

**Yuheng ZHANG, MEng**

zhangyuheng@pku.edu.cn

Private Banking Department, China Construction Bank, Beijing, China

**Xin'e LI, PhD (corresponding author)**

lixine1999@sina.com

Management College, Beijing Union University, Beijing, China

## **Design and Application of a Machine Learning Model for Quantitative Stock Selection Incorporating Association Analysis and Hierarchical Attention Mechanisms**

**Abstract.** *The application of machine learning models to assist investors in quantitative stock selection is an emerging topic of interest. However, due to the complex and dynamic nature of stock markets and issues such as market inefficiency, single machine learning models may underperform, and existing ensemble methods risk losing useful information. To address these challenges, this paper designs and optimises a machine learning model tailored to the characteristics of the Chinese stock market and quantitative stock selection. The model incorporates association analysis into the feature engineering process, generating differentiated datasets through information addition. Furthermore, a hierarchical attention mechanism is introduced in the model integration phase, dynamically assigning weights to base models via "upward integration" and "downward allocation." The model's effectiveness is validated through comparative experiments, ablation studies, and hyperparameter analysis on a self-constructed dataset of over 3.34 million records. Backtesting further demonstrates its practical performance in the stock market, yielding promising results, as expected.*

**Keywords:** *association analysis, hierarchical attention mechanism, feature engineering, ensemble learning model, quantitative stock selection.*

**JEL Classification:** C51, C55, C61, G11.

<b>Received: 1 April 2026</b>	<b>Revised: 2 June 2026</b>	<b>Accepted: 5 June 2026</b>
-------------------------------	-----------------------------	------------------------------

### **1. Introduction**

China's stock market is one of the largest markets globally. With the rapid development of China's economy, the stock market has offered a wide range of investment opportunities for investors. However, due to factors such as policy fluctuations and other external influences, the efficiency of China's stock market remains insufficient. The stock selection process struggles to fully capture historical price trends, volatility information, and future fundamental growth prospects. Additionally, high market volatility undermines the stability, effectiveness, and generalisation capability of conventional single machine learning models. Therefore, designing and optimising machine learning models tailored to the unique

---

DOI: 10.24818/18423264/60.2.26.12

© 2026 The Authors. Published by Editura ASE. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

characteristics of China's stock market to assist investors in making quantitative stock selection decisions has become a shared focus in both academia and industry.

The innovations of this paper lie in the following aspects: First, to address the issues of information loss caused by feature selection and the insufficiency of traditional factors in existing ensemble learning models for quantitative stock selection, a correlation analysis-based feature engineering method is proposed. It generates differentiated feature pools via matrix mapping, enhancing diversity among base models through an informational "addition" strategy. Second, a hierarchical attention mechanism is designed to dynamically assign weights to base models based on market volatility and sample characteristics, overcoming the limitation of traditional models that rely on fixed weights and struggle to adapt to complex market conditions. Third, on a self-constructed dataset of over 3.34 million records from historical CSI 300 constituents, evaluation metrics well-aligned with practical scenarios, such as RankIC, are introduced. The rationality and scientific validity of the model are rigorously examined on this large-scale dataset through comparative experiments, ablation studies, hyperparameter analysis, and backtesting.

The organisation of this study after this introductory section is as follows: Section 2 presents a literature review, covering relevant research on machine-learning-based quantitative stock selection models, association analysis, and hierarchical attention mechanisms. Section 3 focuses on model design and dataset construction. Section 4 discusses the application of the proposed model. Finally, Section 5 provides conclusions on the study, as well as comments on the limits of our research and some future research directions.

## **2. Literature Review**

### ***2.1 Quantitative Stock Selection Models***

In the evolution of quantitative stock selection models, pioneers like Markowitz, William Sharpe, Fama, and French designed and empirically tested factors tailored to the US stock market. Some scholars argue that the Chinese A-share market also provides fertile ground for factor research and quantitative stock selection (Zhou, 2021). However, research on factors specific to the Chinese stock market remains limited. Recent years have seen a growing trend in applying machine learning to quantitative investment (Htun et al., 2023). For instance, some researchers have integrated sentiment analysis models based on LSTM-CNN to incorporate news and public opinion (Ge et al., 2024; Liu et al., 2024), while others have used machine learning to aggregate investor sentiment for stock recommendation systems (Qin et al., 2024). Although these approaches improve stock selection models to some extent, they often suffer from small sample sizes and are prone to failure due to market inefficiencies.

### ***2.2 Ensemble learning model***

Ensemble learning models have been widely applied in various fields, including medicine (Adachi et al., 2023) and finance (Ge et al., 2024; Song et al., 2024).

Enhancing diversity appropriately can significantly improve the overall performance of ensemble models (Cunningham & Carney, 2000). The primary goal of feature engineering methods is to increase the diversity of input features for each base model, thereby enhancing their diversity (Zhang et al., 2023). Regarding model integration, one common approach is to combine the results of all base classifiers with equal weights. Although simple, this method overlooks the varying capabilities of different base models. Some researchers have assigned fixed but different weights to base models (Zhang et al., 2023). However, under different market conditions, there is no stable hierarchy of performance among classifiers (Li & Tu, 2024), making it difficult to address the fluctuating stock market using fixed integration weights.

### ***2.3 Association Analysis***

Some scholars have used machine learning algorithms to capture and measure stock correlations. For instance, graph neural networks based on multi-head attention mechanisms (Li et al., 2022) and GNNs with graph attention (Feng et al., 2022) have been developed.

