotes the dimensionality, and L is the objective function. The essence of this formulation is the subtraction

of the computed mean of column vectors in the weight matrix from each individual column vector, achieving gradient centralization.

For a more matrix-oriented representation, one can express the aforementioned operation as:

$$P = I - ee^{T} \qquad (35)$$

$$\downarrow$$

$$\phi_{\text{GC}} \left(\nabla_{\text{W}} L \right) = P \nabla_{\text{W}} L \qquad (36)$$

In this formulation, $\mathbf{W}$ is the weight matrix, $\mathbf{P}$ is the projection matrix that lies within the hyperplane defined by a normal vector in the weight space, and $\mathbf{P} \nabla \mathbf{W} L$ is the gradient projected onto said hyperplane. Once the centralized gradient $\phi_{\text{GC}}(\nabla \text{W}L)$ is obtained, it can be seamlessly used for weight matrix updates.

One potential enhancement to consider when employing GC with Adam or other adaptive optimizers is the careful tuning of hyperparameters. While GC aims to stabilize the training process, the adaptive nature of optimizers like Adam means that they adjust learning rates based on past gradient moments. Combining these two methodologies might necessitate a recalibration of hyperparameters to achieve optimal results.

2) Dynamic Loss Function

The adaptive loss function, a novel approach to loss function formulation, was pioneered in the work of Barron [62]. This approach integrates a continuous parameter signifying robustness into the traditional loss function. Throughout the optimization phase of model training, this dynamic loss function refines the robustness parameters in tandem with the loss minimization procedure, enhancing prediction precision. The generalized loss function is mathematically described as:

$$f(z, \beta, c) = \frac{|\beta - 2|}{\beta} \left[\left(\frac{z^2}{|\beta - 2|c^2} + 1 \right)^{\frac{\beta}{2}} - 1 \right] \qquad (37)$$

In this formulation, the variable z quantifies the residual between observed and predicted values. The positive scalar $c > 0$ adjusts the curvature of the quadratic component at $x = 0$, while β serves as the adaptable parameter steering robustness.

A detailed examination of the equation reveals that the behavior of the adaptive loss function is intricately tied to variations in β. Different settings of β manifest in the loss function as in the Table 9

It is noteworthy that this adaptive loss function is versatile, enveloping a spectrum of loss functions such as Mean Squared Error (MSE), Cauchy, Charbonnier, and Welsch. This versatility is attributed to the modulation of the β parameter, which grants the flexibility to customize the loss function per specific modeling needs.

TABLE 5. Loss Function Value with $\aleph = \frac{1}{2} \left(\frac{z}{c} \right)^2$

$f(z, \beta, c)$ **Value**	**Condition**
$\aleph$	$\beta = 2$
$\log (\aleph + 1)$	$\beta = 0$
$\sqrt{2\aleph + 1} - 1$	$\beta = 1$
$1 - e^{-\aleph}$	$\beta = -\infty$
$\frac{|\beta - 2|}{\beta} \left[\left(\frac{2\aleph}{|\beta - 2|} + 1 \right)^{\frac{\beta}{2}} - 1 \right]$	Otherwise

Incorporating such a loss function into long-term time series forecasting models can be advantageous. Its adaptability allows for nuanced error handling, potentially leading to improved model robustness and sensitivity to various data patterns, especially in scenarios with non-Gaussian noise or outliers.

3) Evaluation Indicators

The mean absolute error (MAE), mean square error (MSE), root mean square error (RMSE), and coefficient of determination (R^2) are used as indicators for the prediction performance of our Enhanced LFTSformer and all the other models. The formulas of each indicators are as follow in the Table 6:

TABLE 6. Evaluation indicators

Indicator	**Formula**
MAE	$\frac{1}{N} \sum_{t=1}^{N} \left| q_{\text{true}}(t) - q_{\text{pred}}(t) \right|$
MSE	$\frac{1}{N} \sum_{t=1}^{N} \left(q_{\text{true}}(t) - q_{\text{pred}}(t) \right)^2$
RMSE	$\sqrt{\frac{1}{N} \sum_{t=1}^{N} \left(q_{\text{true}}(t) - q_{\text{pred}}(t) \right)^2}$
R^2	$1 - \dfrac{\sum_{t=1}^{N} \left(q_{\text{true}}(t) - q_{\text{pred}}(t) \right)^2}{\sum_{t=1}^{N} \left(q_{\text{true}}(t) - \bar{q} \right)^2}$

In addition, in order to be able to analyse and compare the advantages and disadvantages of the models more intuitively, we will draw a total of three heat maps for MSE, RMSE, and R^2 respectively. When MSE, RMSE is smaller or R^2 is larger, the colour of the heat map will be darker. We will also blend the three heatmaps by colour, and the darker the colour of the resulting blended heatmap means that the corresponding model/stock forecast is better.

IV. EXPERIMENT PROCESS AND RESULTS

A. EXPERIMENT CONDITIONS & CONFIGURATION

1) Experiment Environment

Regarding the experimental configuration, our comprehensive investigations are carried out utilizing a computer system endowed with dual NVIDIA GeForce RTX 4090 GPUs, accompanied by an Intel13th Gen Intel(R)Core(TM)i9-13900K CPU and a robust 128GB RAM.

The operational environment is Wiundows 11 23H2 operating system, while the deep learning framework employed throughout our endeavors remains to be PyTorch.

2) Experiment Data

For the data used in the experiment, we have a few points to note

1) **Data Selection:** Our ablation experiments as well as the comparison experiments are conducted on the data of Pudong Development Bank stock (SHA: 600000). After completing the training of the model, we will conduct stability experiments on the remaining 21 stocks.

2) **Data Segmentation:** For all stock data, we set the first 90% of the data as the training set and the last 10% as the test set. As shown in Fig 3.

3) **Parameters of prediction:** For all stocks and different models, we are predicting the Close parameter, while the other parameters are Indexing and Manipulation as described in Section III-A.

4) **Time Scaling:** Since there are some differences in the time span of different stocks, for the sake of convenience, we set the start time as 0 time unit, the end time as 5000 time unit, and the other data are scaled to the corresponding time unit.

3) Training Parameter Configuration

After rigorous experimentation and meticulous debugging, we have judiciously chosen a set of parameters that optimally enhances the predictive capabilities of our model. The salient parameters integral to our model's architecture are delineated in Table 7.

B. PROCESS OF FEATURE ENGINEERING

1) Determination of K Value and Decomposition of Features

The *VMD-MIC* method enhances the conventional *VMD* approach by utilizing the computation of the *MICyy0* metric. When the original stock dataset of Open/Close/High/Low is input into the *VMD-MIC* framework, it is decomposed into K intrinsic mode functions (IMFs). Figure 4 illustrates the relationship between various K values and their corresponding *MICyy0* results.We can determine the optimum k when the histogram reaches a steady state.

Notably, the *MICyy0* metric remains relatively stable for $K = 8, 10, 12$ and 15. Based on this observation, we make the following selection of K values as shown in the Table8:

TABLE 7. Model Parameters Configuration

Parameter	Value
Input sequence length	256
Prediction sequence length	64-128
Num of encoder layers	8
Num of decoder layers	10
Input size of encoder	5-1
Input size of decoder	5-1
Num of heads	2
Dimension of model	512
Probsparse attention factor	3.8
Early stopping patience	10
Dropout	0.2
Epochs	200

TABLE 8. Selection of K Values

Indicator	Selection of K for Corresponding Indicator
Open	15
High	10
Close	12
Low	8

Based on this premise, we have delineated the IMF values for the four features: close, high, low, and open prices, post-decomposition at varying k values as depicted in Fig5. Correspondingly, we also present their associated Power Spectral Density (PSD) values in Fig6. It is discerned that with the incremental growth of k values, their respective trends largely converge. By conducting an in-depth evaluation of the dominant frequencies in these IMF curves, we can discern the efficiency of the VMD-MIC technique in ensuring precise IMF separation.

2) Dimensionality Reduction of Features

The original stock dataset was decomposed into intrinsic mode functions (IMFs) with allocations of 10 for close, 12 for high, 8 for low, and 15 for open values. Direct integration of these sub-sequences into the predictive model could exacerbate computational overhead. To address this, we employed Feature Extraction (FE) to quantitatively assess the intricacy of the IMFs, subsequently reconstituting them based on their complexity profiles. Through rigorous research, parameters $m = 3$ and $r = 0.3std$ were delineated as optimal, striking a balance between model accuracy and computational efficiency. The corresponding visualization in Fig 7 elucidates the FE values for each IMF.

