

Adaptive Hybrid Model for Enhanced Stock Market Predictions Using Improved VMD and Stacked Informer

Jianan Zhang¹ and Hongyi Duan^{2*}

¹School of Mathematics, Shanghai University of Finance and Economics University, Wuchuan Street, Yangpu District, 200437, Shanghai, China.

²Faculty of Electronic and Information Engineering, Xi'an Jiaotong University, West Xianning Road, Xi'an, 710049, Shannxi, China.

*Corresponding author(s). E-mail(s): dann.hiroaki@ieee.org;
Contributing authors: zjaqifei@ieee.org;

Abstract

This paper introduces an innovative adaptive hybrid model for stock market predictions, leveraging the capabilities of an enhanced Variational Mode Decomposition (VMD), Feature Engineering (FE), and stacked Informer integrated with an adaptive loss function. Through rigorous experimentation, the proposed model, termed Adam+GC+enhanced informer (We name it VMGC-former), demonstrates significant proficiency in addressing the intricate dynamics and volatile nature of stock market data. Experimental results, derived from multiple benchmark datasets, underscore the model's superiority in terms of prediction accuracy, responsiveness, and generalization capabilities over traditional and other hybrid models. The research further highlights potential avenues for optimization and introduces future directions to enhance predictive modeling, especially for small enterprises and feature engineering.

Keywords: Stock Market Predictions, Adaptive Hybrid Model, Variational Mode Decomposition, Stacked Informer, Feature Engineering

1 Introduction

Financial markets play a pivotal role in global economic activities, and their operations and dynamic evolutions are intricately linked to a myriad of chaotic and complex factors, including economic configurations, seasonal components, and the international milieu [1] [2]. As the economy progresses and financial markets expand continuously, time series analysis in finance has become indispensable [3]. This analytical approach has significantly advanced the understanding of market dynamics, refined intelligent decision-making processes, and bolstered developments in forecasting investment returns [4][2]. Consequently, it has garnered immense scholarly attention, leading to abundant research contributions in this domain.

In stark contrast to conventional time series prediction endeavors characterizing various scientific domains—such as the temporal allocation mechanisms associated with wind energy integration [5], the granular analysis of protracted energy consumption patterns in architectural structures [6], or the intricate forecasting of load dynamics within thermal frameworks [7]—the sphere of financial time series forecasting is imbued with an elevated level of complexity and unpredictability. This intricacy is accentuated, especially when the temporal projection is extrapolated over extended horizons or when faced with the advent of unanticipated exogenous shocks [8][9]. Yet, even within such a convoluted milieu, advancements have been palpable, courtesy of a myriad of innovative methodologies. These encompass the integration of sophisticated multi-source multi-instance learning paradigms [10], rigorous exploration and juxtaposition of an array of deep learning architectures [11], innovative techniques that transmute traditional time series data into visually interpretable images [12], and the avant-garde strategies leveraging models like ChatGPT to refine decision-making paradigms [13].

In the field of financial econometrics, particularly within the realm of stock predictions, historical literature suggests that numerous scholars have constructed various forecasting models based on econometric and statistical theories. For instance, Robert Engle, while investigating UK inflation rates, introduced the Autoregressive Conditional Heteroscedasticity (ACH) method[14]. This method was later refined and extensively utilized as the ARCH approach[15]. Additionally, David A. Hsieh, when forecasting the fluctuations of five foreign exchange rates, adopted and enhanced statistical models with time-varying mean and variance[16].

With the rapid expansion of financial markets and the explosive growth of the economy in the 21st century, correspondingly, an increasing number of statistical-based financial time series forecasting models have been proposed and continuously refined. In 2004, RG Brown presented the ARIMA model[17], which subsequently found widespread application in the forecasting of both stock and futures markets. Its derivatives include models such as AR, VAR, ARDL, ARCH, GARCH, MIDAS, and other generalized variants[18]. When stock price movements tend to be linear, these models have demonstrated commendable performances in assessing stock prediction risks[19] and forecasting returns[20]. However, for stocks exhibiting frequent fluctuations and influenced by multiple variables, the efficacy of these models for predictions remains a topic of contention.

To address this challenge, with the advent of artificial intelligence, numerous foundational machine learning models and methods (e.g., CNN, ANN, LSTM, GNN, Informer, BP, SVM, etc.) have been progressively incorporated into the forecasting of financial and stock time series. Notably, the primary areas of optimization and research are ensemble methods based mainly on stacking[21], network structure fusion rooted in combined parameter pruning[22][35], and decomposition-reconstruction approaches based on VMD and EMD[24][25] [26].

It is noteworthy that hybrid models have demonstrated significant potential in stock market forecasting. For instance, the model developed by Qiu et al., which integrates RNN, LSTM, and GRU techniques, has exhibited remarkable performance[27]. Jin and colleagues incorporated investor sentiment and leveraged deep learning with attention mechanisms to further enhance accuracy[28]. Moreover, studies have combined convolutional neural networks with LSTM, among other models, to construct more comprehensive forecasting systems[29][30][31]. In recent years, with the dominance of the Transformer architecture in the machine learning realm and the introduction of the Informer model tailored for long-term time series forecasting with the Transformer[32], numerous researchers have embarked on stock predictions using hybrid models based on either the Transformer or the Informer. Notably, Zhang et al. proposed an attention network based on the Transformer architecture to capture the temporal characteristics of financial data, underscoring the intricacies of stock forecasting[33].

In light of the significant accomplishments and potential avenues for enhancement achieved by hybrid models in financial time series forecasting, we seek to further optimize in the following dimensions: enhancing the model’s sensitivity to volatile data, ensuring sustained performance in long-term time series forecasting, and elevating the prediction accuracy. Consequently, we introduce a model amalgamating Adam, GC, and an enhanced version of Informer, which we christen as VMGCformer.

The subsequent organization of this paper is as follows: In the section2, we provide a comprehensive review of the latest advancements in hybrid models for stock and other financial time series forecasting. In the section3, we delve into the research methodologies and the process of model construction utilized in this study. The section4 presents our feature engineering workflow, along with the experimental execution and results. The concluding section5 encapsulates the findings of this research, placing emphasis on potential avenues for further refinement.

2 Related works

In this paragraph we present a brief overview of the different fusion models proposed and innovated in the field of financial time series forecasting and stock prediction, as well as their recent progress in practical applications

2.1 hybrid of Machine Learning and Econometrics

With the rise of machine learning and the high degree of sophistication and refinement of traditional econometric models, the fusion of machine learning and econometrics is also a major research direction in financial time-series forecasting. Kim and Won used

a hybrid model that incorporates a long-short-term memory network (LSTM) and a generalized autoregressive conditional heteroskedasticity model (GARCH) to study the volatility of stock prices[34], and their results show that this hybrid model has a significant improvement in forecasting accuracy. Meanwhile, Yang and Lin apply the empirical mode decomposition (EMD) technique and combine it with ARIMA and support vector regression (SVR) for more accurate financial time series forecasting[35]

2.2 Multi-Hybrid Model Based on Machine Learning Multi-Module Optimization

Hybrid modeling is not limited to the fusion of econometric and machine learning models, but in recent years, the most widely adopted and researched is the fusion of different deep learning or machine learning. For example, Luo et al. not only fused multiple linear and nonlinear models, but also integrated advanced machine learning techniques such as recurrent neural networks (RNN) to achieve accurate prediction of the direction of the FTSE 100 closing price in the face of the non-stationary nature of stock data. [36] This multi-technology fusion approach breaks through the limitations of a single model and provides a new perspective for the analysis of complex financial time series. Further, inspired by quantum computing, Wang et al. proposed a model [37] combining periodic empirical mode decomposition (PEEMD) and quantum neural network (QNN), which opens up new possibilities for financial time series forecasting.

2.3 Diversity and Wide Application of Hybrid Models

The diversity and flexibility of hybrid models make them highly applicable in a variety of financial markets, and their continuous research and wide application have achieved great results and progress. The more typical ones, such as Iftikhar et al. in the day-ahead price forecasting of Brent crude oil, demonstrate the potential application of hybrid models in different financial products[38]. Also for example, Wei Y L utilized the Adaptive Network Fuzzy Inference System (ANFIS) model to achieve excellent forecasting results on the Taiwan Stock Exchange Weighted Equity Index (TAIEX) and Hang Seng Stock Index (HSI)[39]. Meanwhile, other researchers such as Rubio and Alba[40], Liu et al.[41] and Singh et al.[42] have also successively proposed various hybrid models, which have achieved remarkable results in reducing forecasting errors and improving forecasting accuracy.