Research has also identified the "lead-lag effect" in stock markets (Chan & Hameed, 2006), where fluctuations triggered by new information diffuse among companies with certain "associations." For example, geographically associated companies exhibit a "lead-lag effect" in stock price movements relative to target stocks (Parsons et al., 2020), as similar return patterns often reflect analogous underlying conditions (Tsinaislanidis et al., 2014; Lee & Lee, 2024). This effect has similarly been observed in China's A-share market (Duan et al., 2022).

### ***2.4 Attention and Hierarchical Attention Mechanisms***

The attention mechanism is a computational model that simulates how humans allocate attention, focusing selectively on specific components when processing large amounts of information (Niu et al., 2021). Several scholars have applied the attention mechanism to image detection (Zhang, 2024) and the actor-critic framework (Liu et al., 2024), attention learning to intermittent demand forecasting (Praveena & Prasanna Devi, 2024), attention-based visualisation (Gradient-weighted Class Activation Mapping, Grad-CAM) to dementia classification (Ko et al., 2024), and spatial-temporal attention to supply chain financial fraud detection (Xie et al., 2024). In recent years, attention mechanisms have been increasingly applied across various domains (Chai, 2024), yet their integration into ensemble learning models for quantitative stock selection remains underexplored.

### ***2.5 Summary***

In summary, the main shortcomings of existing machine learning methods in addressing quantitative stock selection problems are as follows: Firstly, traditional feature engineering approaches often result in the loss of useful information; Secondly, stock selection factors fail to fully capture historical price trends,

volatility, and future fundamental growth prospects; Thirdly, current ensemble methods struggle to adapt to the dynamic nature of stock markets; Fourthly, sample sizes are often insufficient.

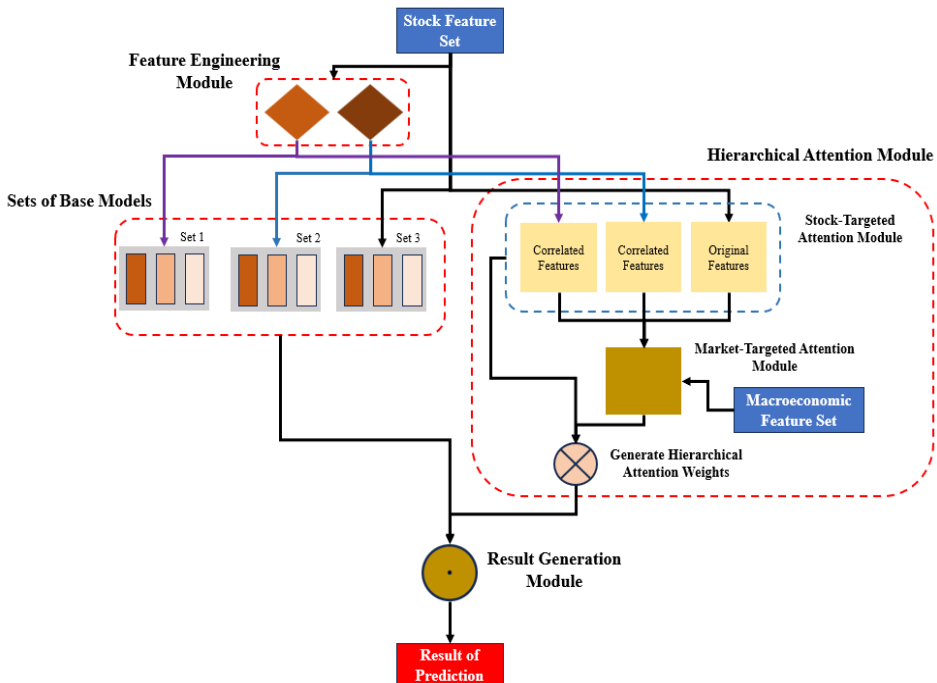
To address these issues, this paper draws on the "lead-lag" effect and association analysis, proposing a novel feature engineering method that incorporates association analysis. Additionally, we design an ensemble method based on a hierarchical attention mechanism to improve model performance and adaptability.

### 3. Model Design and Dataset Construction

#### 3.1 Model design

##### 3.1.1 Incorporating Association Analysis in Feature Engineering

As shown in Figure 1, the machine learning model framework is divided into four main components. The first is the feature engineering module, which incorporates association analysis to process stock selection features comprising fundamental and price-volume characteristics. And within this module, small-order fund synergy and return correlation matrices are designed based on the realities of the Chinese A-share market, and then are used in the feature engineering module.



**Figure 1. Model Framework**  
 Source: Illustration by authors.

Transactions below ¥40,000 are classified as small-order trades, those between ¥40,000 and ¥200,000 as medium-order trades, those between ¥200,000 and ¥1,000,000 as large-order trades, and those above ¥1,000,000 as ultra-large-order trades. The synergy of small-order funds is used to define the association relationship, which is quantified as the proportion of days within the past month where the net flow direction of small-order funds for two stocks was the same. A counting method is employed to construct the association matrix, with the matrix at each time point defined as follows:

$$A = \begin{pmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & \ddots & \vdots \\ a_{n,1} & \cdots & a_{n,n} \end{pmatrix} \quad (1)$$

Each  $a_{i,j}$  in the matrix is defined as:

$$a_{i,j} = \frac{n_{i,j}}{m} \quad (2)$$

Where  $n_{i,j}$  represents the number of days in the past month when the net flow direction of small-order funds for the two stocks was the same, and  $m$  is the total number of trading days in the previous month.