Upon categorization of IMFs by FE value intervals (e.g., 0.0-0.1, 0.1-0.2, etc.), we crafted new composite elements founded on these FE metrics as shown in the Table 9.
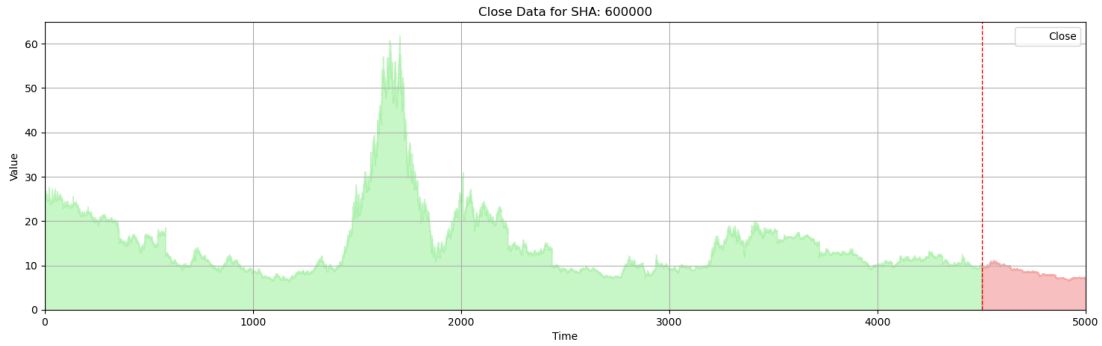
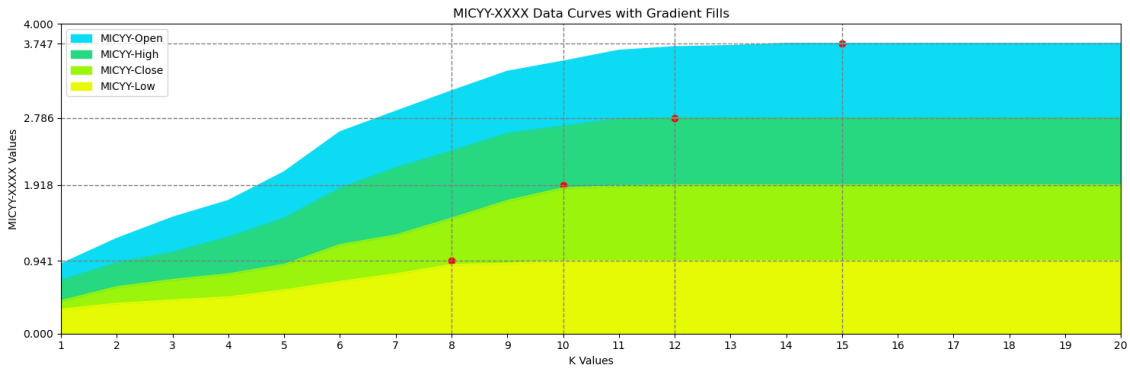**FIGURE 3.** Close Data of SHA: 600000, Split Training and Test Sets at 90% Ratio

FIGURE 4. Relationship Between Different Ks and Their Corresponding MICyy.

TABLE 9. New Composite Elements Founded FE Metrics

New Feature Grouping	FE Value Range	Selection of IMF_N (value of N)
New Feature 1	[0.0,0.1]	3,43
New Feature 2	(0.1,0.2]	2,8,40
New Feature 3	(0.2,0.3]	6,16,21,26,38,39
New Feature 4	(0.3,0.4]	4,5,11,28,30,34,44,45
New Feature 5	(0.4,0.5]	9,42
New Feature 6	(0.5,0.6]	1,7,12,18,19,24,31,32,41
New Feature 7	(0.6,0.7]	13,17,27,29
New Feature 8	(0.7,0.8]	10,14,37
New Feature 9	(0.8,0.9]	15,25,33,35
New Feature 10	(0.9,1.0]	20,22,23,36

Note: New Feature N $= \dfrac{\sum\limits_{\text{Selected N}} IMF_N}{V}$, (V=Number of IMFs selected).

3) Correlation-Based Feature Extraction and Selection

In our endeavor to efficiently extract the residual 25 indicator features, we discerned that the computational efficacy and generalizability of the model could be enhanced by excluding certain irrelevant or redundant attributes from the primary dataset. Consequently, we adopted the Maximal Information

Coefficient (MIC) methodology to scrutinize the interrelations between the remaining 25 stock features and each component, extracting quintessential characteristics of each element through their MIC values. The confusion matrix for MIC is depicted in the subsequent Heatmap of Features (Fig 8).

From the ensuing visualization, it's palpable that each element manifests distinct influential characteristics, encapsulating both holistic correlations and nuanced local attributes. In pursuit of curating features that bear the utmost correlation with the input variables, we instituted an MIC threshold of 0.5 for each component. The outcomes of this selective input feature determination are delineated in the Table 10 below.

C. RESULTS

In this section, in order to assess the superiority of each module in the model and its overall performance, we conduct and present ablation experiments and comparison experiments on SHA: 600000 stocks and their results. In addition, to assess the generality and robustness of the model, we also conduct stability tests on all other 21 stocks to test our model.

1) Ablation Experiment

The purpose of the ablation study is to assess whether complex hybrid models can enhance prediction accuracy
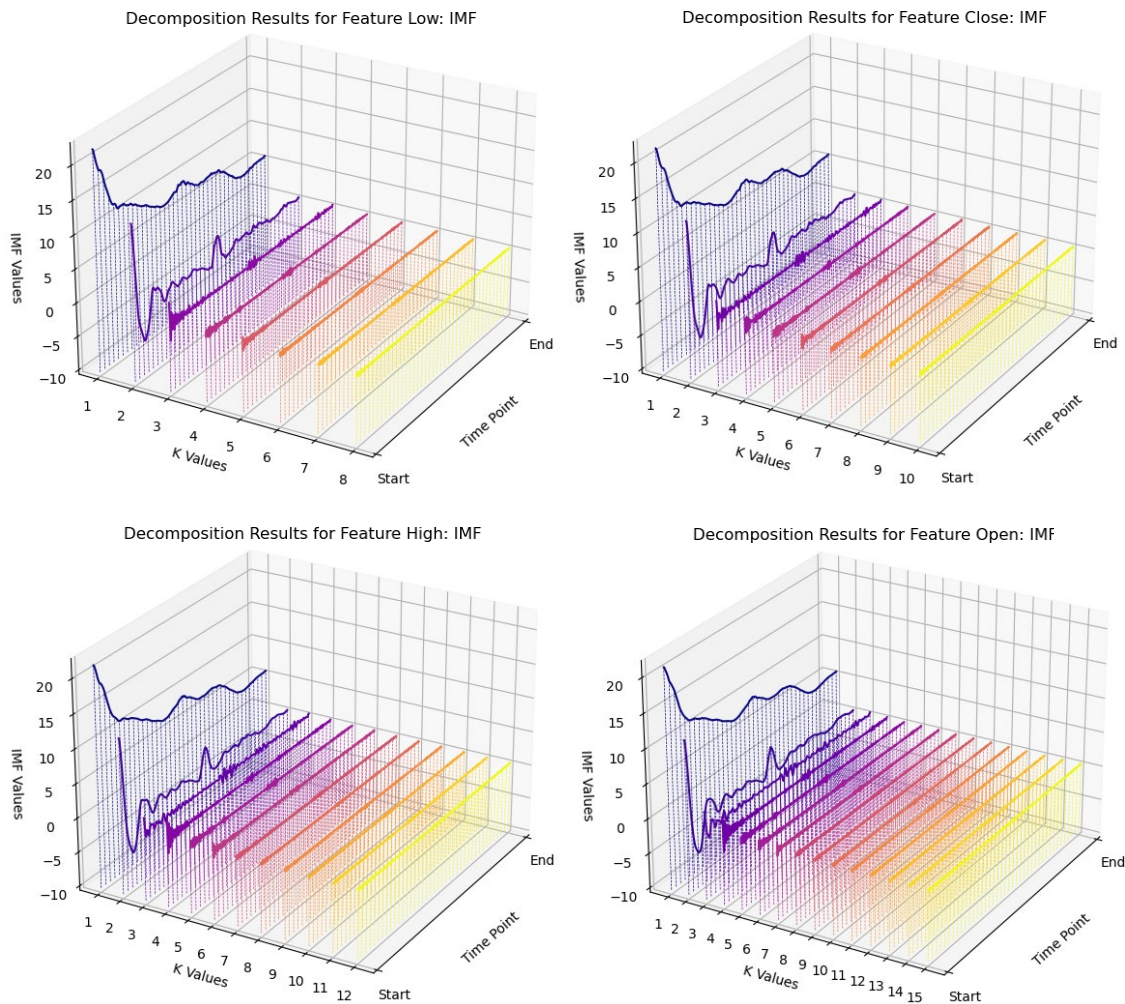
FIGURE 5. The Decomposition Results For Each Feature: IMFs Values for Each K.

over simpler composite and standalone models. We selected benchmark models including the GC-Enhanced Adam+DS Encoder Informer (also referred to as the Enhanced LFTS-former), the Adam+DS Encoder Informer, and the DS Encoder Informer. Parameter optimization was conducted using a grid search method, where the robustness parameter β is adaptively adjusted by the Adam Optimiser. The specific parameters are detailed in Table 7. The sizes of the encoder and decoder inputs are aligned with the number of input variables for the model.