3 Overview of Research Methodology

3.1 Data Processing

3.1.1 Data Indexing

In our study, we procured stock price data for China National Petroleum Corporation (CNPC) from the official records of the Shanghai Stock Exchange, with the stock code sh.601857. The dataset encompasses stock fluctuations from November 6, 2007, to March 21, 2023, with a granularity of 48 data points per day. We constructed our dataset using four distinct metrics: opening and closing prices for each time frame,

along with the highest and lowest prices within that period, resulting in 716164 data entries. We allocated the initial 90% of the data as the training set and designated the final 10% as the test set. All prices are denominated in Renminbi (RMB). The specifics are illustrated in the subsequent Fig1

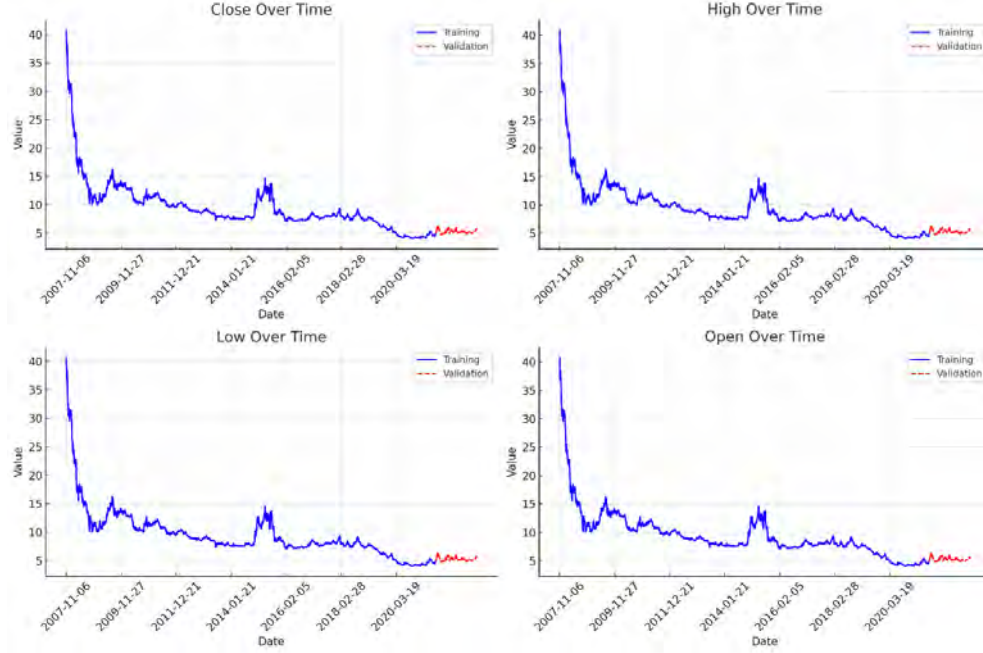


Fig. 1: Data set

Then we meticulously selected a total of 29 indicators. These are bifurcated into fundamental and technical types, encompassing 8 and 21 metrics respectively. Fundamental Indicators:

- Opening price
- Closing price
- Peak price within the designated period
- Trough price within the same interval
- Trading volume in Stock (unit specification contingent on the context)
- Aggregate trading volume
- Weighted price
- Discrepancy between the highest and lowest prices

In the realm of Technical Indicators, our choices were informed by the seminal works of references [43], [44], and [45]. A more detailed tabulation of these indicators will be provided in the subsequent Table 1.

The indicators encompass a diverse set of financial metrics, each designed to encapsulate various facets of stock price dynamics. This includes five distinct return rates

Table 1: Features Used in this Model

Feature name	Number of lags/window size	Number of features
Open, Close, High, Low	1	4
Weighted Price	2	2
Volume(Stock), Volume(Currency)	1	2
High - Low	1	2
Return	from 1 to 5	5
Correlation MA5 and MA30	1	1
Sum3, Sum5	5,3	2
Sum5 - Sum3	1	1
RSI(6, 14)	6,14	2
Rate of change	9,14	2
Williams R	1	1
CCI	1	1
DEMA	1	1

with different lag intervals, a weighted moving average, correlation coefficients between weighted moving averages at various lags, cumulative prices over different timeframes, and other relevant metrics. To elucidate on select indicators: The *Average True Range* (*ATR*) is computed as:

$$ATR = MA(TR, N) \quad (1)$$

where $MA(TR, N)$ is the moving average of TR over N periods, effectively representing the average price over the last N intervals. The true range at the t^{th} period, TR_t , is given by:

$$TR_t = \max(High_t - Low_t, High_t - Close_{t-1}, Close_{t-1} - Low_t) \quad (2)$$

Here, *High*, *Low*, and *Close* denote the highest, lowest, and closing prices for the period, respectively. The *Relative Strength Index* (*RSI*) is defined as:

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

When the price exhibits an upward or downward movement, the following conditions apply:

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

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

For the *Price Return*, the formula is:

$$Return_{i_t} = Close_t - Close_{t-i} \quad (6)$$

The *Rate of Change (ROC)* is given by:

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

Additionally, the *Stochastic Oscillator*, which gauges momentum in the Bitcoin price, is calculated as:

$$K = 100 \frac{Close - Low}{High - Low} \quad (8)$$

$$\text{Stochastic oscillator} = MA(K, N) \quad (9)$$

Lastly, the *CCI (Commodity Channel Index)*, reflecting average price deviation, measures the mean deviation degree given a fixed price interval. It is formulated as:

$$CCI = \frac{TP - MC}{MD \times 0.015} \quad (10)$$

$$TP = \frac{1}{3}(High + Low + Close) \quad (11)$$

$$MC = MA(Close, N) \quad (12)$$

$$MD = MA(MC - Close, N) \quad (13)$$

3.1.2 Data Manipulation

Taking into account the inherent volatility of stock prices in their chronological series, a differencing method was employed. Fig 2 and Table 2 representations elucidate the dataset's autocorrelation. Parallely, the Augmented Dickey-Fuller (ADF) unit root test was executed, validating the series' stability post differencing.



Fig. 2: Autocorrelation Graph

A comprehensive descriptive statistical analysis was undertaken on all indicators, and both the raw and differenced data sets were subjected to the ADF examination. All differenced metrics successfully met the test criteria.

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:

Table 2: ADF Nnit Root Test

<i>Dickly – Fuller</i>	-19.329573
<i>p – value</i>	0.000000

$$new\ close_i = \log\left(\frac{close_i}{close_{i-1}}\right) \quad (14)$$

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

$$new\ value_i = \frac{value_i - E(value)}{\sqrt{D(value)}} \quad (15)$$

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

3.2 Model Architecture and Building

In the process of constructing our stock price prediction model, a series of advanced methodologies were adopted to ensure both accuracy and robustness in forecasting. Initially, the VMD-MIC model was employed for feature decomposition and extraction, ensuring that clear and meaningful information is derived from intricate datasets. Subsequently, Fuzzy Entropy was applied for feature engineering, further refining the feature extraction process and enhancing the model’s generalization capabilities. For the prediction phase, we utilized the stacked Informer framework, an amalgamation of various techniques, adept at capturing the dynamic fluctuations in stock prices with precision. To accurately assess the model’s performance, a Dynamic Loss Function was designed, and in tandem with the Enhanced GC + Adam Optimizer, the model is fine-tuned to ensure optimal prediction outcomes in practical applications.

3.2.1 VMD-MIC Model Framework

Variance Mode Decomposition (VMD) is a sophisticated data decomposition technique, facilitates the transformation of time-domain financial sequences into the frequency domain[46]. This transformation yields a set of K intrinsic mode functions (IMFs). The foundational step in this procedure entails formulating the variational constraint equation:

$$\begin{cases} \min \left\{ \sum_{K=1}^K \left\| \partial_t \left[\left(\delta(t) + \frac{i}{\pi t} \right) * u_K(t) \right] e^{-i w_K t} \right\|_2^2 \right\} \\ \text{s.t. } \sum_{K=1}^K u_K = x(t) \end{cases} \quad (16)$$

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 :

$$\begin{aligned} L(\{u_K\}, \{w_K\}, \lambda) = & \alpha \sum_{K=1}^K \left\| \partial_t \left[\left(\delta(t) + \frac{i}{\pi t} \right) * u_K(t) \right] e^{-iw_K t} \right\|_2^2 \\ & + \left\| x(t) - \sum_{K=1}^K u_K(t) \right\|_2^2 \\ & + \left\langle \lambda(t), x(t) - \sum_{K=1}^K u_K(t) \right\rangle \end{aligned} \quad (17)$$

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:

$$\hat{u}_K^{n+1}(w) = \frac{\hat{x}(w) - \sum_{i \neq K} \hat{u}_i(w) + (\hat{\lambda}(w)/2)}{1 + 2\alpha(w - w_K)^2} \quad (18)$$

$$w_K^{n+1} = \frac{\int_0^\infty w |\hat{u}_K^{n+1}(w)|^2 dw}{\int_0^\infty |\hat{u}_K^{n+1}(w)|^2 dw} \quad (19)$$

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

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)[47]. This integration determines the optimal number of decompositions, K , enhancing the decomposition's adaptability.

The degree of decomposition is judiciously determined via the MIC value between the original sequence, denoted as y , and its reconstructed counterpart, y_0 . A MIC_{yy_0} value closer to one indicates a higher fidelity in the VMD decomposition, translating to minimal information loss and an optimal decomposition result.

3.2.2 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 [48]. 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:

$$D(x) = e^{-\ln(2)\left(\frac{x}{r}\right)^2} \quad (20)$$

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 $i, j = 1, 2, \dots, n - m + 1$, taking an average over each i in D_{ij}^m produces the average similarity function:

$$\phi^m(r) = \frac{1}{N - m + 1} \sum_{i=1}^{N-m+1} \left(\frac{1}{N - m} \sum_{j=1, j \neq i}^{N-m+1} D_{ij}^m \right) \quad (21)$$

Subsequently, the FE can be expressed as:

$$FuzzyEn(m, r, n) = \ln \left(\frac{\Phi^m(r)}{\Phi^{m+1}(r)} \right) \quad (22)$$

3.2.3 stacked Informer structure

In many practical applications, accurate prediction of long sequence time series (LSTF) is paramount, especially in the context of financial time series forecasting. Such predictions offer crucial early warnings for financial markets, ensuring their stability and safety. Commonly employed models, such as Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM), and Transformers, have been applied in forecasting tasks[49][50][51]. However, due to architectural constraints, their performance in LSTF scenarios leaves room for improvement. Several studies [32][52] have indicated that the inherent design of these models struggles to capture long-range dependencies between inputs and outputs, limiting their effectiveness in LSTF tasks.

To address this challenge, Zhou et al. introduced a novel LSTF model based on the Transformer architecture, termed "Informer" [32]. Compared to traditional approaches, the Informer boasts notable advantages:

- (1) It incorporates the probSparse self-attention mechanism, significantly enhancing sequence alignment capabilities.
- (2) Leveraging self-attention distillation techniques, the model effectively reduces the number of stacked layers, thereby improving its efficiency in processing long sequences.
- (3) By employing a generative decoder, it is capable of predicting a substantial number of time series in a single forward pass, markedly accelerating inference speed.