The correlation of daily returns between stocks, which is quantified as the Pearson correlation coefficient of the daily return vectors of the two stocks over the past month, is used to define another association relationship as well. The coefficient is calculated as the association relationship value. A coefficient-based method is employed to construct the association matrix, with the matrix at each time point defined as follows:

$$A = \begin{pmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & \ddots & \vdots \\ a_{n,1} & \cdots & a_{n,n} \end{pmatrix} \quad (3)$$

Each  $a_{i,j}$  in the matrix is defined as:

$$a_{i,j} = \frac{cov(R_i, R_j)}{std(R_i) * std(R_j)} \quad (4)$$

Where  $R_i$  and  $R_j$  are the daily return sequences of the two stocks over the past month.

Based on the small-order fund synergy matrix and the return correlation matrix, linear matrix mapping operations are performed on the original features to generate two association feature pools. The calculation method is shown below.

$$\begin{aligned} Factors &= A_t @ V_t * Adj_t \\ &= \begin{pmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & \ddots & \vdots \\ a_{n,1} & \cdots & a_{n,n} \end{pmatrix} @ \begin{pmatrix} b_1^t \\ \vdots \\ b_n^t \end{pmatrix} * \begin{pmatrix} 1 \\ \frac{\sum_{i=1}^n a_{1,i}}{\sum_{i=1}^n a_{n,i}} \\ \vdots \\ 1 \end{pmatrix} \end{aligned} \quad (5)$$

Where the symbol @ denotes matrix multiplication and \* represents multiplication of elements one by one in a matrix or vector. The term "Factors" on the left-hand side of the equation is the correlation factor values of stocks in the underlying pool obtained after feature engineering, formatted as an  $n \times 1$  vector, where  $n$  is the size of the stock pool. The vector  $V_t$  represents the specific metric value  $b_t^t$  of each stock at the current time point. The vector  $Adj_t$  serves to adjust the dimensional scale of the associated features.

The factors pool which has been generated, along with the original feature pool, result in three distinct sets of features.

Thus, linear matrix mapping is employed to augment the original features with incremental information, generating a new feature pool. In contrast to feature selection methods that derive differentiated datasets by "subtracting" distinct information (Kim & Moon, 2024), the proposed approach introduces additional information to enhance dataset diversity, which is a kind of information "addition".

### 3.1.2 Base Model Selection

In ensemble learning, selecting base models with differing principles and assumptions can enhance the diversity of base model outcomes (Dikici & Bilgin, 2025). Therefore, Naive Bayes, Random Forest, and XGBoost are chosen as the base model group. Three distinct model groups are developed, each corresponding to one of the three feature sets. The complete base model pool is formed by combining all three model groups.

(1) Naive Bayes Model. The naive Bayes model is based on Bayes' theorem and the assumption of conditional independence among features. However, several scholars have noted its "mysterious optimality": even when features are significantly dependent, the classifier can still achieve optimal decisions (Thiruvenkatasamy et al., 2024). This is because the ranking of posterior probabilities often remains correct even when the independence assumption is severely violated. The algorithm is expressed as follows:

$$P(C_k|x_1, x_2, \dots, x_n) = \frac{P(C_k) \prod_{i=1}^k P(x_i|C_k)}{\sum_j P(C_j) \prod_{i=1}^n P(x_i|C_j)} \quad (6)$$

Here,  $x_1, x_2, \dots, x_n$  represent the feature sequence of a given sample. The left side of the equation denotes the probability of the sample being classified as category  $C_k$ , where  $P(C_k)$  is the prior probability of the category, and  $P(x_i|C_k)$  is the probability of feature  $x_i$  occurring given that the sample belongs to  $C_k$ .

(2) Random Forest Algorithm. Random Forest does not assume independence among data or features. Instead, it uses tree structures to capture feature interactions. The algorithm assumes each decision tree is independent, and its predictions are aggregated through voting or averaging.

(3) XGBoost Algorithm. The XGBoost model's objective function consists of two parts: the loss from gradient boosting and a regularisation term. It can be expressed as:

$$L(\varphi) = \sum_{i=1}^n l(y'_i, y_i) + \sum_k \Omega(f_k) \quad (7)$$

The regularisation function  $\Omega$  penalises based on either the absolute or the squared sum of model parameters to prevent overfitting.

### 3.1.3 Attention Mechanism for Sample Features

Different classifiers may exhibit varying predictive capabilities for different samples within the same problem. For some samples, classifier  $A$  may perform better, while for others, classifier  $B$  may be superior (Choi & Lim, 2021; Mesa et al., 2025).

Suggest that  $n$  stocks should be considered in the current period. Assuming that each stock's associated feature vector has a dimension of  $m$ , any association feature pool (so does the original feature pool) can be represented as a matrix  $A$ ,  $A \in \mathbb{R}^{N \times M}$ , where each row represents the feature values of a stock for the current period.

In the model, each association feature pool is used to train three base models, represented by three one-hot vectors. The vector for the  $i$ -th classifier is marked as 1 at the  $i$ -th position and 0 elsewhere. The matrix formed by these three vectors is an identity matrix, denoted as  $C$ ,  $C \in \mathbb{R}^{3 \times 3}$ , where each row represents a classifier in the classifier group. In the attention mechanism module for sample features, the current feature matrix  $A$  acts as the query, and the classifier encoding vectors act as the keys. Before computing their attention scores, both need to be embedded into the same-dimensional space.