As illustrated in Figure 9, the training results for the four models are presented. It is evident that the GC-Enhanced Adam+DS Encoder Informer achieves stability more quickly and exhibits the lowest error rates in both training and testing phases, aligning with our initial expectations.

Figure 10 depicts the results of our ablation experiment. In the top half, we present the predicted outcomes and trends over 500 time units for the bottom 10% of the dataset. To effectively compare the performance between models and to showcase the performance of individual models, we include integrated prediction plots for all models as well as separate prediction plots for each model. For the micro-scale analysis, as illustrated in the bottom half of Figure 10, we have selected a smaller time scale (highlighted by the red box in the figure). This particular segment was chosen because the prediction ranges of the models are most closely aligned during this period, covering a total range of 30 minutes × 200 data points for analysis.

In terms of error quantification, we have employed a paradigm incorporating both heatmaps and blended heatmaps to visualize the values of MAE/MSE and R^2, as depicted on the right side of Fig11. Furthermore, to articulate the absolute error, Box Plots are utilized to delineate the distribution and range of the residuals—predicted values subtracted from actual values—and their outliers, as demonstrated on the right side of Fig12. Synthesizing the above, the following analysis can be articulated:

1) **Comprehensive Assessment of Model Performance**: Empirical evidence suggests that the Enhanced LFTS-former model attains a predictive accuracy exceed-
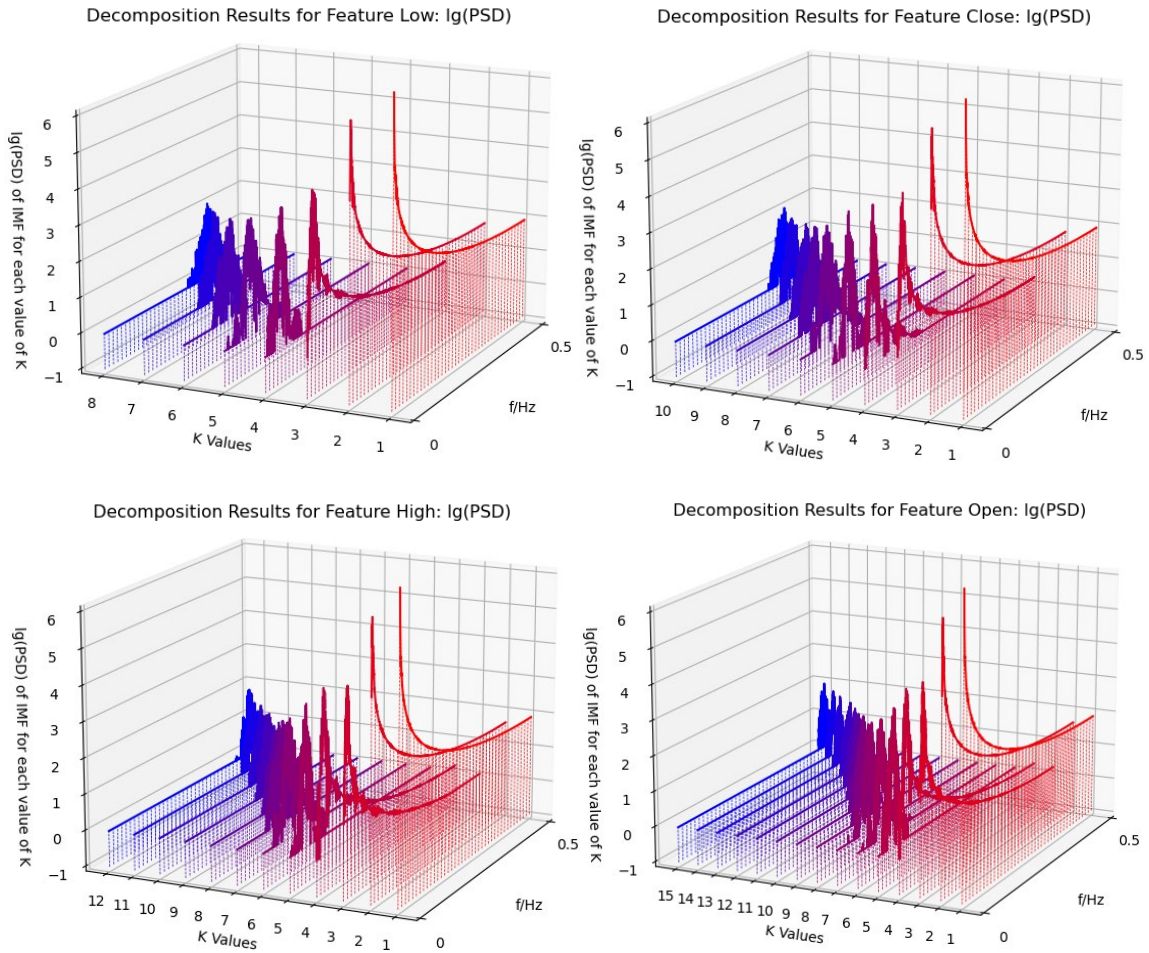
FIGURE 6. The Dcomposition Results For Each Feature: Power Spectral Density (PSD) for Each Corresponding IMFs.
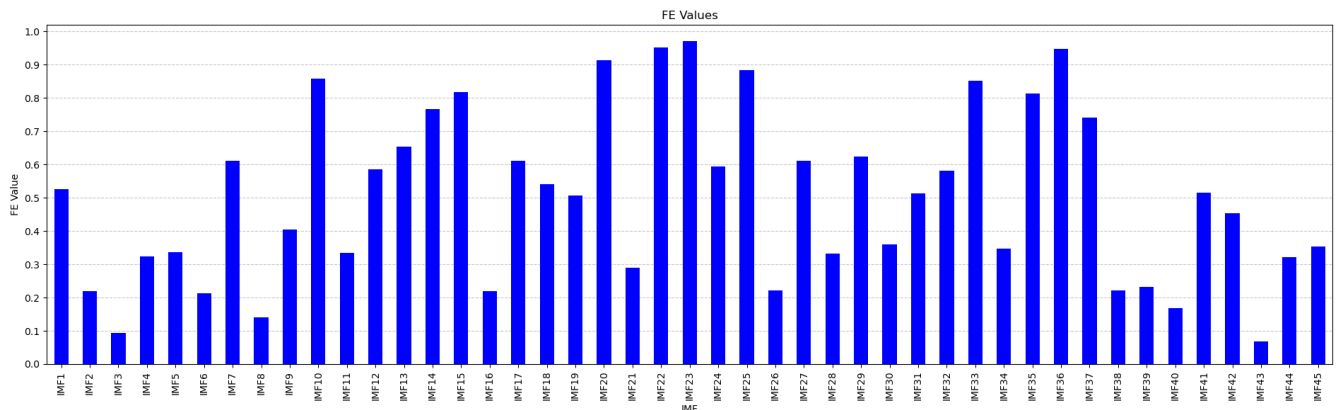

FIGURE 7. FE Values for 45 IMFs.

ing 90% on a holistic scale. Notwithstanding a discernible decrement in accuracy during the transition from macroscopic to microscopic scale analysis, the model's utility and contribution to the field cannot be overstated.