The Informer framework is a sophisticated deep learning construct that primarily incorporates stacked self-attention layers. Intrinsically, this model adheres to the encoder-decoder paradigm, as depicted in Figure 3.

For a designated time t , the model's input is characterized as:

$$x^t = \{x_1^t, \dots, x_{L_x}^t | x_i^t \in \mathbb{R}^{d_x}\} \quad (23)$$

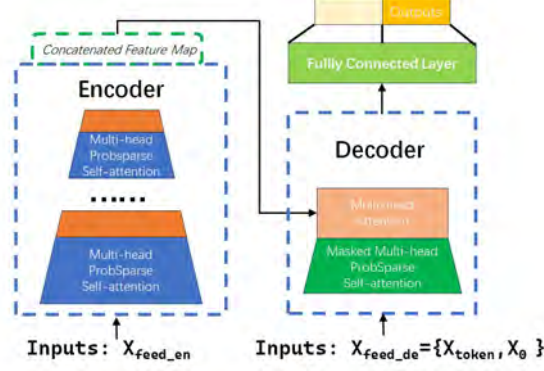


Fig. 3: Informer Model Structure

Simultaneously, the anticipated output of the model is defined as:

$$y^t = \{y_1^t, \dots, y_{L_y}^t | y_i^t \in \mathbb{R}^{d_y}\} \quad (24)$$

The conventional self-attention mechanism operates via the $(Query, Key, Value)$ triplet. This mechanism executes the scaled dot product of these inputs in the following manner:

$$A(Q, K, V) = \text{Softmax} \left(\frac{QK^T}{\sqrt{d}} \right) V \quad (25)$$

Wherein Q, K, V are matrices of dimension $\mathbb{R}^{L \times d}$. The probabilistic representation of the self-attention coefficient, for the i -th Query, is articulated as:

$$A(q_i, K, V) = \mathbb{E}_{p(k_j|q_i)} [v_j] \quad (26)$$

While the inherent time complexity of the self-attention operation stands at $O(L^2)$, empirical studies indicate that its resultant probability distribution assumes a long-tailed configuration. By strategically concentrating on select dot product pairings, computational demands can be optimized with minimal accuracy concessions. To actualize this, the Informer model introduces the ProbSparse self-attention mechanism. The sparsity evaluation for the i -th Query is formulated as:

$$M(q_i, K) = \ln \sum_{j=1}^{L_K} e^{\frac{q_i k_j^T}{\sqrt{d}}} - \frac{1}{L_K} \sum_{j=1}^{L_K} \frac{q_i k_j^T}{\sqrt{d}} \quad (27)$$

The primary term signifies the log-sum-exp operation of the dot products between q_i and all Keys, while the latter term represents their average. Through the implementation of ProbSparse sampling, the self-attention operation's time complexity is efficiently reduced to $O(L \log L)$.

In the encoder phase, the Informer employs Self-attention Distilling to accentuate the dataset’s salient features. This methodology not only prioritizes predominant features but also fosters an improved self-attention feature mapping for ensuing layers. The transition process from layer j to $j + 1$ is represented as:

$$X_{j+1}^t = \text{MaxPool} \left(\text{ELU} \left(\text{Conv1d} \left(X_j^t \right)_{AB} \right) \right) \quad (28)$$

Here, the \llbracket_{AB} notation embodies operations such as multihead ProbSparse computations.

The decoder segment of the Informer adopts a traditional structure, incorporating a pair of consistent multi-head self-attention layers. The input vector for this phase is expressed as:

$$X_{\text{feed_de}}^t = \text{Concat} \left(X_{\text{token}}^t, X_0^t \right) \in \mathbb{R}^{(L_{\text{token}} + L_y) \times d_{\text{model}}} \quad (29)$$

Utilizing Masked Multihead Attention within this layer, the Informer model sidesteps potential autoregressive scenarios. The model’s conclusive predictions are conveyed through a densely connected layer, with the mean square error (MSE) serving as the loss function.

In the realm of long-term time series forecasting, numerous established models exist. Yet, achieving high-precision predictions remains a formidable challenge. To better capture long-range dependencies in time series and enhance forecast accuracy, we introduce an innovative framework: the Stacked Informer Structure.

As illustrated in the Fig 4, we depict the stacked architecture, an enhanced and tiered rendition of the multilayer Informer network. Within the context of the Stacked Informer Structure, consider the original long time series data input, denoted as L . This sequence L undergoes an initial halving (to $L/2$), followed by a subsequent division into quarters (to $L/4$). The input L is then processed through a series of three attention blocks and two convolution layers. Similarly, the $L/2$ sequence traverses two attention blocks and one convolution layer, while the $L/4$ sequence passes through a singular attention block. Ultimately, features extracted from all three $L/8$ dimensional sequences are amalgamated to forge an integrated feature map, which is subsequently channeled to the decoding module. The primary objectives underscoring this stacked architectural configuration are twofold: first, to effectively distill intrinsic temporal features present in the input long time-series sequence; and second, to bolster the robustness and resilience of the forecasted fault sequence output. To achieve these dual aims, the Stacked Informer Structure incorporates a multi-scale input representation, where distinct temporal scales are processed through varied pathways. This ensures the model’s capability to capture diverse features across different temporal scales, thereby enriching the information provision in long-term forecasting tasks. Moreover, by integrating features from varied scales into a cohesive feature map, the model effectively considers interactions and dependencies across scales. For each scale, attention blocks are employed to capture long-range dependencies in the input sequence. The design of these attention blocks is rooted in the Transformer architecture, renowned for its efficacy in capturing long-range dependencies. Additionally, convolution layers are

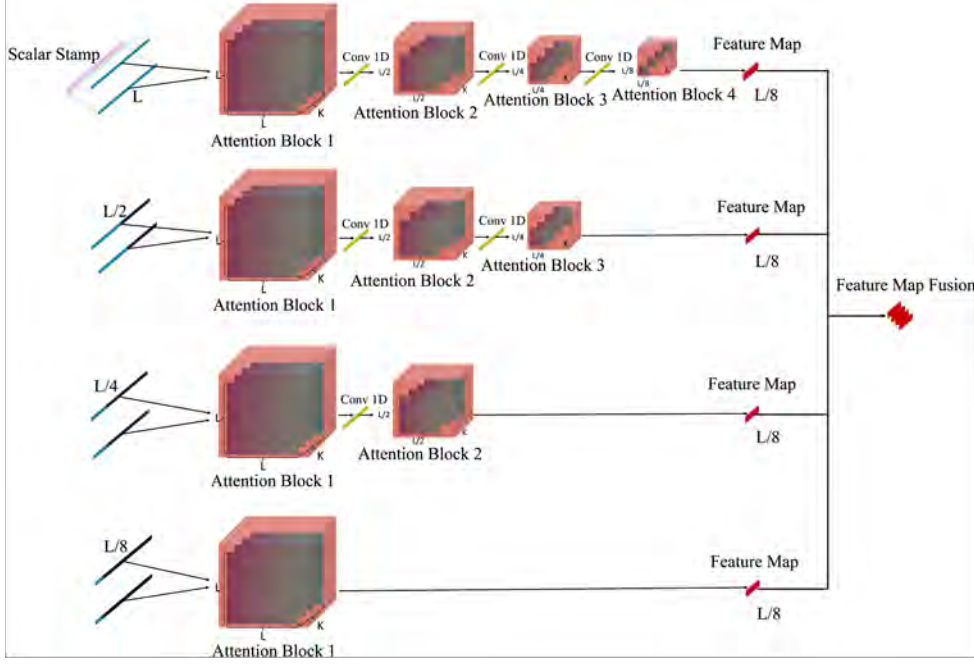


Fig. 4: Proposed Stacked Structure of network

deployed to distill local features, which are subsequently amalgamated with outputs from the attention blocks. In essence, through the adoption of the Stacked Informer Structure, we can adeptly capture intricate patterns and dependencies within long time series, culminating in more precise predictions.

3.2.4 Dynamic Loss Function

The adaptive loss function, a novel approach to loss function formulation, was pioneered in the work of Barron [53]. 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:

$$f(z, \beta, c) = \frac{|\beta - 2|}{\beta} \left(\left(\frac{(z/c)^2}{|\beta - 2|} + 1 \right)^{\beta/2} - 1 \right) \quad (30)$$

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:

$$L(z, \beta, c) = \begin{cases} \frac{1}{2}(z/c)^2 & \text{if } \beta = 2 \\ \log\left(\frac{1}{2}(z/c)^2 + 1\right) & \text{if } \beta = 0 \\ \sqrt{(z/c)^2 + 1} - 1 & \text{if } \beta = 1 \\ 1 - \exp\left(-\frac{1}{2}(z/c)^2\right) & \text{if } \beta = -\infty \\ \frac{|\beta-2|}{\beta} \left(\left(\frac{(z/c)^2}{|\beta-2|} + 1 \right)^{\beta/2} - 1 \right) & \text{otherwise} \end{cases} \quad (31)$$

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.

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.

3.2.5 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 [54]. 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:

$$\phi_{GC}(\nabla w_i L) = \nabla w_i L - \mu \nabla w_i L \quad (32)$$

Here,

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

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:

$$\phi_{GC}(\nabla_W L) = P \nabla_W L, \quad P = I - ee^T \quad (34)$$

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 $\phi_{GC}(\nabla\mathbf{W}L)$ 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.

3.2.6 The Overall Framework of the Prediction Model

Given the intrinsic volatility of financial sequence data, the present research proposes a holistic predictive framework that synergistically integrates five pivotal components: VMD-MIC, Fuzzy Entropy (FE), a stacked Informer structure, a dynamic loss function, and a GC-enhanced Adam optimizer as shown in the Fig5.