Define embedding matrices  $W$  and  $H$ , where  $W \in \mathbb{R}^{M \times D}$ ,  $H \in \mathbb{R}^{3 \times D}$ , to generate latent space representations of the feature matrix and classifier matrix, respectively, with  $D$  representing the dimension of the latent space. Through linear matrix operations, the query and keys are mapped to a  $D$ -dimensional linear space, and their inner products are computed to derive the attention scores. This can be expressed in matrix form as follows:

$$\begin{aligned} Hidden_A &= A@W \\ Hidden_C &= C@H \\ Attention_{feature} &= Softmax(Hidden_A@Hidden_C^T) \end{aligned} \quad (8)$$

The  $Attention_{feature}$  matrix has dimensions  $N * 3$ , where the element in the  $i$ -th row and  $j$ -th column represents the attention score assigned to the  $j$ -th classifier for the  $i$ -th stock. Here,  $@$  denotes the matrix multiplication operation.

### 3.1.4 Attention Mechanism Module for Market Fluctuations

The latent space representation of each stock's associated feature is equally weighted and summed to represent the association relationship. Let  $R$  be the matrix composed of the latent space representations of the three association relationships,  $R \in \mathbb{R}^{3 \times D}$ , The current market fluctuation vector,  $Market$ , acts as the query, while the association relationship representation vectors act as the keys. Define embedding matrices  $W$  and  $H_0$ , where  $W_0 \in \mathbb{R}^{L \times D}$ ,  $H_0 \in \mathbb{R}^{D \times D}$ , to generate latent space representations of the feature matrix and classifier matrix, respectively, where  $D$  is the dimension of the latent space and  $L$  is the length of the  $Market$  vector. Through linear matrix operations, the query and keys are mapped to a  $D$ -dimensional linear space, and their inner products are computed to derive the attention scores. This can be expressed in matrix form as follows:

$$\begin{aligned} Hidden_{Market} &= Market @ W_0 \\ Hidden_R &= R @ H_0 \\ Attention_{Market} &= Softmax(Hidden_{Market} @ Hidden_R^T) \end{aligned} \quad (9)$$

The resulting  $Attention_{Market}$  matrix has dimensions  $1 \times 3$ , where the  $i$ -th element represents the attention score for the  $i$ -th association feature.

### 3.1.5 Loss Function Design

The loss function is composed of two components: (1) Cross-entropy loss to measure prediction accuracy. (2) Cosine similarity to constrain similarity among base classifiers, enhancing ensemble model performance. The three datasets (original features, small-order fund synergy features, and return correlation features) are denoted as  $ori$ ,  $sma$ , and  $ret$ .

The three attention matrices for sample features, denoted as  $Attention_{ori}$ ,  $Attention_{sma}$ , and  $Attention_{ret}$ , are concatenated into  $Attention_{feature}$ . Then the market attention vector  $Attention_{Masked}$  is expanded to  $Attention_{MaskedMarket}$  via repetition. So, the integrated weights for each stock and each base model can be expressed as follows:

$$Weight = Attention_{MaskedMarket} * Attention_{feature} \quad (10)$$

Here, the symbol  $*$  denotes element-wise matrix multiplication, and the matrix weight is used to compute the weighted sum of the predicted probabilities from each classifier for each stock, resulting in an  $N \times 1$  vector denoted as  $Predict$ .

The cross-entropy loss for the prediction of each stock is calculated as:

$$BCE_i = -y_i * \log[\text{sigmoid}(Predict_i)] + (1 - y_i) * \log[1 - \text{sigmoid}(Predict_i)] \quad (11)$$

Here,  $y_i$  represents the true label of the  $i$ -th stock, and  $Predict_i$  represents its predicted probability. The final cross-entropy loss is computed as:

$$BCELoss = \sum_{i=1}^N BCE_i \quad (12)$$

For base models  $p$  and  $q$ , their latent space vectors are denoted as  $\mathbf{h}_p$  and  $\mathbf{h}_q$ . The cosine similarity is calculated as:

$$S(p, q) = \frac{\mathbf{h}_p @ \mathbf{h}_q^T}{|\mathbf{h}_p|^2 * |\mathbf{h}_q|^2} \quad (13)$$

The second part of the loss function is computed as:

$$Diversity = \sum_{1 \leq p < q \leq 9} [1 - S(p, q)] \quad (14)$$

The final loss function is formulated as:

$$Loss = BCELoss - \lambda * Diversity \quad (15)$$

Here,  $\lambda$  adjusts the impact of the penalty for insufficient diversity among base models on the overall loss function.

### 3.2 Dataset Construction

#### 3.2.1 Fundamental Indicators

Fundamental indicators, price-volume indicators, and macroeconomic indicators were selected as experimental metrics. Among these, fundamental indicators are primarily utilised to assess the health status and future development potential of a company, encompassing financial data and operational management metrics.

The Fundamental indicators encompass the historical year-on-year and quarter-on-quarter growth rates of four key indicators: return on equity (ROE), earnings per share (EPS), price-to-earnings ratio (P/E), and net profit (NP). Additionally, the year-on-year and quarter-on-quarter growth rates of analysts' consensus forecast indicators (EPS and NP) are included, resulting in a total of 12 characteristics.

#### 3.2.2 Price-Volume Indicators

For price-volume indicators, a set of 33 representative and effective indicators, including indicators such as the Aroon index, relative strength index, Sentiment intention index, Cumulative Vibration Lifting Index, and so on, is selected to construct the initial indicator pool.

#### 3.2.3 Macroeconomic Indicators

Macroeconomic conditions are the same for all stocks in the entire market, such as gross domestic product, gross national product, inflation and deflation, investment, consumption, etc., which can be used to effectively measure the overall market.