2) **Merits of the Adaptive Function in Prediction**: The volatile nature of financial markets, often accentuated by inflection points, poses significant predictive challenges. The incorporation of GC-Enhanced methodology has demonstrably augmented the model's pre-
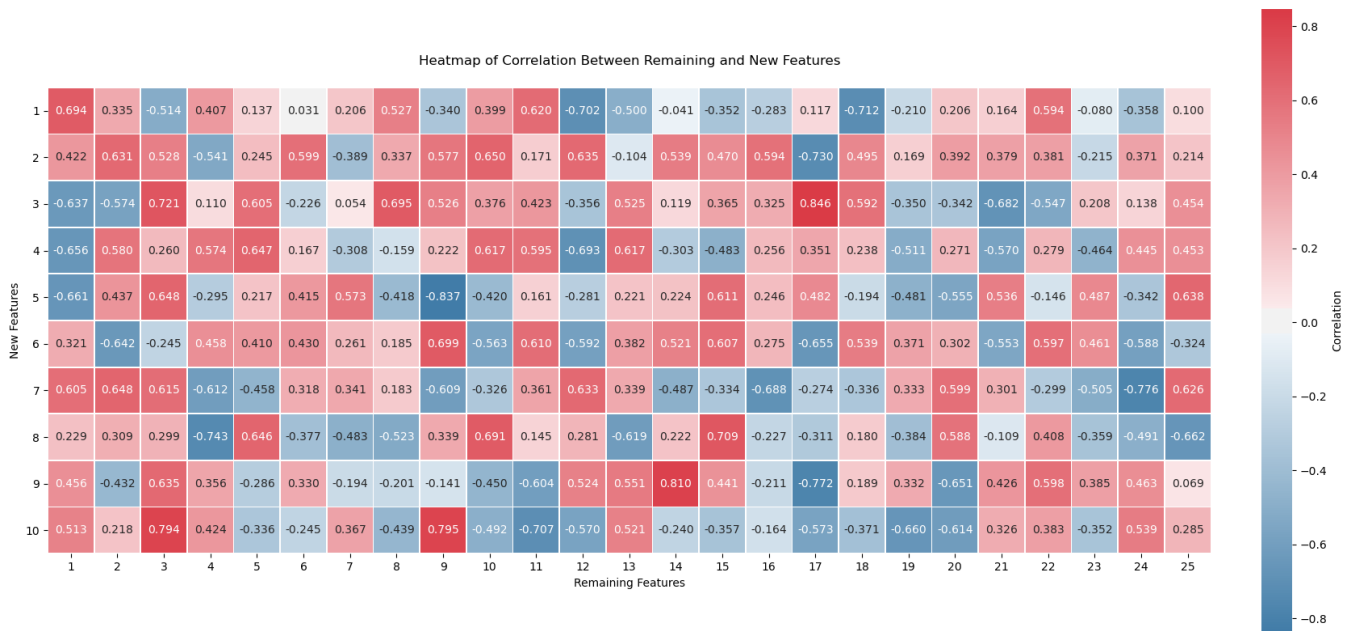
FIGURE 8. Heatmap For Features.
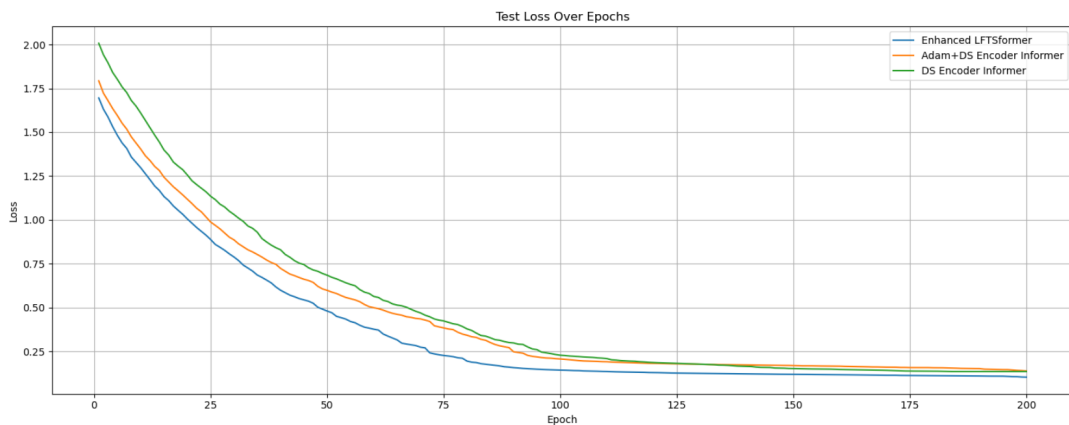
FIGURE 9. Traning Results For Each Model in Ablation Experiments.

cision. Furthermore, it substantiates the efficacy of adaptive functions in mitigating the impact of market volatilities, thereby reducing predictive errors and uncertainties.

3) **Efficiency of Training and the Role of Loss Function**: The pivotal role of a tailored loss function in optimizing both the efficiency and the veracity of model training is incontrovertible. The deployment of our novel adaptive loss function has appreciably curtailed the temporal expenditure required for training the Enhanced LFTSformer model, simultaneously fortifying its robustness and predictive verisimilitude.

Upon contrasting various models, it becomes evident that the Enhanced LFTSformer model possesses the capability to decompose raw stock data into finer granules, thereby delving deeper into the intrinsic dynamics and features of

stocks. This intricate exploration furnishes financial decision-makers with predictions that are both precise and timely.

2) Comparison Experiment

To rigorously assess the overall superiority of the proposed model, comparative experiments were conducted. The selection of benchmark models includes not only the classical CNN and LSTM and the foundational Informer but also other state-of-the-art long temporal series forecasting models based on modifications of the Informer/Transformer, such as Autoformer [63] and Pyraformer [64]. Training was carried out on the SHA: 600000 dataset, with all models, including the Enhanced LFTSformer, sharing an identical set of training parameters. The forecast outcomes are presented in Fig13.

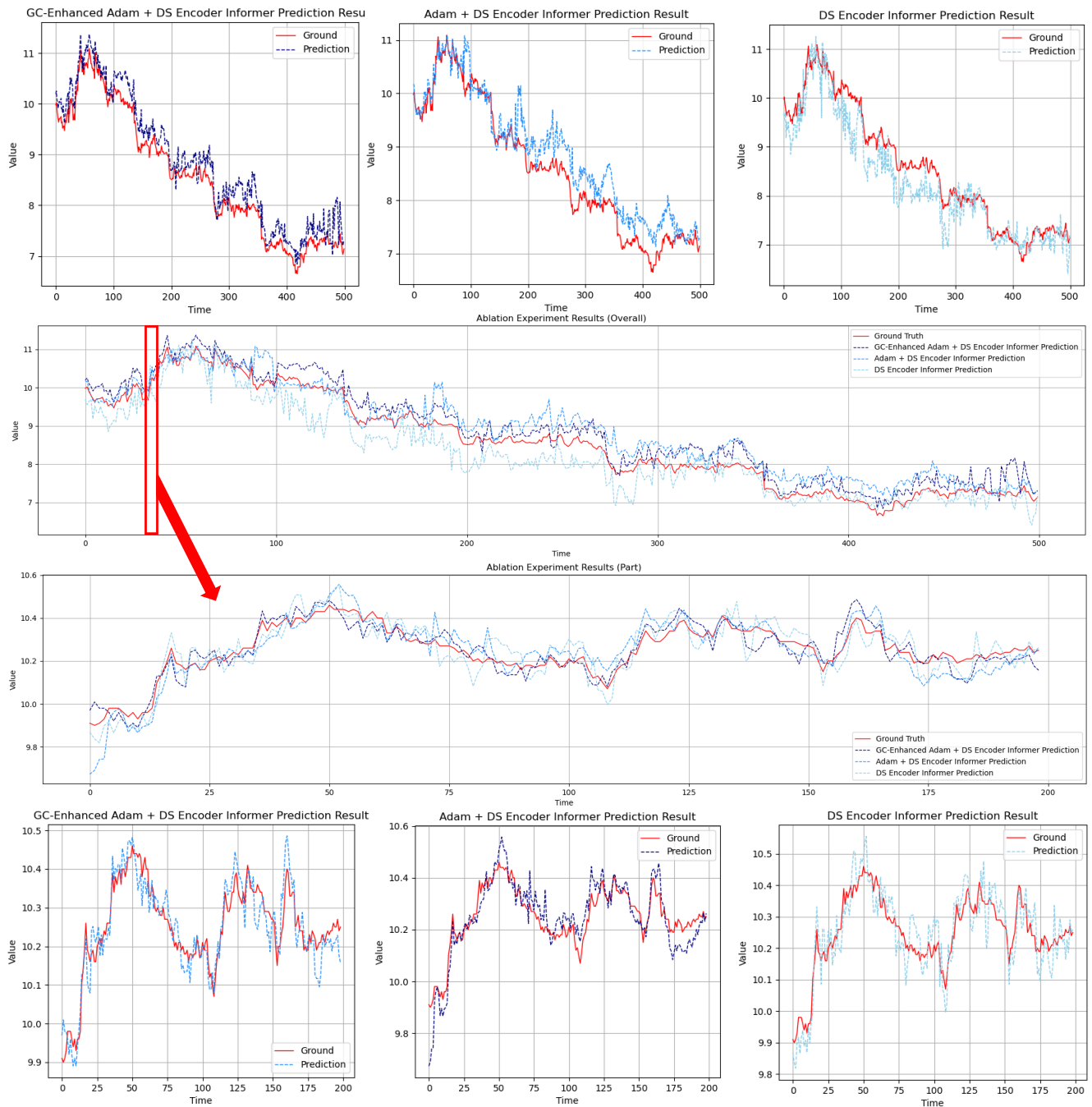For the comparative analysis of errors across different

FIGURE 10. Results of Ablation Experiments for the Whole Test Set and Partial Test Set

models, the same methodologies were employed—heatmaps, blended heatmaps, and Boxplots—as shown on the left side of Figs11 and 12. From this, we can draw the following conclusions:

1) **Outstanding Performance**: Across all metrics, the Enhanced LFTSformer demonstrates improvements over traditional models and basic informer constructs in long-term financial time series forecasting. Although

on certain metrics such as MAE/MSE, the performance of Enhanced LFTSformer is marginally lower compared to Pyraformer and Autoformer, the blended heatmap analysis indicates that the Enhanced LFTSformer exhibits deeper colors, signifying superior outcomes within a more holistic evaluation framework.

