

Retraction

Retracted: Fuzzy Decision-Making Analysis of Quantitative Stock Selection in VR Industry Based on Random Forest Model

Journal of Function Spaces

Received 13 September 2023; Accepted 13 September 2023; Published 14 September 2023

Copyright © 2023 Journal of Function Spaces. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

- [1] J. Zhu, Y. Geng, W. Li, X. Li, and Q. He, "Fuzzy Decision-Making Analysis of Quantitative Stock Selection in VR Industry Based on Random Forest Model," *Journal of Function Spaces*, vol. 2022, Article ID 7556229, 12 pages, 2022.

Research Article

Fuzzy Decision-Making Analysis of Quantitative Stock Selection in VR Industry Based on Random Forest Model

Jia-Ming Zhu ¹, Yu-Gan Geng ², Wen-Bo Li ², Xia Li ², and Qi-Zhi He ³

¹School of Statistics and Applied Mathematics, Anhui University of Finance and Economics, Bengbu 233030, China

²School of Finance, Anhui University of Finance and Economics, Bengbu 233030, China

³School of Statistics and Mathematics, Zhejiang Gongshang University, Hangzhou 310018, China

Correspondence should be addressed to Qi-Zhi He; happyhefei2000@163.com

Received 16 April 2022; Accepted 20 May 2022; Published 7 June 2022

Academic Editor: Muhammad Gulzar

Copyright © 2022 Jia-Ming Zhu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Since Professor L.A. Zadeh published “Fuzzy Set Theory” in the 1960s, the theory of fuzzy mathematics has been formally established and developed and has been gradually introduced into work in all walks of life. At the same time, fuzzy mathematics theory has also been widely used in VR industry selection. In the stock strategy, the advantages of improving unit classification accuracy, screening high-quality stocks, and constructing near-perfect investment portfolios continue to emerge. On the other hand, with the increasing maturity and continuous development of China’s computer and Internet technologies, the VR industry has gained a new round of development space, and its own investment value and the investable space between related industries have been gradually tapped. Different from the analysis of quantitative stock selection by constructing a logistics multifactor stock selection model in the existing research, the research mainly adopts the random forest algorithm based on fuzzy mathematics to construct the initial investment strategy portfolio. Secondly, different from the single effective frontier algorithm, the research is based on the random forest algorithm, calculates the average AUC of the index, and continuously checks and tests the results to obtain the optimal investment portfolio. Finally, select appropriate risk indicators and performance indicators to evaluate the performance of the strategy portfolio. The study concludes that the portfolios selected by the random forest model are highly investable and have good stability.

1. Introduction

1.1. Background of the Selected Topic. In recent years, the rapid development of the VR (Virtual Reality) industry has attracted the attention of more and more investors. From an innovative point of view, the VR industry is a new concept for investors [1], and its investment potential value is huge; in terms of industrial development, the VR industry drives the development of not only the digital media industry but also the computer technology industry on which it relies, which forms a small industry rotation and drives the industrial development of multiple enterprises [2].

In addition, the introduction of the strategy and basic idea of “science and education for the development of the country” and “technology-led development” is a reflection of the increasing importance China is placing on the VR industry. The Chinese government has also introduced a

series of policies to support the development of China’s VR industry and encourage investors to invest in the industry. With the strong support of national policies for the virtual industry, it has led to rising investor confidence in this industry and a growing emphasis on its investment potential.

Ultimately, from the perspective of Moore’s Law, the core technology companies in the VR industry will receive more support, thus promoting the improvement and development of VR technology. However, there are still certain problems with the development of the VR industry, such as the lack of perfect performance of VR all-in-one machines and the difficulty of rapid development of wireless transmission technology in the excessive stage. Therefore, with the continuous development of the key role technology to improve these problems now, front-end technology is more likely to be sought after by capital.

1.2. Literature Review. In recent years, China's economic power has grown and per capita income has increased significantly; in the meantime, income, income disparity, and income inequality will influence the public happiness [3], and thus, not only investors but also scholars pay more attention to investment portfolios, so the research on quantitative stock selection, a key factor in portfolio construction, has also evolved.