Macroeconomic indicators serve as input data for the attention mechanism module aimed at capturing market fluctuations. A total of 45 macroeconomic indicators are selected, 10 for measuring the macroeconomic environment (e.g., the

year-on-year growth rate of the Kestrength Index, the difference between the year-on-year growth rates of M1 and M2, etc.) and 35 for assessing the market trend of broad indices (e.g., CSI 50, Shanghai and Shenzhen industry indices, etc.).

The dataset for model training and prediction consisted of over 1.62 million stock selection feature records, while over 1.72 million records were used to construct the association analysis matrix for the feature engineering module, totalling more than 3.34 million records. Most of the stock selection feature data were sourced from publicly available databases such as CSMAR, Uqer, and Choice, with a small portion derived from internal fund company databases. The year-on-year and month-on-month growth rates for some fundamental factors were calculated from raw data. The dataset contents and descriptions are summarised in Table 1. Before experimentation, the initial indicator pool and the two association feature pools were preprocessed through missing value handling, outlier removal, standardisation, neutralisation, and discretisation.

**Table 1. Description of Dataset Contents**

Feature Category	Feature Nature	Feature Dimension	Module in Model
Fundamental Features	Measures the profitability and other fundamental information of individual stocks	12 dimensions per stock per period	Base Model, Attention Mechanism for Sample Features
Price-Volume Features	Measures historical price and volume changes of individual stocks	33 dimensions per stock per period	Base Model, Attention Mechanism for Sample Features
Macro Features	Measures macroeconomic environment and market volatility trends	45 dimensions per period	Attention Mechanism for Market Volatility

*Source:* Authors' processing.

## 4. Model Evaluation and Application

### 4.1 Evaluation Methods

To achieve profitable trading, the rate of return on stocks purchased and sold must exceed the total transaction cost. The total transaction cost is given by:

$$\text{Total Cost} = 2 \times (c + f + s) + t \quad (16)$$

where  $c$  is the commission rate,  $f$  is the transfer fee rate,  $s$  is the one-way slippage, and  $t$  is the stamp duty rate.

According to the current regulations of the Chinese stock market, the trading commission rate ranges from 0.02% to 0.03%, levied on both buy and sell orders; the stamp duty is 0.1%, levied only on sell orders; and the transfer fee is 0.002%, levied on both buy and sell orders. Furthermore, according to relevant studies, the market impact cost for trading ¥250,000 is approximately 23–24 basis points. Taking

the commission fee as 0.02% and the slippage as 0.23%, the calculated minimum required rate of return is approximately 0.6%.

Thus, stocks with a return greater than 0.6% in the next month are labelled as positive examples, while others are labelled as negative examples.

Precision (PRC) is selected as the evaluation metric. In the context of quantitative stock selection, investors are more concerned with "buying the right stocks" than with "buying every right stock." Consequently, Precision is a more important metric than Recall or Accuracy. Therefore, Precision is selected as the primary metric. Let  $TP$  and  $FP$  denote the numbers of True Positives and False Positives, respectively. The formula for Precision is:

$$Precision = \frac{TP}{TP + FP} \tag{17}$$

Although precision measures the model's overall predictive ability, it does not fully capture the ranking ability that is crucial for portfolio construction. Therefore, RankIC is also adopted to evaluate the model. This metric captures the correlation between predictions and actual returns, rather than just the correctness of predictions. It provides a superior perspective for analysing whether a factor can effectively predict future stock performance. Its formula is:

$$RankIC = Corr(Rank(Predicted Proba), Rank(Actual Return)) \tag{18}$$

Four time periods (as shown in Table 2) were selected as test sets. Periods 1 and 2 represent relatively stable market conditions, while Periods 3 and 4 are characterised by market transitions with significant volatility.

**Table 2. Test Period Selections**

Time	Test Period	Market Characteristics
Period1	July 2020 - December 2020	Stable upward trend
Period2	July 2021 - December 2021	Stable downward trend
Period3	January 2022 - June 2022	Transition from upward to downward
Period4	January 2021 - June 2021	Transition from downward to upward

Source: Authors' processing.

The model's hyperparameters are set as shown in Table 3. The initial learning rate is set to 0.001, with the Adam optimiser used for dynamic adjustment. The Xavier method is employed for initialising the weight matrices.

**Table 3. Model Hyperparameter Settings**

Parameter	Setting
Learning Rate	0.001
Weight Matrix Initialisation	Xavier
Optimiser	Adam
Latent Space Dimensions	16,32,64,96,128
Loss Function $\lambda$ Values	0,0.001,0.01,0.04,0.08

Source: Authors' processing.

## 4.2 Comparative Experiments

Three types of models were used as baselines for the ensemble algorithm: (1) Simple ensemble methods, BAG (Breiman, 1996) and WMV (Bagui,2005); (2) Ensemble methods considering additional factors, MetaSVM (Wolpert,1992) and ConvexDS (Yin et al., 2014); (3) Ensemble methods incorporating dynamic approaches, DDES (Choi & Lim,2021) and AEM (Liu et al., 2019).

As shown in Table 4, the baseline models and the proposed model "Ours" exhibit a clear hierarchical distinction in relatively stable market periods. The "Ours" model outperforms in both periods, with a 1.07% improvement in the PRC over the best baseline model during the stable upward trend and a 0.92% improvement during the stable downward trend.

**Table 4. Comparison in Stable Market Periods**

Model	Period 1: Upward		Period 2: Downward	
	PRC	RankIC	PRC	RankIC
BAG	0.5390	0.0495	0.5371	0.0501
WMV	0.5433	0.0527	0.5308	0.0456
MetaSVM	0.5378	0.0444	0.5414	0.0494
ConvexDS	0.5605	0.0516	0.5628	0.0547
DDES	0.5802	0.0595	0.5665	0.0580
AEM	0.5814	0.0613	0.5581	0.0599
Ours	0.5876	0.0631	0.5717	0.0622

*Source:* Authors' processing.