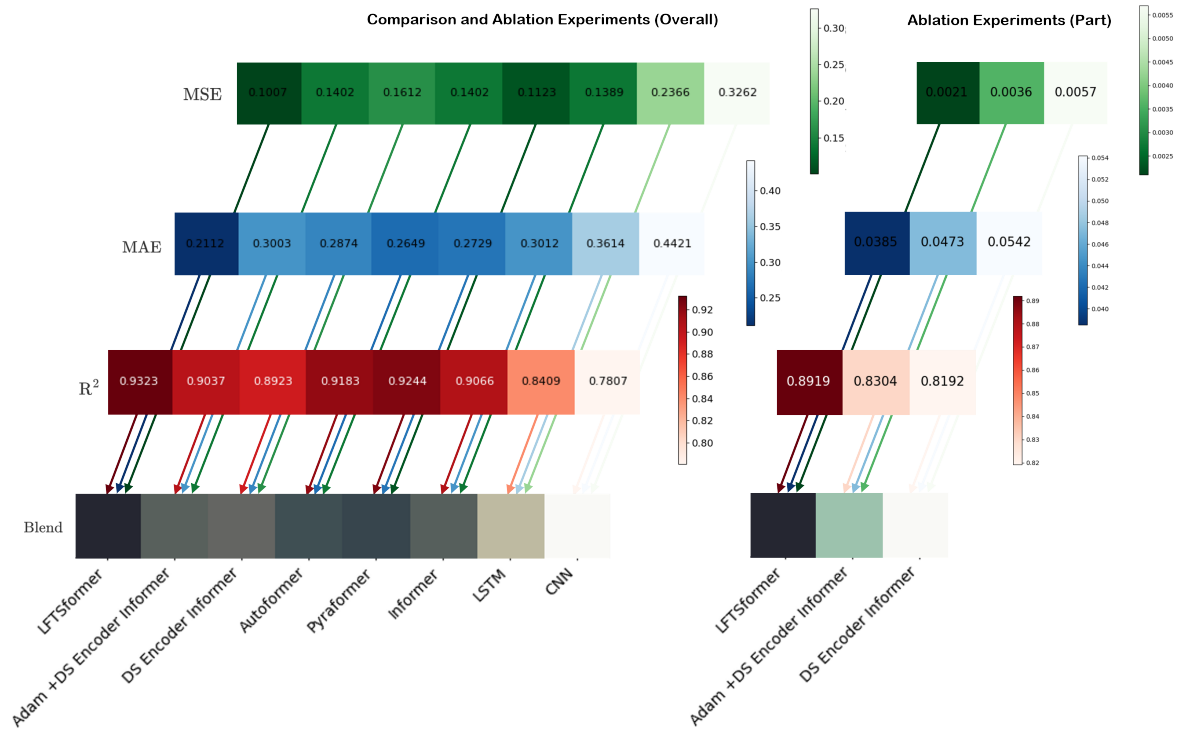2) **Decomposition Capability**: Curve analysis under similar data processing conditions reveals that the VMD

FIGURE 11. Heat Map for MAE/MSE, R^2 and Their Mix for Different Models in Ablation Experiments and Comparison Experiments. (Darker colours in the blend heat map represent a better combination of predicted performance)
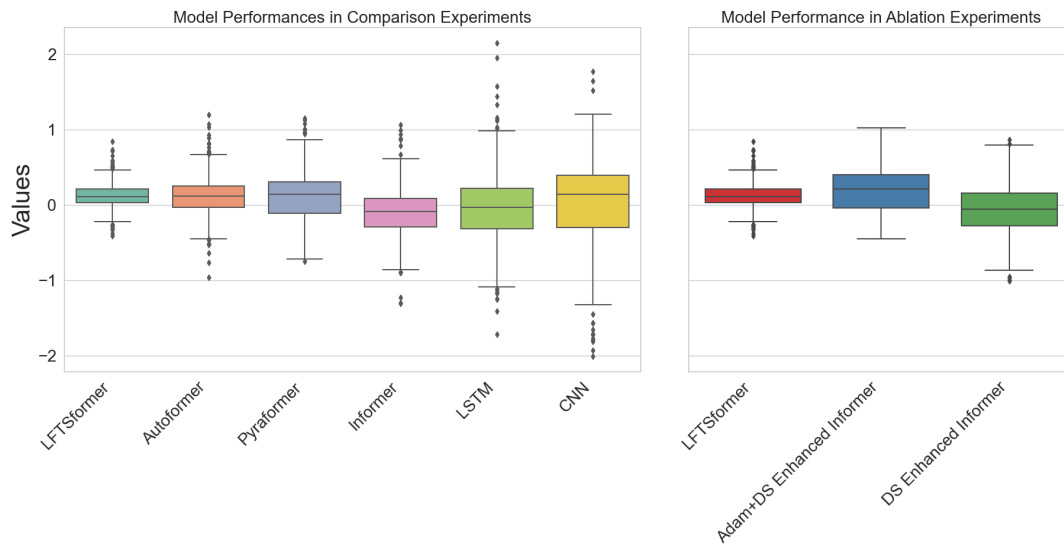

FIGURE 12. Box Plot of $q_{real} - q_{prediction}$ for Different Models in Ablation Experiments and Comparison Experiments

algorithm exhibits enhanced data decomposition capabilities over conventional algorithms like EMD. This is evidenced by its effectiveness in attenuating non-smooth features within the original data, thus improving the accuracy of stock predictions.

3) **Long-Term Predictive Proficiency**: Further data examination indicates divergences in curve fitting among different models as the temporal scale of prediction extends. While most curves initially maintain a high

degree of fit, they begin to deviate from actual values over time. However, the Enhanced LFTSformer maintains its predictive accuracy, underscoring its capacity to address the challenges of long-term sequential forecasting.

In conclusion, Enhanced LFTSformer represents a superior predictive combination model, holding significant practical implications, especially in domains like long-term investment decisions, risk management, and asset allocation.
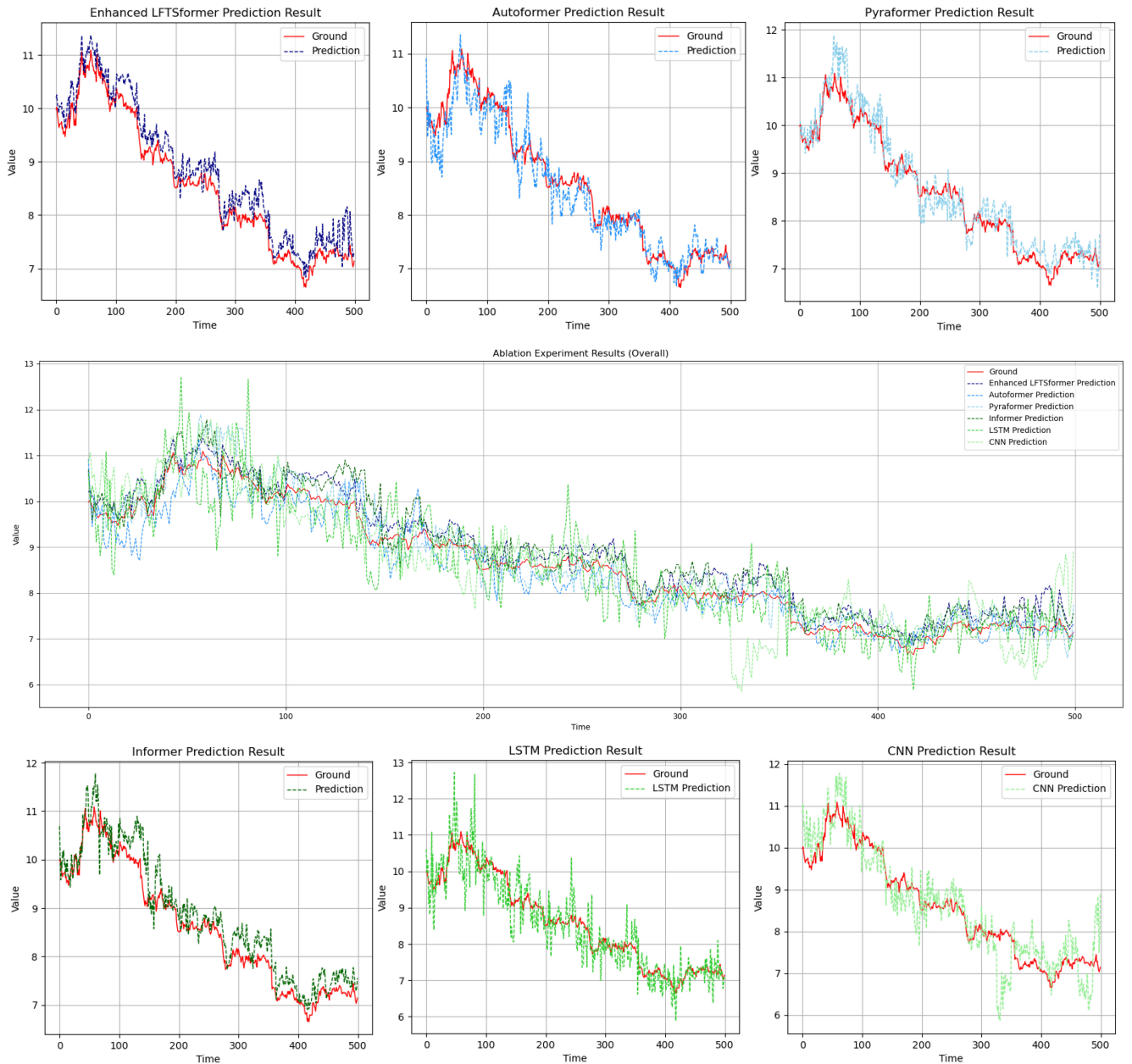
FIGURE 13. Results of Comparison Experiments for the Whole Test Set

3) Stability Experiment

Previous experimental outcomes have established that the Enhanced LFTSformer delivers superior performance on the SHA: 600000 dataset compared to other benchmark models, demonstrating its formidable capacity for long-term sequence prediction. Nevertheless, stock prices, as intricately complex financial time series indices, exhibit unpredictability due to shifts in industrial domains, unforeseen events, fluctuations in market sentiment, and the interplay of chaotic factors.