[4]. One of the most common methods is multifactor quantitative stock selection [5]; however, the multifactor quantitative stock selection is the same, the research angle is different, the construction of the portfolio is also different, and the actual portfolio return obtained in the end is also different [6].

The most common perspective is the multifactor quantitative stock selection method based on the Boosting algorithm (XGBoost algorithm): Wang mentioned in his quantitative stock selection research on this algorithm that compared to the traditional static multifactor stock selection model, the use of rolling training methods to build a dynamic stock selection model is more capable of improving the profitability of the final stock selection model, thus improving the investor's desired investment strategy [7, 8]. Further, Zhu combined the LightGBM algorithm with the XGBoost algorithm, compared the relative advantages and disadvantages of the two in quantitative stock selection by testing the model validity and decision tree structure [9], and finally concluded that although both have good prediction ability, the prediction ability of the LightGBM algorithm is obviously stronger than that of the XGBoost algorithm.

Network models are applied broadly to the stock price recently. Zhang et al. [10, 11] give empirical tests on the stock price vulnerability using topology network. In addition, more studies on quantitative stock selection models have used classifiers such as GRU neural network models or integrated tree models [12–14], and the most comprehensive one is the Stacking method, which combines the above-mentioned neural networks, gradient boosting trees, and XGBoost to form a new algorithmic model, RGXB-Stacking stock selection model, and the research results show that this model has significantly better back-testing effect on constituent stock data than other models [15–17]. Based on the above analysis of existing research, it is found that most of the research on quantitative stock selection nowadays is only at the level of optimizing the selected impact factors, and there is no good solution for the relationship between a large number of impact factors and the expected return of the model [18]. Therefore, the study extracts the random forest model algorithm from the Stacking method and conducts an empirical analysis to verify the role of the random forest model in quantitative stock selection by taking the VR industry [19–22], which is developing rapidly in today's society, as the research object [23–25].

2. Theoretical Elaborations

2.1. Phenomena Related to Industry Rotation. The study of economic cycles is essential for the most realistic evaluation

and analysis of a country's economy. Therefore, the basis of industry rotation is the economic cycle, which is used as a benchmark to make basic judgments about the relationships between industries and their ups and downs, as well as the interaction between individual industries and the overall economy [2].

The theory also guides investors to enter the capital market for investment in different economic operation cycles, so as to obtain excess returns [26]. The Merrill Lynch investment clock model theory studies the selected security research indicators and gives the corresponding analysis summary according to their development trends and finally obtains the four stages of the economic cycle, namely, recession, recovery, overheating, and stagflationary periods.

In addition, the theory predicts the average return of various types of assets in these four phases, so as to obtain the relationship between the investment return of each industry and the overall market investment return during the phase rotation of different economic cycles and also the best asset class to adapt to the market in different economic cycles. Details are shown in Table 1.

2.2. Random Forest Building Process. In the random forest building process, a decision tree is first built for each training subset, and the process is repeated N times to form a "forest" of decision trees, where each decision tree grows naturally without pruning during the development phase ([27]; Zhang et al. [10, 11]). In this process, two aspects need to be noted: first, the node splitting process. For the random forest algorithm, node splitting is the core step to generate each decision tree, and only after the complete node splitting process can a complete decision tree be formed [28].

In addition, in the process of generating branches, each decision tree selects the corresponding attributes according to some specific splitting rules, which mainly include maximizing the information gain (rate) and minimizing the Gini coefficient [29, 30]. It should be noted that the corresponding splitting algorithms are different using different rules, and the resulting decision trees are also different [31, 32]. Next is the random selection of input variables. This process refers to the process of random forest generation, in order to improve the variability between different decisions, while ensuring the classification accuracy of the random forest, so as to improve the operational performance of the whole "forest" [33, 34]. The specific generation process and classification method are shown in Figures 1 and 2.