For the RankIC metric, which measures stock selection capability, the "Ours" model maintains a comparative advantage, achieving 0.0631 and 0.0622 in the two periods, representing improvements of 2.94% and 3.84% over the best-performing AEM model, respectively.

As shown in Table 5, the baseline models and the "Ours" model still exhibit a clear hierarchical distinction in volatile market periods. The "Ours" model maintains its advantage, with a 1.98% improvement in PRC over the best baseline model in Period 3 and a 2.33% improvement in Period 4, significantly larger than the improvements observed in Periods 1 and 2. For the RankIC metric, the "Ours" model shows substantial improvements, with increases of 0.0064 and 0.0059 over the best-performing models in the two periods, representing percentage improvements exceeding 10%. This indicates that the "Ours" model has significantly enhanced predictive capability and stability during periods of high market volatility.

**Table 5. Comparison in Volatile Market Periods**

Model	Period 3: Upward to Downward		Period 4: Downward to Upward	
	PRC	RankIC	PRC	RankIC
BAG	0.5014	0.0379	0.4947	0.0380
WMV	0.4997	0.0331	0.5017	0.0413

Model	Period 3: Upward to Downward		Period 4: Downward to Upward	
	PRC	RankIC	PRC	RankIC
MetaSVM	0.5016	0.0402	0.5102	0.0409
ConvexDS	0.5357	0.0444	0.5474	0.0509
DDES	0.5453	0.0527	0.5529	0.0512
AEM	0.5514	0.0503	0.5416	0.0519
Ours	0.5623	0.0591	0.5658	0.0578

Source: Authors' processing.

### 4.3 Ablation Experiments

Three ablation models were designed as follows: (1) Model *A*: The full model without the attention module for market volatility, retaining the feature engineering method and the attention module for sample features. (2) Model *B*: Without the attention modules for market volatility and sample features, retaining only the feature engineering method, it degenerates into a stock selection model similar to the BAG algorithm. (3) Model *C*: Further removing the feature engineering method from Model *B*, resulting in a BAG ensemble model using only raw features.

Tests were conducted in the four selected periods, and the results were compared with the full model using PRC and RankIC metrics, as shown in Figure 2 and Figure 3.

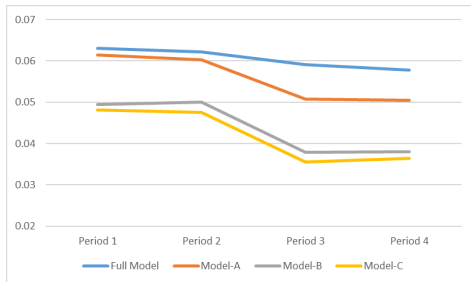
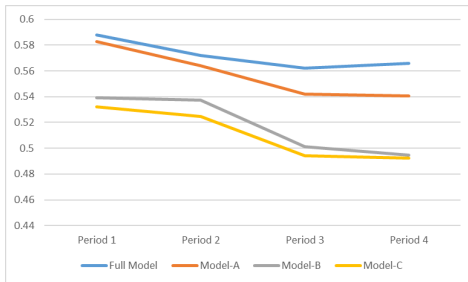


Figure 2. Ablation Experiment (PRC)

Figure 3. Ablation Experiment (RankIC)

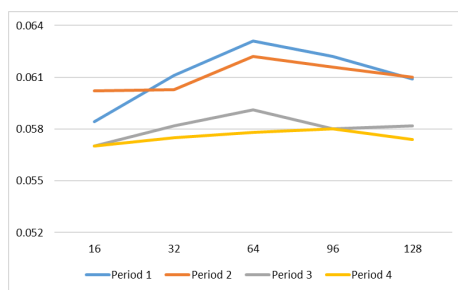
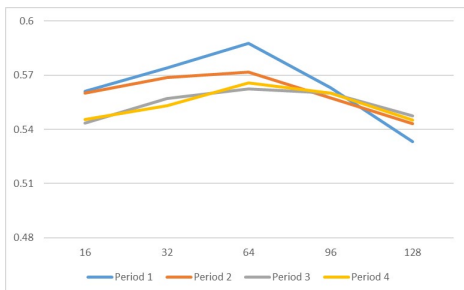
Source: Illustration by authors.

The results indicate the following: (1) Compared to the full model, Model *A* exhibits a decline in both PRC and RankIC metrics across all periods. (2) Compared to model *A*, a significant performance degradation of model *B* is observed, with noticeable declines in PRC and RankIC metrics. (3) In contrast to Model *B*, Model *C* shows a further decline in performance, with continued decreases in PRC and RankIC values, suggesting that the model lacks sufficient stock selection capability overall.

Therefore, the ablation experiments demonstrate that the model designed in this study exhibits strong stability and overall predictive performance.

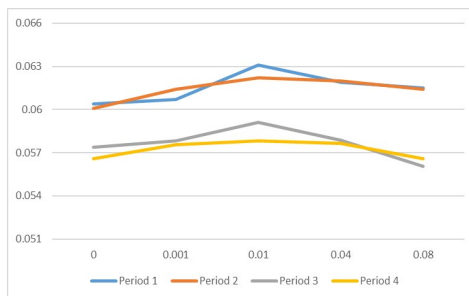
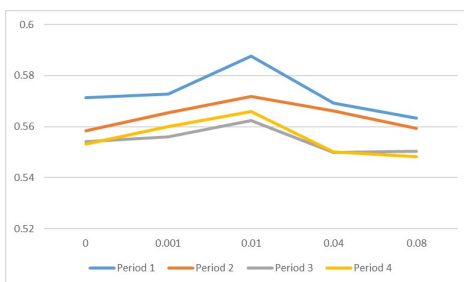
### 4.4 Hyperparameter Analysis

Using PRC and RankIC metrics, the impact of hyperparameters on model performance was evaluated in the four selected periods. As shown in Figures 4 and the 5, both the PRC and RankIC metrics increase as the latent space dimension grows from 16 to 32, peaking at 64, after which they decline.