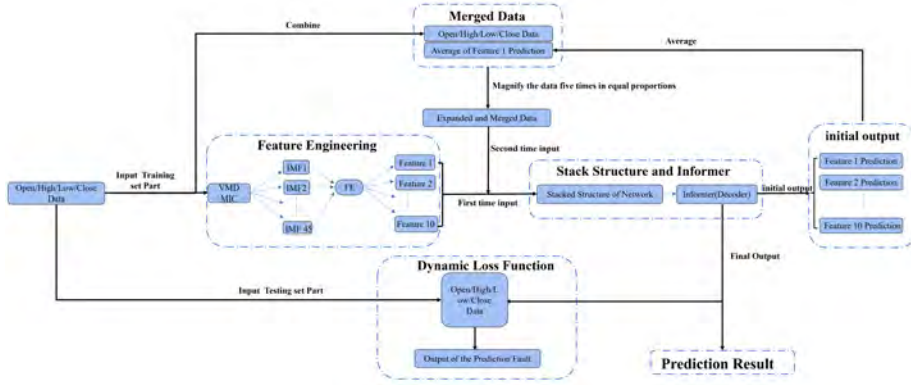


Fig. 5: Overall Framework of the Prediction Model

1. **Data Decomposition with VMD-MIC:** At the inception of our data preprocessing, the original financial data, comprising 29 unique features, is subjected to Variance Mode Decomposition (VMD), resulting in a series of K Intrinsic Mode Functions (IMFs). The VMD-MIC method stands out for its ability to transform time-domain financial sequences into the frequency domain, offering a nuanced perspective on the inherent patterns of the data.
2. **Complexity Measurement via Fuzzy Entropy:** Post decomposition, the Fuzzy Entropy (FE) technique is employed to gauge the complexity of each individual IMF. This quantification not only provides insights into the intricacies of the data structure but also facilitates the amalgamation of IMFs that bear similar characteristics, leading to the generation of reconstructed components.

3. **Feature Selection and Input Preparation:** Leveraging the MIC algorithm, salient features are meticulously selected from the processed data. These curated features serve as the foundation upon which the subsequent predictive modeling is constructed.
4. **Prediction with Stacked Informer:** The heart of our framework is the stacked Informer structure. Built upon advanced architectures like Long Short-Term Memory (LSTM) and Transformers, this model is specifically tailored for long sequence time series forecasting, ensuring impeccable accuracy.
5. **Optimization via GC-Enhanced Adam:**

As we delve into the training phase, the GC-enhanced Adam optimizer comes into play. By incorporating the novel Gradient Centralization technique, this optimizer promises rapid convergence and robust model performance, setting it apart from traditional optimization methods.

To synthesize the outputs from the various components, a weighted ensemble approach is adopted. Specifically, the outputs are combined using a weighted average, where the weights are determined based on each model's performance on a validation set. This ensures that models with superior performance have a more significant influence on the final prediction, providing a refined forecasting curve that captures the intricate nuances of financial time series.

3.3 Evaluation indicators

The mean absolute error (MAE), mean square error (MSE), root mean square error (RMSE), and coefficient of determination (R²) are used as indicators for the prediction performance of our VMGCformer and all the other models. The formulas of each indicators are as follow:

$$MAE = \frac{1}{N} \sum_{t=1}^N |q_{\text{true}}(t) - q_{\text{pred}}(t)| \quad (35)$$

$$MSE = \frac{1}{N} \sum_{t=1}^N (q_{\text{true}}(t) - q_{\text{pred}}(t))^2 \quad (36)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{t=1}^N (q_{\text{true}}(t) - q_{\text{pred}}(t))^2} \quad (37)$$

$$R^2 = 1 - \frac{\sum_{t=1}^N (q_{\text{true}}(t) - q_{\text{pred}}(t))^2}{\sum_{t=1}^N (q_{\text{true}}(t) - \bar{q})^2} \quad (38)$$

4 Feature extraction and experimental results

4.1 Experimental environment and configuration

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 Intel Xeon E5-2650 v3 @2.3GHz CPU and a robust 128GB RAM. The operational environment is governed by the Ubuntu22 operating system, while the deep learning framework employed throughout our endeavors remains PyTorch.

4.2 Feature decomposition and feature engineering

4.2.1 Determination of K Value and decomposition of features

The *VMD-MIC* method enhances the conventional *VMD* approach by utilizing the computation of the *MIC_{yy0}* metric. When the original stock dataset is input into the *VMD-MIC* framework, it is decomposed into K intrinsic mode functions (IMFs). Figure 6 illustrates the relationship between various K values and their corresponding *MIC_{yy0}* results.

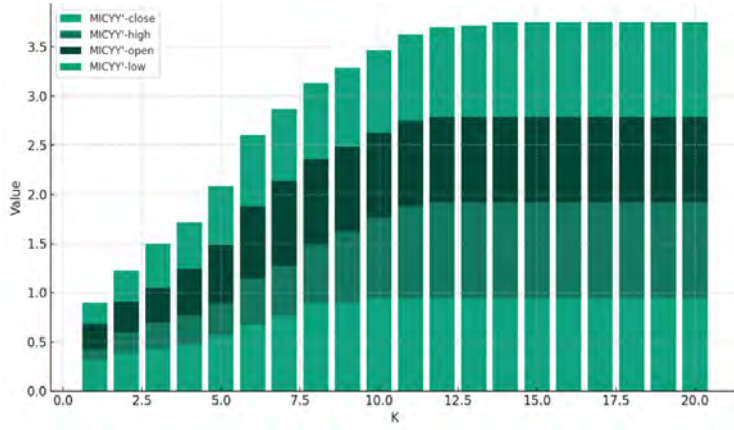


Fig. 6: Relationship Between Different K s and Their Corresponding *MIC_{yy0}*

Notably, the *MIC_{yy0}* 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 Table 3:

Table 3: Selection of K Values

indicator	the selection of K for corresponding indicator
Open	15
Close	10
High	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 Fig7. Correspondingly, we also present their associated Power Spectral Density (PSD) values in Fig8. 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.

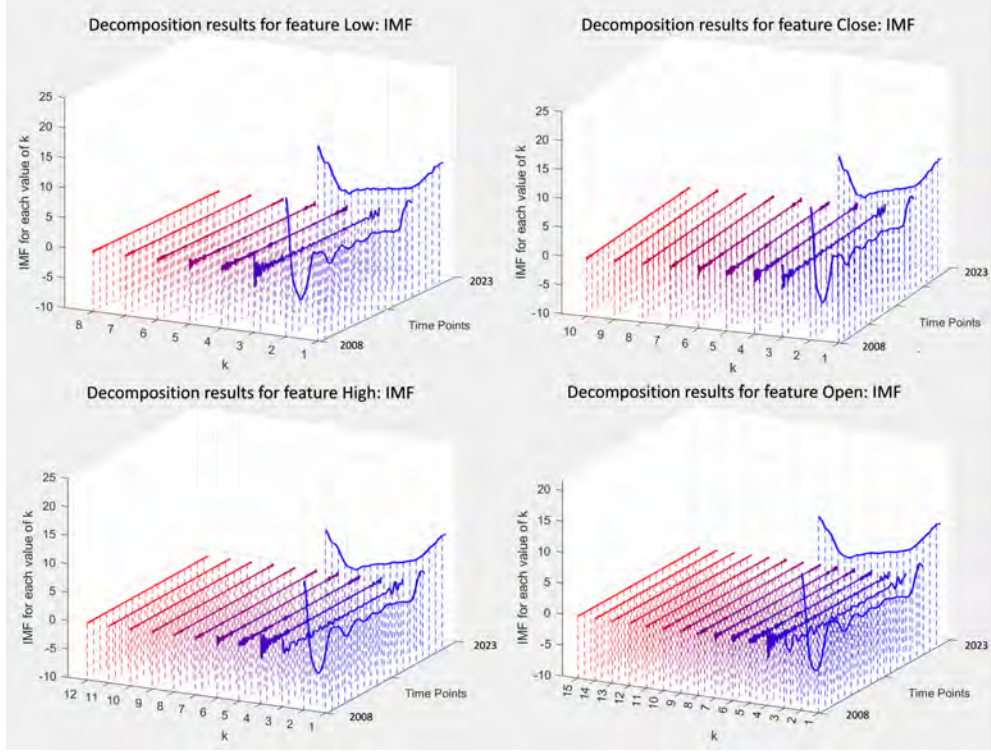


Fig. 7: The Decomposition Results For Each Feature: IMFs Values for Each k .

4.2.2 Dimensionality reduction operation 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 $r = 0.3\text{std}$ were delineated as optimal, striking a balance between model accuracy and computational efficiency. The corresponding visualization in Fig 9 elucidates the FE values for each IMF.

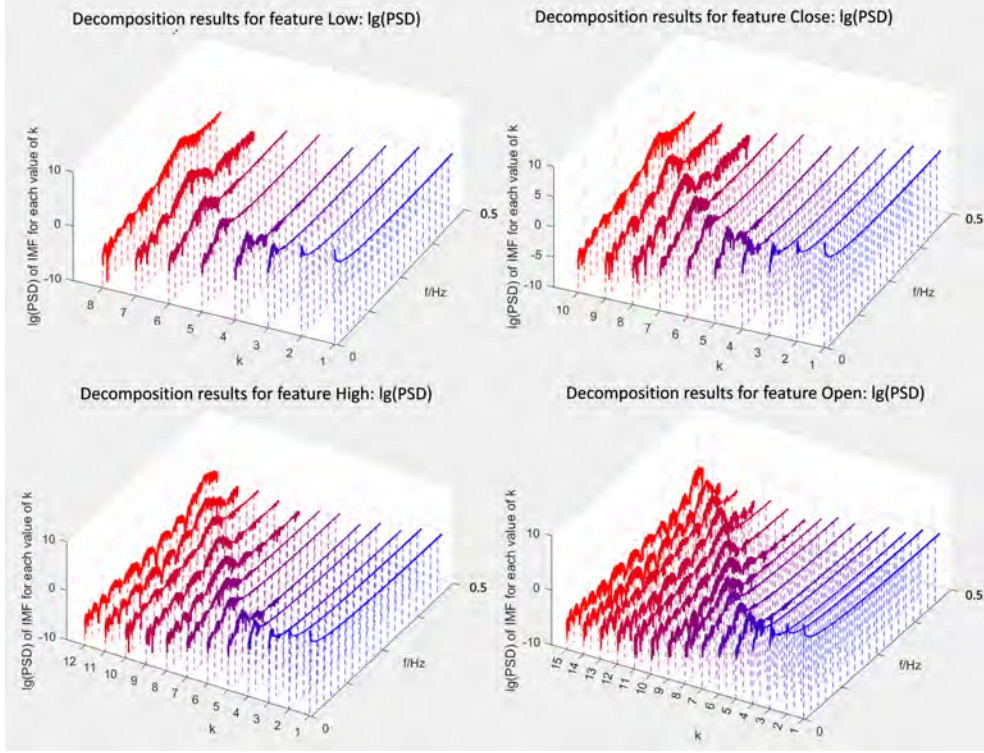


Fig. 8: The Dcomposition Results For Each Feature: Power Spectral Density (PSD) for Each Corresponding IMFs.