To further ascertain the model's stability and adaptability,

it was retrained and its predictions reassessed on an additional 21 stocks listed on the Shenzhen or Shanghai Stock Exchanges, with all training configurations and parameters remaining consistent. The results are illustrated in Fig14 (due to space constraints, only 16 out of 21 are randomly showcased). Error heatmaps and blended heatmaps are displayed in Fig15.

Analysis of the tabulated results indicates that the Enhanced LFTSformer achieved commendable predictive outcomes across various stocks in different industry sectors. Nonetheless, it is imperative to acknowledge that our model
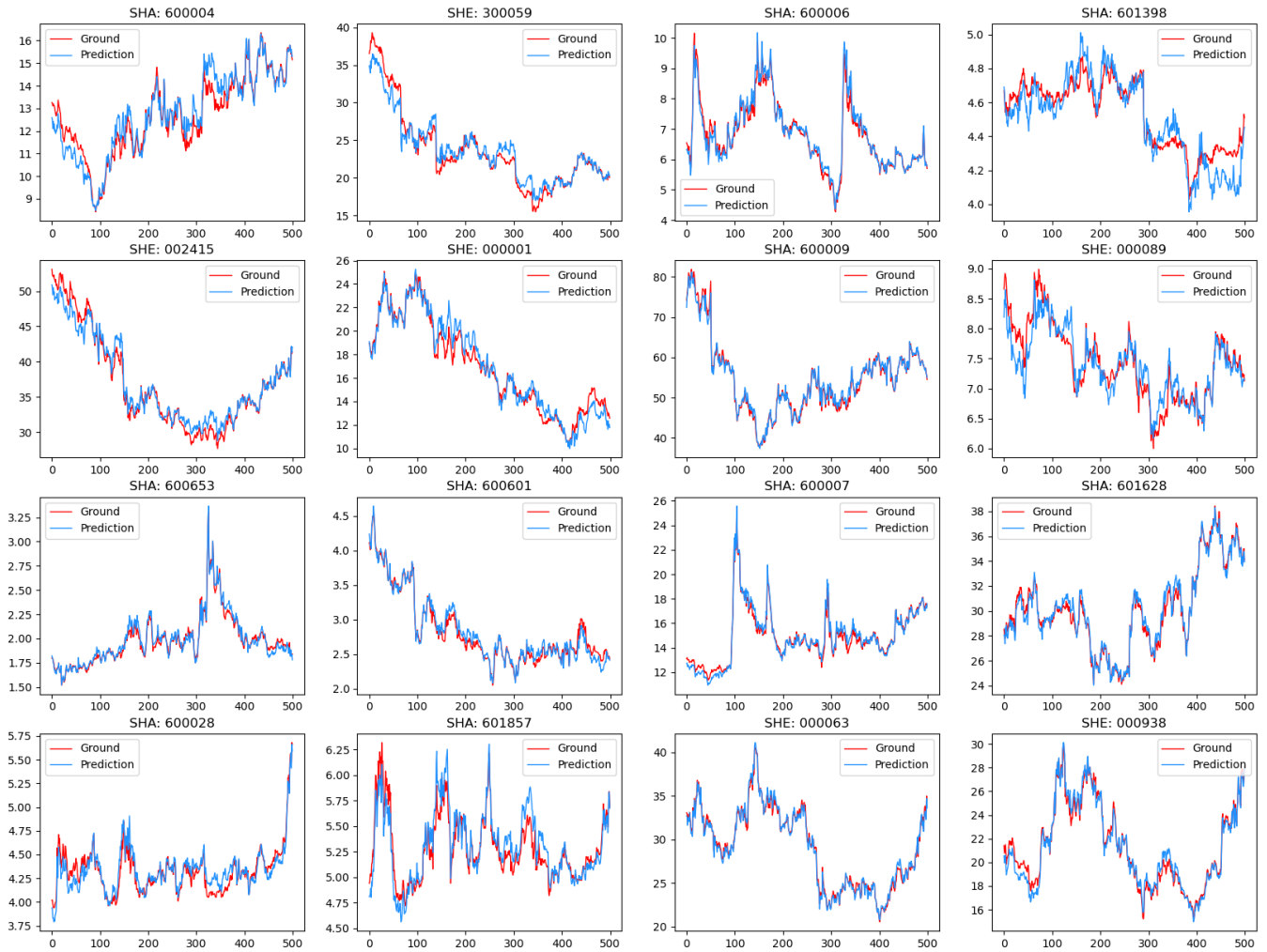
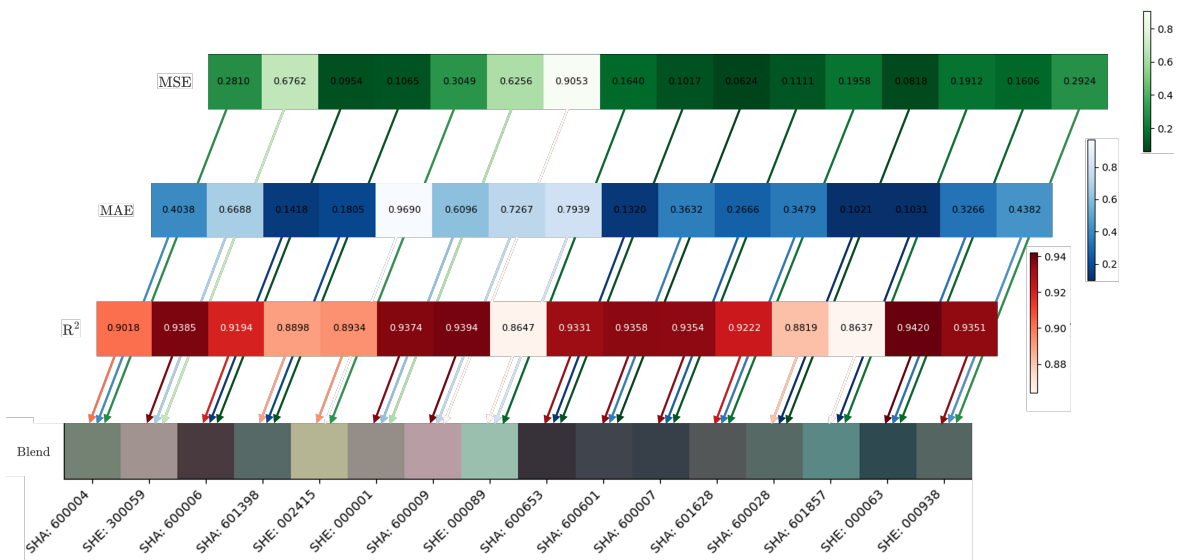**FIGURE 14.** Results of Stability experiments for the whole test set

FIGURE 15. Heat Map for MAE/MSE, R^2 and Their Mix for Different Stocks in Stability Experiments. (Darker colours in the blend heat map represent a better combination of predicted performance)

TABLE 10. Reconstructed Feature Values Based On Heatmap

Reconstruct Features	Included Remaining Feature M (M value)
Reconstruct Feature 1	1, 8, 11, 22
Reconstruct Feature 2	2, 3, 6, 9, 10, 12, 14, 16
Reconstruct Feature 3	3, 5, 8, 9, 13, 17, 18
Reconstruct Feature 4	2, 4, 5, 10, 11, 13
Reconstruct Feature 5	3, 7, 15, 21, 25
Reconstruct Feature 6	9, 11, 14, 15, 18, 22
Reconstruct Feature 7	1, 2, 3, 12, 20, 25
Reconstruct Feature 8	5, 10, 15, 20
Reconstruct Feature 9	3, 12, 13, 14, 22
Reconstruct Feature 10	1, 3, 9, 13, 24

Note 1: The features corresponding to all areas greater than 0.5 in the heatmap are selected into the corresponding group.