**Figure 4. Latent Space Dimension (PRC)** **Figure 5. Latent Space Dimension (RankIC)**  
 Source: Illustration by authors.

Keeping other hyperparameters unchanged, the value of the  $\lambda$  coefficient was adjusted and tested. As illustrated in Figure 6 and Figure 7, the model achieves the best performance when the  $\lambda$  coefficient is set to 0.01.



**Figure 6. lambda Coefficient (PRC)** **Figure 7. lambda Coefficient (RankIC)**  
 Source: Illustration by authors.

### 4.5 Backtesting

The model proposed in this study is applied for a total of 3.5 years from January 2020 to June 2023, during which time the stock market has experienced a long period of intense volatility. The backtest parameters are set in Table 6.

**Table 6. Backtesting Parameters**

Parameter	Setting	Parameter	Setting
Underlying Assets	CSI 300 Constituents	Initial Capital	100 million CNY
Stock Selection Period	Jan 2020 - Jun 2023	Commission Rate	3‰
Rebalancing Frequency	Last trading day of month	Stamp Duty	0.1%
Portfolio Construction	Equal-weighted buying	Transaction Fee	0.2‰
Number of Stocks	30	Slippage	0.1%

*Source:* Parameters used by authors.

The portfolio is rebalanced monthly, holding 30 stocks. The model is retrained every six months. The initial capital is set at 100 million CNY, with a commission rate of 3‰, stamp duty of 0.1%, transaction fee of 0.2‰, and slippage of 0.1% to simulate real market conditions.

The selected quantitative strategy evaluation metrics include total return, monthly average return, annualised return, volatility, Sharpe ratio, and maximum drawdown. The Quantisation Strategy Metrics and the backtesting results of the model are presented in Table 7. The backtesting results indicate that the quantitative strategy constructed based on the stock selection model proposed in this study demonstrates strong practical value.

**Table 7. Quantisation Strategy Metrics**

Metric	Value	Metric	Value
Total Return	133.14%	Total Excess Return (vs CSI 300)	131.84%
Monthly Average Return	3.33%	Annualised Return	27.36%
Annualised Volatility	23.90%	Sharpe Ratio	1.02
Maximum Drawdown	18.73%	-	-

*Source:* Authors' processing.

## 5. Conclusions

By incorporating association analysis into feature engineering and introducing hierarchical attention mechanisms in model integration, the proposed model dynamically allocates weights to base models, effectively adapting to changing market conditions. Experimental results demonstrate the model's strong performance.

However, the study has the following limitations:

Firstly, due to computational constraints and data availability, the model was tested only on historical constituents of the CSI 300 index. A broader validation across the entire A-share market, comprising over 5,000 stocks, would provide stronger evidence.

Secondly, the model relies on a set of fundamental, price-volume, macroeconomic, and market index metrics. While reasonable, the dataset is limited by available features. Higher-quality features could further enhance model performance.

Thirdly, the ensemble framework assumes the effectiveness of base models. While the current selection of nine base models is reasonable, future advancements in machine learning may offer superior alternatives, potentially improving the ensemble model's performance.

## References

---

- [1] Adachi, A., Yamashita, T., Kanaya, S., Kosugi, Y. (2023), *Ensemble Machine Learning Approaches Based on Molecular Descriptors and Graph Convolutional Networks for Predicting the Efflux Activities of MDR1 and BCRP Transporters*. *The AAPS Journal*, 25(5), 1-10.
- [2] Bagui, S.C. (2005), *Combining Pattern Classifiers: Methods and Algorithms*. *Technometrics*, (47), 517-518.
- [3] Breiman, L. (1996), *Bagging Predictors*. *Machine Learning*, (24), 123-140.
- [4] Chai, Y. (2024), *Emotion Intensity Detection in Online Media: An Attention Mechanism Based Multimodal Deep Learning Approach*. *Tehnicki vjesnik–Technical Gazette*, 31(2), 587-595.
- [5] Chan, K., Hameed, A. (2006), *Stock Price Synchronicity and Analyst Coverage in Emerging Markets*. *Journal of Financial Economics*, 80(1), 115-147.
- [6] Choi, Y.R., Lim, D.J. (2021), *DDES: A Distribution-Based Dynamic Ensemble Selection Framework*. *IEEE Access*, (9), 40743-40754.
- [7] Cunningham, P., Carney, J. (2000), *Diversity versus Quality in Classification Ensembles Based on Feature Selection*. *European Conference on Machine Learning*. (Jan), 1-9.
- [8] Dikici, S., Bilgin, T.T. (2025), *Advancements in Automated Program Repair: A Comprehensive Review*. *Knowledge and Information Systems*, 67(6), 4737-4783.
- [9] Duan, B.L., Wang, R.F., Zhang, R. (2022), *Economic Links and Stock Returns in Chinese A—Share Market*. *Journal of Financial Research*, (2), 171-188.
- [10] Feng, S.B., Xu, C., Zuo, Y., Chen, G., Lin, F., XiaHou, J.B. (2022), *Relation-Aware Dynamic Attributed Graph Attention Network for Stocks Recommendation*. *Pattern Recognition*, (121), 108-119.
- [11] Ge, Y.B., Liu, W.J., Gu, Y.C. (2024), *Stock Price Prediction Method by Fusing Sentiment Analysis and GAN-TrellisNet*. *Computer Engineering and Applications*, 60(12), 314-324.
- [12] Htun H.H., Biehl M., Petkov N. (2023), *Survey of Feature Selection and Extraction Techniques for Stock Market Prediction*. *Financial Innovation*, 9(1).
- [13] Kim H, Moon K.S. (2024), *Optimal Data Construction in Supervised Machine Learning for Financial Prediction*. *Economic Computation and Economic Cybernetics Studies and Research*, 58(1), 5-20.
- [14] Ko, H., Wang, B., Lim, J.S. (2024), *Grad-CAM-Based Feature Selection and Dementia Classification Algorithm Using Voice Data*. *Tehnicki vjesnik–Technical Gazette*, 31(6), 2036-2044.