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 4.

4.2.3 Feature extraction and selection based on correlation

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 10).

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 0.8 for each component. The outcomes of this selective input feature determination are delineated in the Table 5 below.

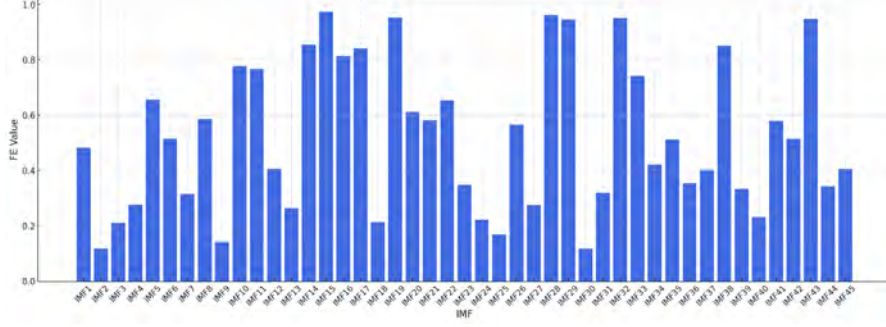


Fig. 9: FE Values for 45 IMFs

Table 4: New Composite Elements Founded FE Metrics

New feature grouping	FE Value range	Selection of IMF _n (value of n)
New feature1	[0.0,0.1]	none
New feature2	(0.1,0.2]	2,9,25,30
New feature3	(0.2,0.3]	3,4,13,18,24,27,40
New feature4	(0.3,0.4]	7,23,31,36,37,39,44,45
New feature5	(0.4,0.5]	1,12,34
New feature6	(0.5,0.6]	6,8,21,26,35,41,42
New feature7	(0.6,0.7]	5,20,22
New feature8	(0.7,0.8]	10,11,33
New feature9	(0.8,0.9]	14,16,17,38
New feature10	(0.9,1.0]	15,19,28,29,32,43

Note2: new feature $N = \frac{\sum_{\text{selected } N} \text{IMFN}}{V}$, V is the total number of selected IMFs.

4.3 Model parameter configuration

In this section, we delineate the rationale behind the selection of our model’s parameters. 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 6.

4.4 Experiments and results

In the experimental segment of this investigation, we integrated ten previously extracted features into our model, procuring predictions for these ten features. Subsequently, we computed their average and amalgamated this mean value with one of the four financial metrics: "open", "high", "low", or "close". The consolidated data was then linearly amplified by a factor of five and re-input into our model to forecast the aforementioned financial indicators. This procedure is graphically delineated in

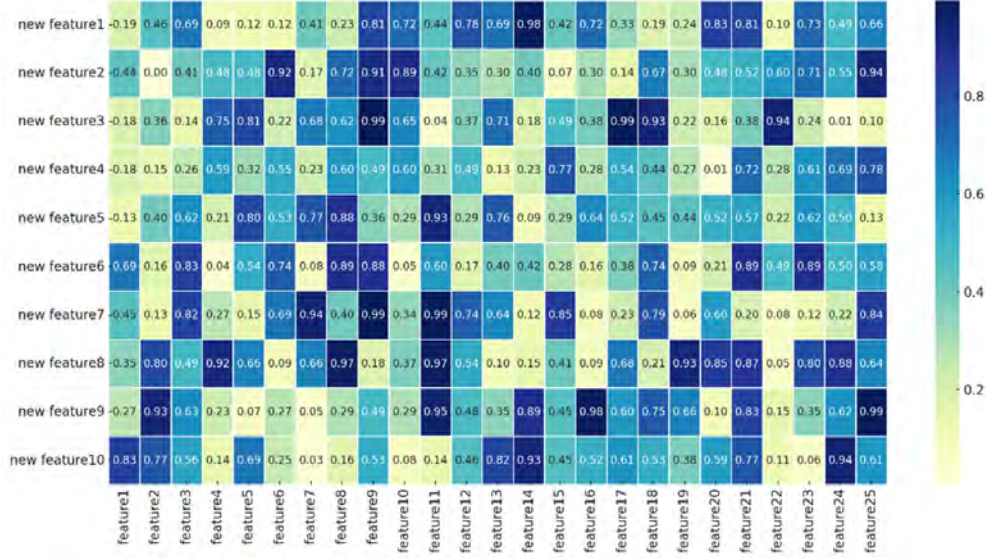


Fig. 10: Heatmap For Features

Table 5: Selection of Reconstructed Feature Values Based On Heatmap

New feature grouping	Included New Feature M (M value)	Included Feature N (N value)
New feature1 reconstruct	1	2,3,7,9,10,11,12,13,14,15,16,20,21,23,24,25
New feature2 reconstruct	2	3,4,5,6,8,9,10,11,14,18,20,31,22,23,24,25
New feature3 reconstruct	3	4,5,7,8,9,10,13,15,17,18,22
New feature4 reconstruct	4	4,6,8,9,10,12,15,17,18,21,23,24,25
New feature5 reconstruct	5	2,3,5,6,7,8,11,13,16,17,18,19,20,21,23,24,25
New feature6 reconstruct	6	1,3,5,6,8,9,11,13,14,18,21,22,23,24,25
New feature7 reconstruct	7	3,6,7,8,9,11,12,13,15,18,20,25
New feature8 reconstruct	8	2,3,4,5,7,8,11,12,15,17,19,20,21,23,24,25
New feature9 reconstruct	9	2,3,9,11,12,14,15,16,17,18,19,21,24,25
New feature10 reconstruct	10	1,2,3,5,9,12,13,14,15,16,17,18,20,21,24,25

Note1: The features corresponding to all areas greater than 0.4 in the heatmap are selected into the corresponding group.

Note2:
$$\text{new feature } N \text{ reconstruct} = \frac{\text{New Feature } M}{V} + \frac{\sum \text{Feature } N}{\text{selected } N} \cdot \frac{N}{V}$$

$$V \text{ is the total number of selected parameters.}$$

the architecture presented in Fig. 11. We then input the results into Laoo Function to calculate the error.

To evaluate the superiority of each module within the model as well as its overall performance, we conducted ablation studies and comparative experiments. Furthermore, to assess the generalizability and robustness of our model, we performed stability tests on several additional stocks. Due to space constraints, we presented only the

Table 6: Model Parameters Configuration

Parameter	Value
Input sequence length	200
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	8
Dimension of model	512
Probsparse attention factor	3.8
Early stopping patience	10
Dropout	0.2
Epochs	150

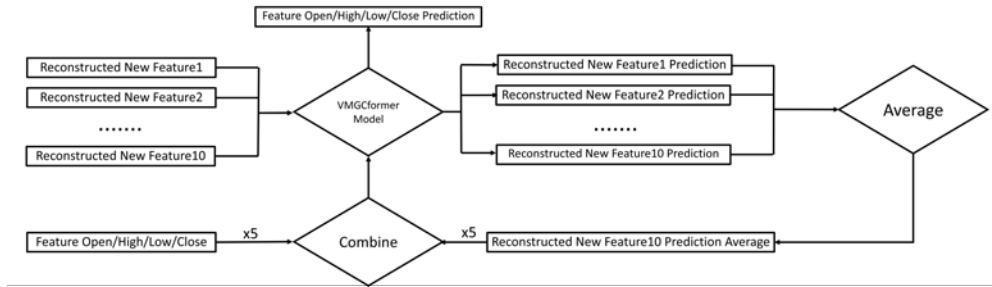


Fig. 11: Structure of The Experiment

prediction results for the "open" indicator in the ablation studies, selected the "high" indicator for the comparative experiments, and showcased the "low" indicator in the stability tests. The indicators chosen typically demonstrated median predictive performance in their respective experiments, a selection approach designed to minimize potential biases arising from random factors.

4.4.1 Ablation Experiment

The objective of the ablation study is to verify whether the intricate hybrid model enhances prediction accuracy in comparison to simpler composite models and standalone models. The benchmark models we selected include Adam+GC+enhanced informer (In the ablation study, for ease of description and comparison, we refer to VMGCformer as Adam+GC+enhanced informer) , Adam+GC+informer, Adam+enhanced informer, and Adam+informer. The corresponding parameters were obtained via grid search methods, where the robustness parameter β is adaptively adjusted by the Adam optimizer. Specific parameters are detailed in the Table 6. Both the encoder and the decoder have input sizes equivalent to the number of input variables of the model.

As depicted in Fig. 12, the training outcomes for the four models are presented. It is evident that the forecasting approach utilizing Adam+GC+informer consistently achieves stability more rapidly and attains the lowest error rates on both the training and test sets, aligning with our initial expectations.