Note 2: $RCF_N = NF_N * \left(\prod_{\text{Included } M} RMF_M * C_{NM} \right)$

Abbreviation	Meaning
RCF_N	Reconstruct Feature N
NF_N	New Feature N
RMF_M	Remaining Feature M
C_{NM}	Correlation Between NF_N and RMF_M

exhibited significant predictive performance declines in certain stocks, such as SHE: 000089 (Shenzhen Airport) and SHA: 600009 (Shanghai Airport), which are characterized by market capitalizations not exceeding 100 billion RMB and are more susceptible to singular factors.

V. CONCLUSIONS & FUTURE WORK
A. CONCLUSIONS
The operation of stock markets is influenced by a plethora of factors, including economic conditions, seasonal changes, and the global financial climate. These factors may introduce a multitude of anomalies and non-smooth characteristics into the data of stock markets, posing substantial challenges to improving the accuracy and performance of stock price forecasts. Consequently, this paper proposes an adaptive hybrid model for stock prediction, leveraging an enhanced VMD, Feature Engineering (FE), Stacked & Distributed (DS) Informer, and an adaptive loss function. The Enhanced LFTSformer model, as a promising hybrid model, is capable of adaptively predicting inherently volatile stock data. The primary advantages of this methodology are summarized as follows:

1) **Hybrid Integration**: The Enhanced LFTSformer represents a comprehensive model that integrates advanced feature engineering, an improved encoder in the Informer, and GC-Enhanced optimization. It amalgamates the strengths of various techniques, delivering superior accuracy and robustness. This superiority is especially prominent in long-term time series forecast-

ing.

2) **Data Decomposition**: Compared to conventional decomposition methods, the MIC-enhanced VMD excels at extracting non-linear features from raw data, effectively augmenting data quality and streamlining the forecasting process.

3) **Adaptive Response**: A comprehensive analysis of experimental results illustrates the adaptive loss function's rapid response to volatile financial time series data, skillfully managing anomalies and predicting evolutionary trends.

4) **Generalization Capability**: Experiments on additional stock datasets highlight the exemplary predictive performance of our proposed model, confirming the Enhanced LFTSformer's significant generalization capability and its promising prospects in stock data forecasting.

In conclusion, hybrid stock prediction models that integrate the advantages of multiple algorithms demonstrate increased predictive accuracy and robustness. Notably, their effectiveness is sustained in long-term time series forecasting.

B. FUTURE WORK
While this study has made preliminary advancements, several areas warrant further investigation to deepen our understanding and enhance the practical applications of stock market prediction models. The following research directions are proposed:

1) **Predictive Modeling for Small Enterprises**: A notable limitation of this study is its focus predominantly on large-scale enterprise stocks. The stock prices of smaller firms, however, may exhibit greater sensitivity and vulnerability to unforeseen events. Future research should explore the integration of natural language processing and real-time event analysis to more accurately forecast stock price fluctuations of these smaller entities. It is crucial to develop specific parameters or variables that can effectively capture the nonlinear impacts of such events on stock prices.

2) **Non-linear Optimization in Feature Engineering**: Another limitation identified in this study is the primarily linear methodology employed in our feature engineering. Future research could investigate non-linear techniques, such as activation functions in neural networks or dynamically adjusted weight coefficients, to perpetually optimize the generation of feature values during iterative learning. Such approaches are anticipated to enhance model performance and adaptability.

3) **Further Optimizations and Refinements**: This study has highlighted two main areas for potential enhancement. First, our feature selection was limited to correlation factors, overlooking important metrics such as the F-score and sensitivity. Subsequent studies should incorporate these metrics to evaluate the relationships

between variables and stock characteristics more thoroughly, aiming to reduce data redundancy. Second, the selection of parameters in our models may lack precision and comprehensiveness. Future studies should consider the application of optimization algorithms to address the sensitivity issues inherent in the parameter selection of deep learning networks.

By exploring these proposed avenues, we anticipate further improvements in the precision, utility, and comprehensiveness of stock market prediction models. These advancements will not only propel academic research in the financial domain but also provide practitioners with more accurate decision-support tools, facilitating navigation through the complex challenges of financial markets.

REFERENCES

[1] Analysis of Financial Time Series, John Wiley & Sons, 2005.
[2] The (Mis)Behaviour of Markets: A Fractal View of Risk, Ruin, and Reward, Profile Books, 2010.
[3] Introduction to Time Series and Forecasting, Springer, 2002.
[4] The Empirical Mode Decomposition and the Hilbert Spectrum for Nonlinear and Non-Stationary Time Series Analysis, Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences, 1998, 454(1971), pp. 903-995.
[5] Financial Time Series Forecasting with Deep Learning: A Systematic Literature Review: 2005–2019, Applied Soft Computing, 2020, 90: 106181.
[6] A Review on Singular Spectrum Analysis for Economic and Financial Time Series, Statistics and Its Interface, 2010, 3(3), pp. 377-397.
[7] Wind Power Generation Forecast Based on Multi-Step Informer Network, Energies, 15(18), 2022, pp. 6642, MDPI.
[8] Improving the Accuracy of Multi-Step Prediction of Building Energy Consumption Based on EEMD-PSO-Informer and Long-Time Series, Computers and Electrical Engineering, 110, 2023, pp. 108845, Elsevier.
[9] Load Forecasting of District Heating System Based on Informer, Energy, 253, 2022, pp. 124179, Elsevier.
[10] Financial Analysis, Planning, and Forecasting, In: Essentials of Excel VBA, Python, and R: Volume II: Financial Derivatives, Risk Management and Machine Learning, Springer, 2023, pp. 433–455.
[11] Effects of the Validation Set on Stock Returns Forecasting, Expert Systems with Applications, 150, 2020, pp. 113271, Elsevier.
[12] Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation, Econometrica: Journal of the Econometric Society, 1982, pp. 987–1007, JSTOR.
[13] Generalized Autoregressive Conditional Heteroskedasticity, Journal of Econometrics, 31(3), 1986, pp. 307–327, Elsevier.
[14] The Statistical Properties of Daily Foreign Exchange Rates: 1974–1983, Journal of International Economics, 24(1-2), 1988, pp. 129–145, Elsevier.
[15] Nelson, C.R. Inflation and Rates of Return on Common Stocks. The Journal of Finance, 1976, Vol. 31, No. 2, pp. 471-483.
[16] Brown, Robert Goodell. Smoothing, Forecasting and Prediction of Discrete Time Series. Courier Corporation, 2004.
[17] Wang, Jujie; Cui, Quan; Sun, Xin; He, Maolin. Asian Stock Markets Closing Index Forecast Based on Secondary Decomposition, Multi-Factor Analysis and Attention-Based LSTM Model. Engineering Applications of Artificial Intelligence, Vol. 113, 2022, Article 104908, Elsevier.
[18] Wang, Fang; Li, Menggang; Mei, Yiduo; Li, Wenrui. Time Series Data Mining: A Case Study with Big Data Analytics Approach. IEEE Access, Vol. 8, 2020, pp. 14322–14328, IEEE.
[19] Herwartz, Helmut. Stock Return Prediction under GARCH—An Empirical Assessment. International Journal of Forecasting, Vol. 33, No. 3, 2017, pp. 569–580, Elsevier.
[20] Sutton, R. S. Learning to Predict by the Methods of Temporal Differences. Machine Learning, 1988, 3: 9-44.
[21] Mahfoud, S., Mani, G. Financial Forecasting Using Genetic Algorithms. Applied Artificial Intelligence, 1996, 10(6): 543-566.
[22] Kim, K., Han, I. Genetic Algorithms Approach to Feature Discretization in Artificial Neural Networks for the Prediction of Stock Price Index. Expert Systems with Applications, 2000, 19(2): 125-132.