- [15] Lee, W.S., Lee, H.S. (2024), *Asymmetric Volatility Connectedness Among G7 Stock Markets. Economic Computation and Economic Cybernetics Studies and Research*, 58(1),87-103.
- [16] Li, B., Tu, X.Y. (2024), *Research on Portfolio Selection Based on Machine Learning and Asset Characteristics. Systems Engineering - Theory & Practice*, 44(1), 338-359.
- [17] Li, X.H., Wang, J., Jia, H.D., Xiao, L. (2022), *Stock Market Volatility Prediction Method Based on Graph Neural Network with Multi-Attention Mechanism. Journal of Computer Applications*, 42(7), 2265-2273.
- [18] Liu, A.Y., Yue, D.Z., Chen, J.L., Chen, H. (2024), *Deep Learning for Intelligent Production Scheduling Optimization. International Journal of Simulation Modelling*, 23(1), 172-183.
- [19] Liu, H.Z., Du, Y.P., Wu, Z.H. (2019), *AEM: Attentional Ensemble Model for Personalized Classifier Weight Learning. Pattern Recognit*, 96(7), 1-8.
- [20] Liu, Z., Liu, X., Zhou, L. (2024), *A Hybrid CNN and LSTM based Model for Financial Crisis Prediction. Tehnicki vjesnik–Technical Gazette*, 31(1), 185-192.
- [21] Mesa, J.P., Montoya, A., Ramos-Pollan, R., Toro, M. (2025), *Machine-Learning Component for Multi-Start Metaheuristics to Solve the Capacitated Vehicle Routing Problem. Applied Soft Computing*, (173), 112916-112916.
- [22] Niu, Z., Zhong, G., Yu, H. (2021), *A Review on the Attention Mechanism of Deep Learning. Neurocomputing*, (452), 48-62.
- [23] Parsons, C.A., Sabbatucci, R., Titman, S. (2020), *Geographic Lead-Lag Effects. The Review of Financial Studies*, 33(10), 4721-4770.
- [24] Qin, C., Chang, J., Tu, W.T., Yu, C.R. (2024), *Follow AKO Investor: Stock Recommendation by Hearing Voices from all Kinds of Investors with Machine Learning. Expert Systems with Applications*, 249(B).
- [25] Song, Z.Y., Zhou, Z.B., Yu, L.A., Ren, T.T. (2024), *Portfolio Optimization Strategy with a hybrid Ensemble Forecasting Algorithm and Black-Litterman Model. Chinese Journal of Management Science*, (5), 1-12.
- [26] Praveena, S., Prasanna Devi, S. (2024), *A Sporadic Classification and Regression-Based Approach to Intermittent Demand Forecasting in Smart Supply Chain. Tehnicki vjesnik–Technical Gazette*, 31(5), 1544-1552.
- [27] Thiruvenkatasamy, S., Sivaraj, R., Vijayakumar, M. (2024), *Blockchain Assisted Fireworks Optimization with Machine Learning based Intrusion Detection System (IDS). Tehnicki vjesnik–Technical Gazette*, 31(2), 596-603.
- [28] Tsinaslanidis, P.E., Kugiumtzis, D. (2014), *A Prediction Scheme Using Perceptually Important Points and Dynamic Time Warping. Expert Systems with Applications*, 41(15), 6848-6860.
- [29] Wolpert, D.H. (1992), *Stacked Generalization. Neural Networks*, 5(2), 241-59.
- [30] Xie, W., He, J., Huang, F., Ren, J. (2024), *Supply Chain Financial Fraud Detection Based on Graph Neural Network and Knowledge Graph. Tehnicki vjesnik–Technical Gazette*, 31(6), 2055-2063.

- [31] Yin, X.C., Huang, K.Z., Yang, C., Hao H.W. (2014), *Convex Ensemble Learning with Sparsity and Diversity*. *Information Fusion*, (20),49-59.
- [32] Zhang, Q., Gong, Y.X., Zhang, X., Cai, X.L., Zheng, J. (2023), *Intelligent Passenger Flow Analysis System for Urban Rail Transit Trains Based on Heterogeneous Ensemble Learning Method*. *Railway Computer Application*, 32(7), 73-78.
- [33] Zhang, X. (2024), *Visual Communication Design of Weak and Small Target Images Based on Image Processing Model and Data Fusion*. *Tehnicki vjesnik–Technical Gazette*, 31(3), 815-823.
- [34] Zhou, L. (2021), *Research on Multi-Factor Stock Investment Based on Random Forest Model*. *Financial Theory & Practice*, (7), 97-103.