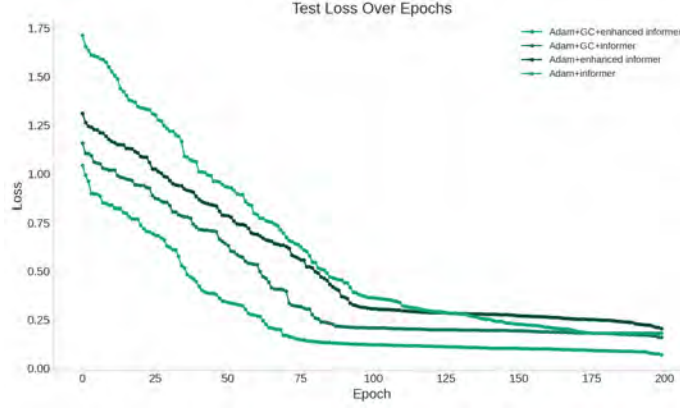


Fig. 12: Training Results For Each Model

Within the prediction results of the ten features, the Adam+GC+enhanced informer demonstrated superior performance in terms of accuracy and robustness against drastic feature value fluctuations, as illustrated in Fig. 13. In the endeavor to predict the four financial indicators: "open", "low", "high", and "close", and to visually represent the outcomes, a comprehensive portrayal of the model's predictive prowess was imperative. To this end, all data points in the results were uniformly divided into 200 segments, chronologically, with each segment comprising 89 data points. A single data point was randomly selected from each segment to ensure a representative sample of the dataset's overall characteristics. Utilizing this methodology, the ablation study results for the "open" indicator, based on four distinct models, are displayed on the left side of Fig. 14. Moreover, for time intervals in the overall results exhibiting significant fluctuations, a similar visualization strategy was employed, partitioning them into 200 segments, as shown on the right side of Fig. 14.

To accentuate the disparities in predictive outcomes across the models, individual model predictions were also distinctly presented. Fig. 15 delineates the predictions from all four models across the entire time frame, while Fig. 16 showcases their respective predictions during periods of pronounced changes.

Based on the visualization of the prediction results and in conjunction with the training times and errors presented in Table 7 for different models, we can draw the following analysis:

- **Advantages of the Adaptive Function:** When predicting the dynamic changes in the financial market, inflection points often pose significant challenges. In this regard, the Adam-GC Informer model exhibited a pronounced accuracy over the

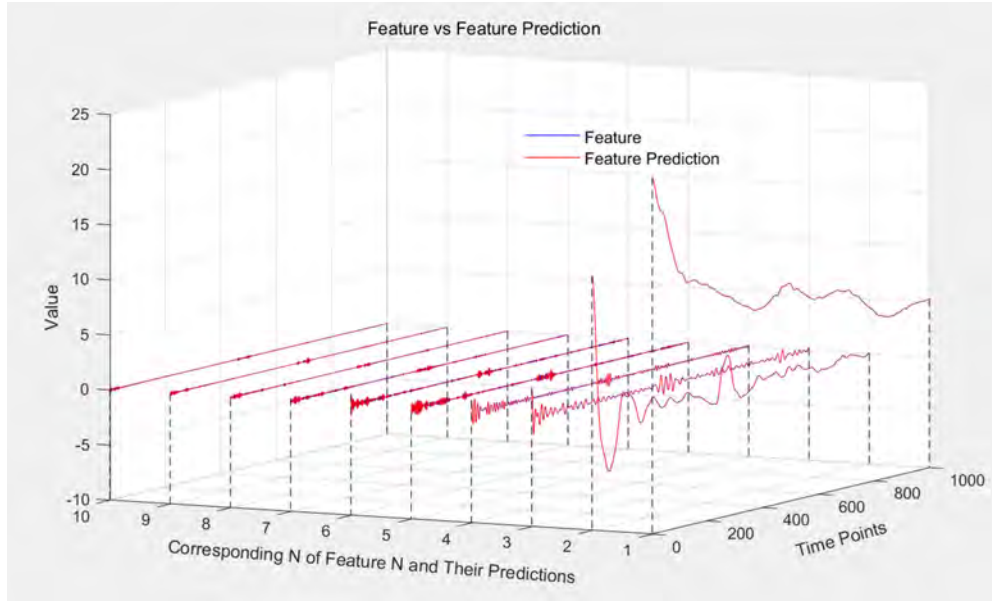


Fig. 13: Prediction Results For Each Feature

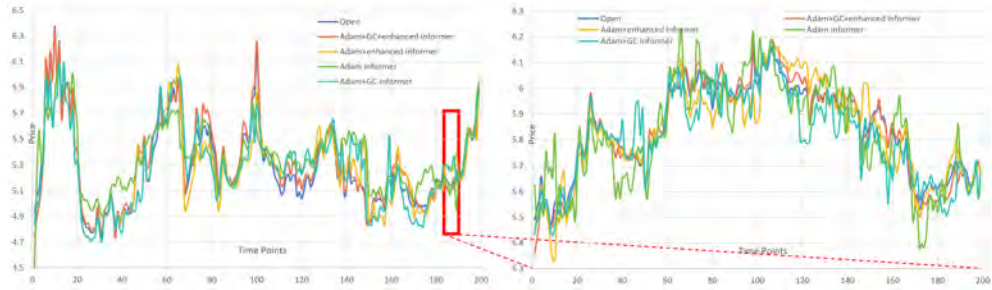


Fig. 14: Visualized Results of Ablation Experiments

Adam-Informer model. This observation underscores the effectiveness of the introduced adaptive function in mitigating errors and uncertainties at abrupt market shifts.

- **GC Data Processing and Temporal Lag:** Data handling and preprocessing stand as pivotal steps in the predictive pipeline. Our evaluations reveal that by employing the Adam+GC+enhanced model, predictions align more closely with actual market data, emphasizing the role of data processing techniques in mitigating temporal lags in predictions.
- **Training Efficiency and Loss Function:** An efficacious loss function is pivotal in ensuring model training efficiency and accuracy. The introduction of an adaptive

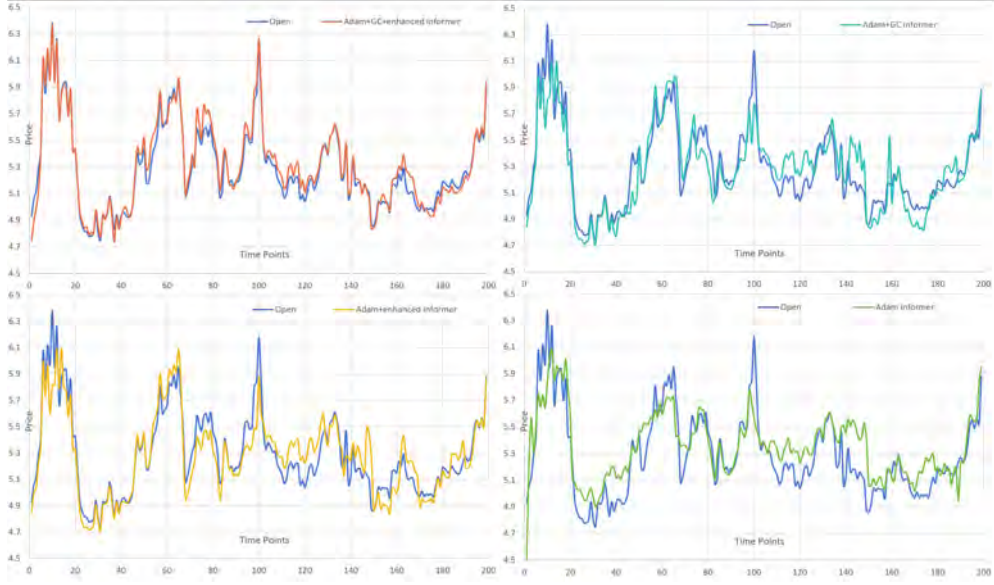


Fig. 15: Visualized Results of Ablation Experiments for Individual Model: Overall Time Frame

loss function facilitated a noticeable reduction in training time for the Adam-GC Informer model, simultaneously bolstering its robustness and predictive fidelity.

- **Performance Evaluation of the Hybrid Model:** The hybrid model introduced in this study exhibited commendable performance across various assessment metrics. In comparison to the Adam-Informer, Adam+enhanced, and Adam+GC informer models, the hybrid model manifested a substantial improvement in MAE, with reductions of 63.15%, 55.86%, and 62.59% respectively. Furthermore, it also demonstrated reductions in RMSE by 57.71% and 30.7% respectively, while showcasing enhancements in R^2 values by 13.75%, 17.60%, and 29.37% respectively.

Upon contrasting various models, it becomes evident that the Adam+GC+enhanced 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.

Table 7: Forecasting Errors of Different Models in Ablation Experiment

Model	MAE	RMSE	R^2	Time(s)
Adam+GC+enhanced informer	0.049	0.063	0.960	1237.39
Adam+GC informer	0.102	0.091	0.828	1026.94
Adam+enhanced informer	0.111	0.114	0.791	1543.84
Adam+informer	0.133	0.149	0.678	1836.41



Fig. 16: Visualized Results of Ablation Experiments for Individual Model: Volatile Time Frame

4.4.2 Comparative Experiment

To rigorously assess the overall superiority of the proposed model, we benchmarked against traditional models including Informer, LSTM, ANN, and CNN, all under parameter settings consistent with VMGCformer. The prediction outcomes are depicted in Fig. 17. In addition, we further compared our model with other hybrid models like EMD-FE-Informer, IVMD-F-Informer, and CNNInformer, with results presented in Fig. 18. And the error values for the predictions of all models can be found in Table 8. The box plot of prediction errors for selected models compared to VMGCformer is depicted in Fig. 19. It is evident that VMGCformer outperforms other models in terms of training error, and it converges more rapidly to a stable level.