2.3. Analysis of the Advantages of Random Forest Stock Selection Models. For quantitative stock selection, nowadays, more research models are used based on Boosting algorithm, LightGBM algorithm, and quantitative stock selection model based on neural network [35]. Although the algorithms of these models are different, they are all based on multifactor stock selection models. For multifactor stock selection models, the main process is to select a number of factors that affect the movement of stock market value and study the intrinsic linkage between these factors and the impact of different combinations on stock returns, so as to find the portfolio with the highest return for investment. Although these

TABLE 1: The performance of different industries within the four phases.

Stage	Economic growth rate	Best asset class	Best industry segment
Recession	Decline	Bonds	Defensive growth
Recovery period	Rise	Stock	Cyclical growth
Overheating period	Rise	Popular goods	Cyclical growth
Stagflationary period	Decline	Cash	Defensive growth

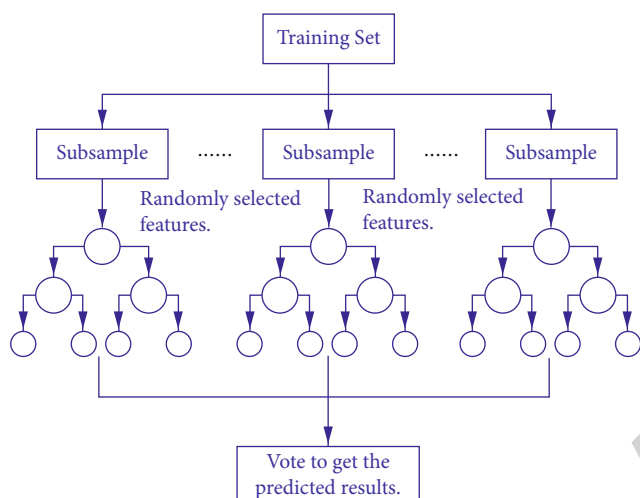


FIGURE 1: Random forest generation structure.

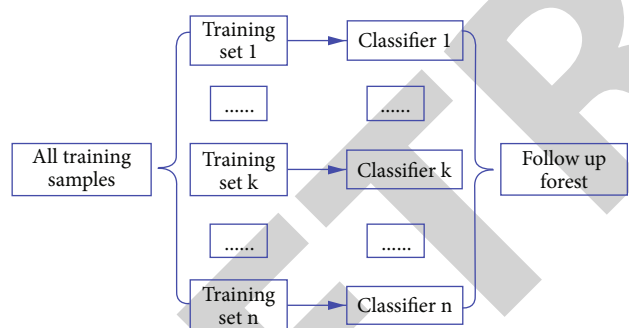


FIGURE 2: Random forest classification process.

models can intuitively reflect the impact of different factors on stock returns, they are susceptible to the influence of multicollinearity among different factors and are only applicable to the case where there are few influencing factors, and if there are more influencing factors, it is impossible to make accurate portfolio prediction [36].

The random forest model effectively solves this problem. The efficient classification process of the random forest model not only ensures that the impact of different factors on the expected return of the portfolio is accurately and intuitively reflected but also allows the same class of influencing factors to be classified, so that the impact of the same class of factors on the expected return of the portfolio can be derived [37]. This not only ensures the accuracy of the constructed quantitative stock selection model but also effectively avoids the interference of multicollinearity among different factors; in addition, in the case of a large

number of impact factors, the effective classification of them before testing the impact factors can ensure that the final constructed portfolio has less error, so that effective investment can be made.

3. Quantitative Stock Selection

3.1. Research Idea. Considering that there are hundreds of companies in the VR industry-related sector, the study decided to select 60 stocks with high total turnover from all the VR industry sector companies in the first round for preliminary screening, and the selected stocks need to meet the characteristics of high growth and high profitability. After that, 34 factors such as profitability factor, trading factor, volatility factor, and cash flow factor were selected through the establishment of quantitative stock selection research factors, and suitable research factors were selected to analyze the stocks as experimental subjects and construct suitable stock portfolios [38].