[23] Trafalis, T. B., Ince, H. Support Vector Machine for Regression and Applications to Financial Forecasting. Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks. IJCNN 2000. Neural Computing: New Challenges and Perspectives for the New Millennium. IEEE, 2000, 6: 348-353.
[24] Lee, M. C. Using Support Vector Machine with a Hybrid Feature Selection Method to the Stock Trend Prediction. Expert Systems with Applications, 2009, 36(8): 10896-10904.
[25] Chen, K., Zhou, Y., Dai, F. A LSTM-based Method for Stock Returns Prediction: A Case Study of China Stock Market. In Proceedings of the 2015 IEEE International Conference on Big Data (Big Data). IEEE, 2015, pp. 2823-2824.
[26] Li, Z. M., Xu, S. W., Cui, L. G., et al. The Short-term Forecast Model of Pork Price Based on CNN-GA. Advanced Materials Research, 2013, 628, pp. 350-358.
[27] Ding, X., Zhang, Y., Liu, T., et al. Deep Learning for Event-Driven Stock Prediction. In Proceedings of the Twenty-fourth International Joint Conference on Artificial Intelligence, 2015.
[28] Dase, R. K., Pawar, D. D. Application of Artificial Neural Network for Stock Market Predictions: A Review of Literature. International Journal of Machine Intelligence, 2010, Vol. 2, No. 2, pp. 14-17.
[29] Dambalkar, S. S., Shirure, O. S., Bankar, V. K., et al. An In-Depth Review on Stock Market Prediction Through Deep Learning. no, 2023, 4: 1695-1699.
[30] Shah, D., Isah, H., Zulkernine, F. Stock Market Analysis: A Review and Taxonomy of Prediction Techniques. International Journal of Financial Studies, 2019, Vol. 7, No. 2, Article 26.
[31] Vaswani, A., Shazeer, N., Parmar, N., et al. Attention Is All You Need. Advances in Neural Information Processing Systems, 2017, 30.
[32] Alshemali, B., Kalita, J. Improving the Reliability of Deep Neural Networks in NLP: A Review. Knowledge-Based Systems, 2020, 191: 105210.
[33] Han, K., Wang, Y., Chen, H., et al. A Survey on Vision Transformer. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 45(1): 87-110.
[34] Wen, Q., Zhou, T., Zhang, C., et al. Transformers in Time Series: A Survey. arXiv preprint arXiv:2202.07125, 2022.
[35] Zhou, H., Zhang, S., Peng, J., et al. Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting. Proceedings of the AAAI Conference on Artificial Intelligence, 2021, 35(12): 11106-11115.
[36] Pomponi, Jary; Scardapane, Simone; Uncini, Aurelio. Structured Ensembles: An Approach to Reduce the Memory Footprint of Ensemble Methods. Neural Networks, Vol. 144, 2021, pp. 407–418, Elsevier.
[37] Tung, Frederick; Mori, Greg. CLIP-Q: Deep Network Compression Learning by In-Parallel Pruning-Quantization. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 7873–7882.
[38] Yang, Heng-Li; Lin, Han-Chou. An Integrated Model Combined ARIMA, EMD with SVR for Stock Indices Forecasting. International Journal on Artificial Intelligence Tools, Vol. 25, 2016, Article 1650005, World Scientific.
[39] Alfonso, Gerardo; Carnerero, A Daniel; Ramirez, Daniel R; Alamo, Teodoro. Stock Forecasting Using Local Data. IEEE Access, Vol. 9, 2020, pp. 9334–9344, IEEE.
[40] Bukhari, Ayaz Hussain; Raja, Muhammad Asif Zahoor; Sulaiman, Muhammad; Islam, Saeed; Shoaib, Muhammad; Kumam, Poom. Fractional Neuro-Sequential ARFIMA-LSTM for Financial Market Forecasting. IEEE Access, Vol. 8, 2020, pp. 71326–71338, IEEE.
[41] Büyükşahin, Ümit Çavuş; Ertekin, Şeyda. Improving Forecasting Accuracy of Time Series Data Using a New ARIMA-ANN Hybrid Method and Empirical Mode Decomposition. Neurocomputing, Vol. 361, 2019, pp. 151–163, Elsevier.
[42] Ren, S., Wang, X., Zhou, X., et al. A Novel Hybrid Model for Stock Price Forecasting Integrating Encoder Forest and Informer. Expert Systems with Applications, 2023, 234: 121080.
[43] Ding, Y., Zhang, C., Zhang, C. An Informer Based Method for Stock Intraday Price Prediction. In Proceedings of the 2023 19th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), IEEE, 2023: 1-6.
[44] Liu, H., Chen, D., Wei, W., et al. Long-term Stock Price Forecast Based on PSO-Informer Model. In Proceedings of the Fifth International Conference on Computer Information Science and Artificial Intelligence (CISAI 2022), SPIE, 2023, 12566: 288-294.
[45] Sababipour Asl, G. Stock Volatility Forecasting with Transformer Network. Journal of 2023.

[46] Abdulsahib, H. M., Ghaderi, F. GLAD: Global–Local Approach; Disentanglement Learning for Financial Market Prediction. IET Signal Processing, 2023, 2023.

[47] Xiong, F. Research on Stock Price Prediction Model and Quantitative Trading Based on INFORMER Mechanism. Master's Thesis, Dalina Jiaotong University, 2023. DOI:10.26990/d.cnki.gsltc.2023.000058.

[48] Feng, Y. Research on Stock Quantitative Trading Algorithm Based on Informer. Master's Thesis, Dalina Jiaotong University, 2022. DOI:10.27734/d.cnki.ggdsx.2022.000963.

[49] Wang, X., Xia, M., Deng, W. MSRN-Informer: Time Series Prediction Model Based on Multi-Scale Residual Network. IEEE Access, 2023, 11: 65059-65065.

[50] Wang, N., Zhao, X. Enformer: Encoder-Based Sparse Periodic Self-Attention Time-Series Forecasting. IEEE Access, 2023, 11: 112004-112014.

[51] Liu, M., Wang, W., Hu, X., et al. Multivariate long-time series traffic passenger flow prediction using causal convolutional sparse self-attention MTS-Informer. Neural Comput & Applic, 2023, 35: 24207–24223.

[52] Yang, Z., Liu, L., Li, N., Tian, J. Time Series Forecasting of Motor Bearing Vibration Based on Informer. Sensors, 2022, 22(15): 5858.

[53] Guo, L., Li, R., Jiang, B. A Data-Driven Long Time-Series Electrical Line Trip Fault Prediction Method Using an Improved Stacked-Informer Network. Sensors, 2021, 21(13): 4466.

[54] Li, F., Wan, Z., Koch, T., Zan, G., Li, M., Zheng, Z., Liang, B. Improving the accuracy of multi-step prediction of building energy consumption based on EEMD-PSO-Informer and long-time series. Computers and Electrical Engineering, 2023, 110: 108845.

[55] Achelis, S. B. Technical Analysis from A to Z. 2001.

[56] Kara, Y., Boyacioglu, M. A., Baykan, Ö. K. Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul Stock Exchange. Expert Systems with Applications, 2011, 38(5): 5311-5319.

[57] Guresen, E., Kayakutlu, G., Daim, T. U. Using artificial neural network models in stock market index prediction. Expert Systems with Applications, 2011, 38(8): 10389-10397.

[58] Dragomiretskiy, K., Zosso, D. Variational mode decomposition. IEEE Transactions on Signal Processing, 2013, 62(3): 531-544.

[59] Lin, G., Lin, A., Gu, D. Using support vector regression and K-nearest neighbors for short-term traffic flow prediction based on maximal information coefficient. Information Sciences, 2022, 608: 517-531.

[60] Chen, W., Zhuang, J., Yu, W., Wang, Z. Measuring complexity using fuzzyen, apen, and sampen. Medical Engineering & Physics, 2009, 31(1): 61-68.

[61] Yong, H., Huang, J., Hua, X., Zhang, L. Gradient centralization: A new optimization technique for deep neural networks. In: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part I, vol. 16, pp. 635-652. Springer, 2020.

[62] Barron, J. T. (2019). A general and adaptive robust loss function. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 4331-4339). IEEE/CVF.

[63] Chen M., Peng H., Fu J., et al. (2021). Autoformer: Searching Transformers for Visual Recognition. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 12270-12280). IEEE/CVF.

[64] Liu S., Yu H., Liao C., et al. (2021). Pyraformer: Low-complexity Pyramidal Attention for Long-range Time Series Modeling and Forecasting. In International Conference on Learning Representations.

HONGYI DUAN (M'2023) Hongyi Duan was born in Shanghai, China, on September 2, 2003. He is currently pursuing the Bachelor's degree in computer science and artificial intelligence at Xi'an Jiaotong University, Xi'an, Shaanxi, People's Republic of China, expected in 2025. Mr. Duan has been an IEEE Student Member since 2023.

Despite being early in his academic career, Mr. Duan has already amassed numerous research achievements. Among his numerous publications, two are particularly representative: "Application and Analysis of Machine Learning Based Rainfall Prediction," presented at the 2023 8th International Conference on Intelligent Computing and Signal Processing (ICSP), doi: 10.1109/ICSP58490.2023.10248891; and "Comparative study of microgrid optimal scheduling under multi-optimization algorithm fusion," presented at the 2023 10th International Forum on Electrical Engineering and Automation (IFEEA), doi: 10.1109/IFEEA60725.2023.10429466.

⋯
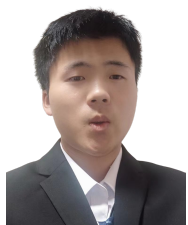
JIAN'AN ZHANG (M'2021) Zhang Jian'an was born on June 21, 2003, in Pingxiang, Jiangxi Province, China. He is currently pursuing a Bachelor's degree at the School of Mathematics, Shanghai University of Finance and Economics, with a minor in Finance at Fudan University. Since 2021, he has been a student member of IEEE. He has presented papers at several international computer conferences and has served as a presenter at ICAACE, as well as a reviewer for related papers. His primary research focuses on using deep learning techniques to solve practical problems such as time series analysis and option pricing