Table 8: Forecasting Errors of Different Models in Comparative Experiment

Model	MAE	RMSE	R^2	Time(s)
VMGCformer	0.036	0.053	0.965	1237.39
EMD-FE-Ad-Informer	0.066	0.089	0.902	1004.23
Pyraformer	0.076	0.108	0.857	1124.58
IVMD-FE-Informer	0.102	0.138	0.765	1382.24
Autoformer	0.101	0.148	0.758	420.25
TimesNet	0.115	0.153	0.712	363.14
Informer	0.124	0.162	0.678	143.35
LSTM	0.139	0.183	0.587	524.35



Fig. 17: Single Model Prediction Results

Based on the data presented in the charts and the analyzed characteristics, we can derive the following insights:

1. **Outstanding Performance:** The VMGCformer model consistently excels across all evaluation metrics, surpassing both individual predictive models and other composite models. Specifically, the MAE has been reduced by at least 45.45% compared to other models, the RMSE by at least 40.04%, and the R^2 has seen an approximate increase of 7.09%.
2. **Decomposition Prowess:** Analyzing curve characteristics reveals that, under similar data processing conditions, the IVMD algorithm demonstrates superior data decomposition capabilities compared to the traditional EMD algorithm. This enhances the stock prediction accuracy by effectively attenuating non-smooth features in the original data.
3. **Comparative Superiority:** Under the same data processing methods, VMGCformer significantly outperforms both the Informer and LSTM, with R^2 values of 0.925, 0.889, and 0.808 respectively.
4. **Time Efficiency Considerations:** While the prediction speed of VMGCformer is comparable to other models, the time overhead slightly increases due to the execution of ten Ad-Informer prediction modules post the IVMD-FE data preprocessing.
5. **Long-Term Forecasting Capability:** Further data analysis indicates a trend wherein, during extended forecasting, the curve fit of different models begins to diverge. While most curves maintain a high degree of fit initially, they start deviating from the actual values over time. However, the VMGCformer maintains its



Fig. 18: Mixture Model Prediction Results

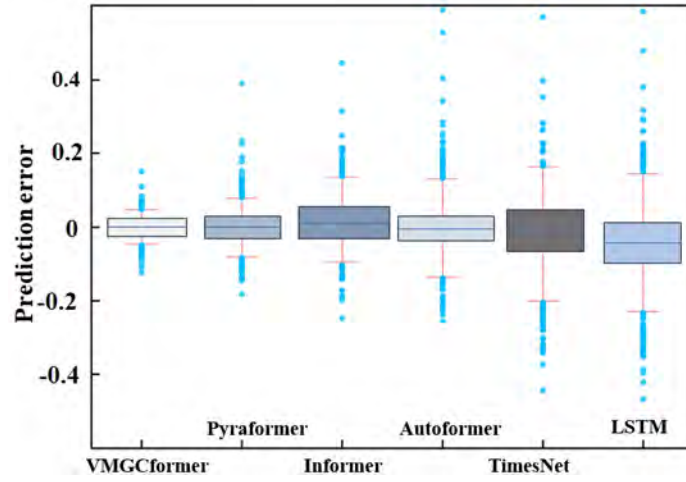


Fig. 19: Box Plot of Prediction Errors For Some Models

prediction accuracy, highlighting its proficiency in addressing the challenges of long-time series forecasting.

In conclusion, VMGCformer represents a superior predictive combination model, holding significant practical implications, especially in domains like long-term investment decisions, risk management, and asset allocation.

4.4.3 Stability Experiment

Previous experimental results indicate that VMGCformer outperforms other benchmark models on the China Petroleum stock dataset, demonstrating considerable capabilities in long-time series forecasting. However, stock prices, as intricate financial time series indices, can exhibit unpredictability due to variations in their industrial sectors, unforeseen events, shifts in market sentiment, and the interplay of chaotic factors. To further scrutinize the model’s stability and applicability, we expanded our study to include four stocks from diverse sectors, each with distinct price trends and variances. These stocks are: ZTE Corporation (sz.000063), China Petroleum & Chemical Corporation (sh.600028), Shanghai Pudong Development Bank (sh.600000), and Dongfeng Motor Corporation (sh.600006).

Under the premise of retaining the original parameter settings, the results of the predictions for the four stocks using VMGCformer are illustrated in Fig. 20. The corresponding error values can be found in Table 9

Table 9: Forecasting Errors for Selected Stocks using VMGCformer

Stock	MAE	RMSE	R^2
sz.000063	1.24991	1.64381	0.961253
sh.600028	0.0399661	0.0509714	0.958723
sh.600000	1.15268	1.48779	0.947264
sh.600006	0.181266	0.231195	0.970125

Based on the analysis of charted results, stocks from four major corporations spanning the chemical, technology, transportation, and financial sectors have demonstrated commendable accuracy under the predictions of the VMGCformer model. However, it is noteworthy that for stocks sz.000063 and sh.600000, the MAE and RMSE values are significantly elevated. This disparity can largely be attributed to the considerably higher stock prices of these two entities compared to the others. Nonetheless, it is reasonable to conclude that the VMGCformer model, as proposed in this study, exhibits high stability and generalization capabilities across diverse datasets.

5 Conclusions & future work

5.1 Conclusions

The operation of stock markets is influenced by various factors, such as economic conditions, seasonal fluctuations, and the global financial environment. These elements can introduce numerous anomalies and non-smooth characteristics into stock market data, posing significant challenges to enhancing the accuracy and performance of stock price predictions. Hence, this paper introduces an adaptive hybrid model for stock predictions, based on an improved VMD, FE, stacked Informer, and adaptive loss function. The Adam+GC+enhanced informer model emerges as a promising hybrid



Fig. 20: Results For Stability Experiment

model capable of adaptively predicting stock data with inherent volatility. The main advantages of this approach are summarized as follows:

- **Hybrid Integration:** The Adam+GC+enhanced informer serves as an integrated model, amalgamating the strengths of various techniques to offer superior accuracy and robustness compared to foundational models like CNN, LSTM, and other hybrid counterparts. This superiority is notably retained in long-time series forecasting.
- **Data Decomposition:** The MIC-enhanced VMD excels in extracting non-linear features from raw data compared to conventional decomposition methods, effectively enhancing data quality and simplifying the prediction process.
- **Adaptive Response:** Comprehensive analysis of experimental results showcases the rapid responsiveness of the adaptive loss function to volatile financial time series data, adeptly handling anomalies and predicting evolving trends.

- **Generalization Capability:** Experiments conducted on additional stock datasets underscore the exemplary predictive outcomes of our proposed model, affirming the significant generalization capabilities and promising prospects of the Adam+GC+enhanced informer in stock data forecasting.

In summary, hybrid stock prediction models, which amalgamate the advantages of multiple algorithms, demonstrate heightened prediction accuracy and robustness. Particularly, their efficacy remains undiminished in long-time series forecasting.

5.2 Future Work

While this study has made preliminary advancements, various aspects warrant further exploration to enrich our understanding and practical application of stock market prediction models. Below are potential research directions:

- **Predictive Modeling for Small Enterprises:** One limitation of our study is the primary focus on large-scale enterprise stocks. However, stock prices of smaller firms may be more sensitive and vulnerable to unforeseen events. Future research could emphasize integrating natural language processing and real-time event analysis to more precisely predict stock fluctuations of smaller firms. Given the non-linear impact of such sudden events on stock prices, it would be judicious to introduce specific parameters or variables to aptly capture these effects.
- **Non-linear Optimization in Feature Engineering:** Another shortcoming in our study pertains to the predominantly linear approach employed in our feature engineering design. Subsequent studies might consider non-linear methods, such as activation functions in neural networks or dynamically learned weight coefficients. This approach can perpetually optimize feature value generation during iterative learning, bolstering model performance and adaptability.
- **Further Optimizations and Refinements:** We've identified two potential avenues for enhancement. Firstly, our feature selection only considered correlation factors, neglecting metrics like F-score and sensitivity. Future endeavors will attempt to utilize F-score and sensitivity to analyze relationships between other variables and stock characteristics, aiming to curtail data redundancy. Secondly, the parameter selection in our study may lack precision and comprehensiveness. To address this, we'll contemplate the integration of optimization algorithms to mitigate sensitivity issues inherent in deep learning network parameter selection.

With a deeper dive into the aforementioned research directions, we anticipate further improvements in the precision, utility, and comprehensiveness of stock market prediction models. This advancement will not only propel academic research in the financial domain but also furnish practitioners with more accurate decision-support tools, aiding in navigating the multifaceted challenges of financial markets.

References

- [1] Ruey S. Tsay, 2005. Analysis of financial time series. John Wiley & Sons.

- [2] Benoit B. Mandelbrot, Richard L. Hudson, 2010. The (mis) behaviour of markets: a fractal view of risk, ruin and reward. Profile books.
- [3] Peter J. Brockwell, Richard A. Davis, 2002. Introduction to time series and forecasting. Springer.
- [4] Norden E Huang, Zheng Shen, Steven R Long, Manli C Wu, Hsing H Shih, Quanan Zheng, Nai-Chyuan Yen, Chi Chao Tung, Henry H Liu, 1998. 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 454(1971), 903–995. The Royal Society.
- [5] Xiaohan Huang, Aihua Jiang, 2022. Wind Power Generation Forecast Based on Multi-Step Informer Network. Energies 15(18), 6642. MDPI.
- [6] Feiyu Li, Zhibo Wan, Thomas Koch, Guokuan Zan, Mengjiao Li, Zhonghai Zheng, Bo Liang, 2023. 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, 108845. Elsevier.
- [7] Mingju Gong, Yin Zhao, Jiawang Sun, Cuitian Han, Guannan Sun, Bo Yan, 2022. Load forecasting of district heating system based on Informer. Energy 253, 124179. Elsevier.
- [8] John Lee, Jow-Ran Chang, Lie-Jane Kao, Cheng-Few Lee, 2023. Financial Analysis, Planning, and Forecasting. In: Essentials of Excel VBA, Python, and R: Volume II: Financial Derivatives, Risk Management and Machine Learning, 433–455. Springer.
- [9] Tomas Jerez, Werner Kristjanpoller, 2020. Effects of the validation set on stock returns forecasting. Expert Systems with Applications 150, 113271. Elsevier.
- [10] Xi Zhang, Siyu Qu, Jieyun Huang, Binxing Fang, Philip Yu, 2018. Stock market prediction via multi-source multiple instance learning. IEEE Access 6, 50720–50728. IEEE.
- [11] Akshit Kurani, Pavan Doshi, Aarya Vakharia, Manan Shah, 2023. A comprehensive comparative study of artificial neural network (ANN) and support vector machines (SVM) on stock forecasting. Annals of Data Science 10(1), 183–208. Springer.
- [12] Omer Berat Sezer, Ahmet Murat Ozbayoglu, "Algorithmic financial trading with deep convolutional neural networks: Time series to image conversion approach", Applied Soft Computing, 70, 525–538, 2018. Elsevier.