Since the characteristic factors for stock analysis and research are rather redundant, they are categorized and studied by constructing a random forest, so that suitable stocks can be screened for investment [39]. Eventually, considering that the random forest model may still have some errors, it is improved by building a deep forest model to obtain the initial screening of stocks [40].

3.2. Establishment of Stock Pool. In order to further improve the accuracy of the stock screening process and to reduce the error in the subsequent classification using random forest, the 30 stocks after the above initial screening were reprocessed and 20 stocks with ROE greater than 10% were selected for the subsequent study, as shown in Table 2.

3.3. Factor Pool Establishment and Data Preprocessing

3.3.1. Establishment of the Factor Pool. In the process of selecting factors, the study focused on the correlation between basic technical characteristics of stocks and stock returns, and the following 34 factors were selected from several perspectives, such as trading volume factor, market capitalization factor, and *hls* factor, as detailed in Table 3.

For the general multifactor stock selection model, investors are more concerned about the validity test of the research factors in the process of stock selection research; the main method is as follows: gradually select a single factor in the ranking of the score of multiple stock returns to select a more suitable stock portfolio, keep repeating several cycles, and finally, calculate the return of the stock portfolio, so as to determine if the stock portfolio can generate continuously. The effective factors that can generate continuous

TABLE 2: Top 20 stocks with net return (ROE) greater than 10%.

Stock name	Rongtai Health	Perfect World	Kaiying Network	Songcheng Performing Arts	Chinese Media	Skyworth Digital	Mason Culture
ROE	56.71%	27.13%	47.19%	17.03%	13.04%	22.18%	12.19%
Stock name	Lingnan Garden	SuperMap Software	Endeavor Technology	Zhongke Chuangda	Common Progress Shares	Tianyin Holdings	Shunnet Technology
ROE	13.11%	10.97%	15.03%	13.20%	10.47%	11.12%	25.14%
Stock name	Xinwanda	San Qi Entertainment	Jebsen shares	Goer Corporation	Anjie Technology	Crystal Photonics	—
ROE	22.20%	27.15%	13.11%	17.10%	16.41%	10.08%	—

TABLE 3: Results of the study factor selection.

Factor type	Factor name
Profit factor	Nct ProfitGrow rate, GrossIncome ratio, ROE, ROA
Trading factors	vr, vol10, vol120, vol240, vol5, vol60, vol20
Fluctuation factor	FiftyTwoWeekHigh, DVRAT, HBETA, skewness
Cash flow factor	MoneyFlow 20, current ratio
Momentum factor	MTM
Dividend factor	DilutedEPS, EPS
Leverage factor	Equity to asset, FixAsset ratio, BLEV
Size factor	LETO, accounts PayablesT rate
Averaging factor	MA10, MA120, MA20, MA5
Valuation factor	PB, PE, Hurst

returns are determined. As for the study, each stage of the model construction process consists of a random forest, so each factor in the factor pool can be categorized by itself, which has the effect of screening factors [41].

3.3.2. Data Preprocessing. During the initial acquisition of data and the corresponding modeling process, the data generated often have factors such as missing data for some study factors, unit discrepancies among factors, and incorrect values for some factors. If the first-hand data are used directly for modeling analysis, the final accuracy of the model prediction will be affected. Therefore, before model building, the first-hand data need to be preprocessed to complete the missing values, change the incorrect values, and standardize the data to unify the study factors.

3.3.3. Completing Missing Data Values. Since the data for the study mainly comes from the company's website financial statements, web data, and market research data, there may be some important data missing. For the study factors, they may contain certain missing values. For the treatment of this situation, in general, stock funds with more missing values of research factors are directly excluded, while stocks with fewer missing values of research factors can be considered for supplementary treatment using the interpolation method.

$$L_n(x) = \sum_{i=0}^n l_i(x)y_i, l_i(x) = \prod_{j=0, j \neq i}^n \frac{x - x_j}{x_i - x_j}, \quad (1)$$

where x is the position of the corresponding missing value and the approximate value of the missing value.