- [13] Adam Zaremba, Ender Demir, 2023. ChatGPT: Unlocking the future of NLP in finance. Available at SSRN 4323643.
- [14] Robert F. Engle, 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: Journal of the econometric society* , 987–1007. JSTOR.
- [15] Tim Bollerslev, 1986. Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics* 31(3), 307–327. Elsevier.
- [16] David A. Hsieh, 1988. The statistical properties of daily foreign exchange rates: 1974–1983. *Journal of international economics* 24(1-2), 129–145. Elsevier.
- [17] Robert Goodell Brown, 2004. Smoothing, forecasting and prediction of discrete time series. Courier Corporation.
- [18] Wang, Jujie, Cui, Quan, Sun, Xin, He, Maolin, 2022. Asian stock markets closing index forecast based on secondary decomposition, multi-factor analysis and attention-based LSTM model. *Engineering Applications of Artificial Intelligence* 113, 104908. Elsevier.
- [19] Fang Wang, Menggang Li, Yiduo Mei, and Wenrui Li, 2020. Time series data mining: A case study with big data analytics approach. *IEEE Access* 8, 14322–14328. IEEE.
- [20] Helmut Herwartz, 2017. Stock return prediction under GARCH—An empirical assessment. *International Journal of Forecasting* 33(3), 569–580. Elsevier.
- [21] Jary Pomponi, Simone Scardapane, Aurelio Uncini, 2021. Structured Ensembles: An approach to reduce the memory footprint of ensemble methods. *Neural Networks* 144, 407–418. Elsevier.
- [22] Frederick Tung, Greg Mori, 2018. Clip-q: Deep network compression learning by in-parallel pruning-quantization. *Proceedings of the IEEE conference on computer vision and pattern recognition* 7873–7882.
- [35] Heng-Li Yang, Han-Chou Lin, 2016. An integrated model combined ARIMA, EMD with SVR for stock indices forecasting. *International Journal on Artificial Intelligence Tools* 25, 1650005. World Scientific.
- [24] Gerardo Alfonso, A Daniel Carnerero, Daniel R Ramirez, Teodoro Alamo, 2020. Stock forecasting using local data. *IEEE Access* 9, 9334–9344. IEEE.
- [25] Ayaz Hussain Bukhari, Muhammad Asif Zahoor Raja, Muhammad Sulaiman, Saeed Islam, Muhammad Shoaib, Poom Kumam, 2020. Fractional neuro-sequential ARFIMA-LSTM for financial market forecasting. *IEEE Access* 8, 71326–71338. IEEE.

- [26] Ümit Çavuş Büyüksahin, Şeyda Ertekin, 2019. Improving forecasting accuracy of time series data using a new ARIMA-ANN hybrid method and empirical mode decomposition. *Neurocomputing* 361, 151–163. Elsevier.
- [27] Yue Qiu, Hao-Yu Yang, Shan Lu, Wei Chen, 2020. A novel hybrid model based on recurrent neural networks for stock market timing. *Soft Computing* 24, 15273–15290. Springer.
- [28] Zhigang Jin, Yang Yang, Yuhong Liu, 2020. Stock closing price prediction based on sentiment analysis and LSTM. *Neural Computing and Applications* 32, 9713–9729. Springer.
- [29] Wenjie Lu, Jiazheng Li, Jingyang Wang, Lele Qin, 2021. A CNN-BiLSTM-AM method for stock price prediction. *Neural Computing and Applications* 33, 4741–4753. Springer.
- [30] Yufeng Chen, Jinwang Wu, Zhongrui Wu, 2022. China’s commercial bank stock price prediction using a novel K-means-LSTM hybrid approach. *Expert Systems with Applications* 202, 117370. Elsevier.
- [31] Qiutong Guo, Shun Lei, Qing Ye, Zhiyang Fang, 2021. MRC-LSTM: a hybrid approach of multi-scale residual CNN and LSTM to predict bitcoin price. In *2021 International Joint Conference on Neural Networks (IJCNN)*, 1–8. IEEE.
- [32] H. Zhou, S. Zhang, J. Peng, S. Zhang, J. Li, H. Xiong, W. Zhang, 2021. Informer: Beyond efficient transformer for long sequence time-series forecasting. *Proceedings of the AAAI conference on artificial intelligence*, 35, 12, 11106–11115. AAAI Press.
- [33] Qiuyue Zhang, Chao Qin, Yunfeng Zhang, Fangxun Bao, Caiming Zhang, Peide Liu, 2022. Transformer-based attention network for stock movement prediction. *Expert Systems with Applications* 202, 117239. Elsevier.
- [34] Ha Young Kim, Chang Hyun Won, "Forecasting the volatility of stock price index: A hybrid model integrating LSTM with multiple GARCH-type models", *Expert Systems with Applications*, 103, 25–37, 2018. Elsevier.
- [35] Heng-Li Yang, Han-Chou Lin, "An integrated model combined ARIMA, EMD with SVR for stock indices forecasting", *International Journal on Artificial Intelligence Tools*, 25(02), 1650005, 2016. World Scientific.
- [36] Zhidan Luo, Wei Guo, Qingfu Liu, Zhengjun Zhang, "A hybrid model for financial time-series forecasting based on mixed methodologies", *Expert Systems*, 38(2), e12633, 2021. Wiley Online Library.
- [37] Pei Du, Jianzhou Wang, Wendong Yang, Tong Niu, "Container throughput forecasting using a novel hybrid learning method with error correction strategy",

Knowledge-Based Systems, 182, 104853, 2019. Elsevier.

- [38] Hasnain Iftikhar, Aimel Zafar, Josue E Turpo-Chaparro, Paulo Canas Rodrigues, Javier Linkolk López-Gonzales, "Forecasting Day-Ahead Brent Crude Oil Prices Using Hybrid Combinations of Time Series Models", Mathematics, 11(16), 3548, 2023. MDPI.
- [39] Liang-Ying Wei, 2016. A hybrid ANFIS model based on empirical mode decomposition for stock time series forecasting. Applied Soft Computing, 42, 368–376. Elsevier.
- [40] Lihki Rubio, Keyla Alba, 2022. Forecasting selected colombian shares using a hybrid ARIMA-SVR model. Mathematics, 10(13), 2181. MDPI.
- [41] Shuqi Li, Xiaoming Liu, Aijing Lin, 2020. Fractional frequency hybrid model based on EEMD for financial time series forecasting. Communications in Nonlinear Science and Numerical Simulation, 89, 105281. Elsevier.
- [42] Sarbjit Singh, Kulwinder Singh Parmar, Jatinder Kumar, 2021. Soft computing model coupled with statistical models to estimate future of stock market. Neural Computing and Applications, 33, 7629–7647. Springer.
- [43] Yakup Kara, Melek Acar Boyacioglu, Ömer Kaan Baykan, 2011. 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 38(5), 5311–5319. Elsevier.
- [44] Erkam Guresen, Gulgun Kayakutlu, Tugrul U Daim, 2011. Using artificial neural network models in stock market index prediction. Expert systems with Applications 38(8), 10389–10397. Elsevier.
- [45] S.B. Achelis, 2005. Technical Analysis From A To Z. Vision Books. <https://books.google.com/books?id=J1YDPgAACAAJ>
- [46] K. Dragomiretskiy, D. Zosso, 2013. Variational mode decomposition. IEEE transactions on signal processing, 62(3), 531–544. IEEE.
- [47] G. Lin, A. Lin, D. Gu, 2022. Using support vector regression and K-nearest neighbors for short-term traffic flow prediction based on maximal information coefficient. Information Sciences, 608, 517–531. Elsevier.
- [48] W. Chen, J. Zhuang, W. Yu, Z. Wang, 2009. Measuring complexity using fuzzyen, apen, and sampen. Medical Engineering & Physics, 31, 1, 61–68. Elsevier.
- [49] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, L. Kaiser, I. Polosukhin, 2017. Attention is all you need. Advances in neural

information processing systems, 30. NIPS.

- [50] S. Selvin, R. Vinayakumar, E.A. Gopalakrishnan, V.K. Menon, K.P. Soman, 2017. Stock price prediction using LSTM, RNN and CNN-sliding window model. 2017 international conference on advances in computing, communications and informatics (icacci), 1643–1647. IEEE.
- [51] Q. Ye, X. Yang, C. Chen, J. Wang, 2019. River water quality parameters prediction method based on LSTM-RNN model. 2019 Chinese Control And Decision Conference (CCDC), 3024–3028. IEEE.
- [52] R. Pascanu, T. Mikolov, Y. Bengio, 2013. On the difficulty of training recurrent neural networks. International conference on machine learning, 1310–1318. Pmlr.
- [53] J. T. Barron, 2019. A general and adaptive robust loss function. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 4331–4339. IEEE/CVF.
- [54] Hongwei Yong, Jianqiang Huang, Xiansheng Hua, Lei Zhang, 2020. Gradient centralization: A new optimization technique for deep neural networks. Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part I 16 , 635–652. Springer.