3.3.4. Removing Data Outliers. For the treatment of outliers, first of all, the concept of quartiles should be introduced; i.e., the left quarter and the right quarter of the whole data range become a quartile point, and the standard line at both ends is the quartile line, and the data falling outside the 3/4 and 1/4 quartile lines of the overall data value is the abnormal data, which will be rejected. However, in general, when the amount of data is small, it is necessary to correct the data at this time instead of deleting it, otherwise it will affect the overall data structure.

3.3.5. Data Normalization Process. Considering that the selected study factors may have inconsistent data units, they need to be quantified, and considering the existence of quantitative data among the factors, the corresponding data can be zero-mean standardized to improve the accuracy of the study.

$$x^* = \frac{x - \min x}{\max x - \min x}. \quad (2)$$

In addition, the zero-mean standardization was also made standard deviation normalized, and all data after processing had a mean of 0 and a standard deviation of 1.

$$x^* = \frac{x - \mu}{\sigma}. \quad (3)$$

3.4. Modeling Industry Rotation Strategies

3.4.1. Delineating the Economic Cycle. According to the economic cycle-related theory of industry rotation, different investment tools are needed at different stages of the economic cycle in order for investors to obtain higher excess returns after making investments. First, during the economic recovery period, the industries with better prospects in that period are the financial industry with assets such as stocks, bonds, and commodities, while economic inflation causes lower interest rates in the industry, so it is necessary to grasp the development of stock-based industries. Second, in the period of economic overheating, there are problems such as rapid economic development and unstable economic growth rate, and there is also a certain amount of inflation, so this period should focus on the commodity industry. Again, in the period of economic stagnation, the economic development in this period begins to gradually go downhill, the phenomenon of economic recession, but there is still inflation in the market, and gradually strengthen; the prices of major asset classes gradually decline, so this period should grasp the cash. Finally, in the period of economic decline, the market economy in this period has been in sharp recession and the inflation phenomenon is more serious, and the growth of the market economy needs to rely on the continued reduction of commodity interest rates to ensure that the market bonds and other industries maintain a rebound.

In summary, in the process of stock selection and investment by investors based on different economic development cycles, it is most important to invest in different categories of asset stocks; therefore, it is especially critical to delineate the appropriate economic cycle. For the division of the economic cycle, factors such as inflation indicators and economic growth need to be considered, which also relates to the effect of the investor's final asset allocation. For a country's macroeconomic development, its leading and lagging indices can best visualize a country's overall economic situation and forecast the subsequent development accordingly [2].

Therefore, the study selects two indicators, leading and lagging indices, to classify the economic cycle. The study simultaneously selects a one-year time period on the basis of preselected data and analyzes the applicability of sector rotation in the Chinese stock market accordingly based on the investment clock principle to arrive at the monthly indexes in Figure 3.

According to Figure 3, it visually reflects the changes of the two indices in a year, and it can also visualize the inflection point time of the two indices. According to the division of the economic cycle, it can be seen that in the economic recovery period, the leading index rises and the lagging index falls; in the economic overheating period, the leading index rises and the lagging index rises; in the economic stagnation period, the leading index falls and the lagging index rises; in the economic recession period, the leading index falls and the lagging index falls. Therefore, the economic

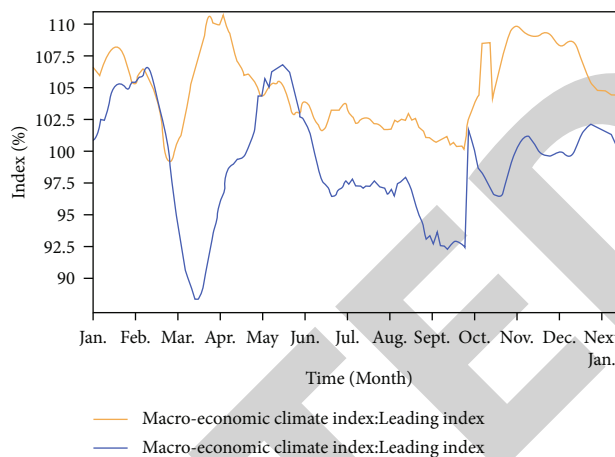


FIGURE 3: Monthly data fluctuations of leading and lagging indices.

cycle can be divided accordingly according to the combination of the changes of the two indices in different periods mentioned above. The details are shown in Table 4.

3.4.2. Allocation of Rotatable Sectors. After dividing the economic cycle as described above, it is necessary to divide the studied stock objects into industry indices. For the 20 groups of VR industry stocks selected for the study, they can be divided into 6 industry indices such as energy, medicine and health, film and media, information technology, telecommunication business, and public utility industry, as shown in Table 5. In addition, the above results and investment clock theory make a corresponding comparison of the industry index classification; the fit is high, so it can be well applied to the study of industry rotation and can be descriptive statistics and ADF test between different industry indices.

As shown in Table 6, it represents the monthly returns of 6 industries. From a local perspective, the average monthly return of the film and television industry is the highest, being top 1 of the 6 industries. However, in terms of its volatility, its minimum value is -25.13% and 17.14% , and its average monthly return is 1.12% , which can be seen that the film and television industry is not very risky, but its development level is correspondingly low. In addition, from an overall perspective, the industry index skewness of each industry is below 0, the kurtosis is mostly above 1, the part below 1 is mostly greater than 0.5, and the part above 1 is basically at the position of 3. Therefore, it can be concluded that the overall return pattern is all spiky and thick-tailed left-skewed. Meanwhile, the two columns of ADF1 and ADF2 in Table 6 are the unit root tests of the raw price series of the industry indices, and from their significance, it can be concluded that all six industries are first-order single-integer indices.

3.4.3. Cointegration Tests of the Two Macroeconomic Indices. Unlike foreign markets, one of the most important factors influencing the macroeconomic and capital market changes in China is the Chinese national policy, which also has a certain influence on the Chinese stock market. In addition, for

TABLE 4: Economic cycle segmentation results.

Indicators	Period
Economic recovery period	February-March; late May-June; early to mid-September
Economic overheating period	March-April; July-mid-July; early September; mid-September-early October
Period of economic stagnation	April-May; November-December
Economic recession	January-February; May; June-July; mid-July-September; early October-November

TABLE 5: Industry index results.

Industry index	Related companies
Energy	Xinwanda, Crystal Photonics
Medicine and health	Rongtai Health
Film & Media	Perfect World, Songcheng Performing Arts, Chinese Media, Skyworth Digital, Meisheng Culture, Sanjiu Entertainment, Jie Cheng
Information technology	Kaiying Network, Chaotu Software, Endeavor Technology, CKC, Gongjin, Shunnet Technology, Anjie Technology
Telecom business	Tianyin Holdings, Geer Corporation
Public utilities	Lingnan Garden

TABLE 6: Descriptive statistics and ADF tests by industry.

Category	Average value	Standard deviation	Minimum value	Maximum value	Skewness	Kurtosis	ADF1	ADF2
Energy	0.1%	8.72	-30.73%	25.01%	-0.13	1.35	-3.1007	-4.9875
Medicine	0.8%	6.94	-21.01%	12.83%	-0.47	0.09	-2.4273	-4.4157
Film & TV	1.12%	7.11	-25.13%	17.14%	-0.52	0.71	-2.1413	-4.9316
Information	0.35%	8.84	-27.4%	20.07%	-0.39	0.83	-2.0121	-5.1009
Telecommunications	0.09%	9.01	-31.08%	36.16%	-0.1	2.59	-2.4149	-4.1733
Public	0.27%	7.03	-27.91%	29.08%	-0.1	4.27	-2.4813	-3.9115

the Chinese industry indices, there are certain errors in the analysis of the economic cycle; therefore, in order to reduce the errors, the relationship between the industry indices and the economic cycle needs to be processed before the research is conducted, so as to increase the feasibility of the research and to ensure that the historical data counted in the research process fits well with the real situation [2]. The results of the cointegration test between industry indices and economic cycles are shown in Table 7.

The experimental results reflected in Table 7 examine the correlation between the industry indices and the related economic cycles from an econometric perspective.

3.4.4. Industry Index Rotation Strategy Configuration Analysis. The degree of correlation between the two macro-economic indicators is confirmed by the cointegration test between the industry indices and the economic cycle in Table 7 above. In addition, in order to observe and analyze the industry indices in different economic cycles from various aspects, the monthly average returns are selected for statistical purposes, as detailed in Table 8.

From Table 8, it can be seen that different industries do not have the same status in different economic cycles, so it is possible to construct the corresponding industry rotation strategy models for the industry indices and to count their

basic returns in the sample interval. In addition, in order to more intuitively reflect the effect of the industry rotation strategy model construction, the study additionally selects three industry indices for separate purchase strategies, in addition to the VR industry version of the index as the industry index benchmark in order to have a more intuitive standard when making comparisons. The details are shown in Table 9.

As seen in Table 9, the sector rotation strategy produced a significant annualized return of 18.05% over the sample period; however, its maximum retracement rate and corresponding Sharpe ratio were 51.04% and 1.80%, respectively. In addition, in the separate sector index buy-and-hold strategy, pharmaceuticals and health care performed better, but the effect was not significant compared to the sector rotation strategy. Thus, based on the sector rotation strategy model established above, investors can select the sector rotation strategy to handle the underlying assets when making investments ([23, 24]).

3.5. Building a Random Forest Stock Selection Model

3.5.1. Analysis of Construction Ideas. The study selects stock data for the corresponding year, where the stock data during the half-year period is used for the training factors of

TABLE 7: Industry index and economic cycle cointegration test results.

Category	Eigenvalue	Statistical quantities	5% critical value	<i>P</i> value
Energy	0.201	37.28	29.79	0.0061***
Medical health	0.193	37.31	29.79	0.0061***
Film & media	0.157	31.71	29.79	0.00891*
Information technology	0.243	40.57	29.79	0.0024***
Telecom business	0.231	40.72	29.79	0.0018***
Public utilities	0.142	38.75	29.79	0.0041***

TABLE 8: Monthly average returns for each sector index over different economic cycles.

Category	Recovery period	Overheating period	Stagflationary period	Recession
Energy	-0.27%	2.61%	-0.29%	-1.44%
Medical health	1.94%	1.99%	1.97%	-0.85%
Film & media	2.77%	2.84%	1.83%	-0.25%
Information technology	1.21%	1.52%	0.82%	-0.21%
Telecom business	2.25%	0.77%	0.71%	-0.77%
Public utilities	0.26%	0.68%	1.81%	-1.16%

TABLE 9: Sector rotation strategy model and individual index strategy returns.

Category	Industry rotation strategy	Energy	Medicine and health	Film & media	VR industry
Annualized yield	18.05%	3.92%	15.35%	7.13%	5.13%
Earnings standard deviation	11.21%	12.03%	9.41%	11.52%	9.87%
Maximum retracement	-51.04%	-60.89%	-50.40%	-60.29%	-60.88%
Sharpe ratio	1.80	0.37	1.67	0.65	0.52

the random forest and the optimization parameters, and then predicts the forecast of the data model for the next year. Finally, before constructing the random forest, we need to use the industry rotation strategy and the investment clock theory to divide the economic cycle accordingly, so as to count the returns of each stage, and select the top 2 industries under different stages for the random forest model according to their results, as shown in Table 10.

3.5.2. Random Forest Construction Process. For the random forest model, the main components are the number of decision trees B and the number of splitting points A in each decision tree, which will have some influence on the random forest algorithm process, where B represents the number of decision trees in the random forest and A represents the number of splitting points in each decision tree; in general, the more trees in the random forest, the higher the accuracy of the model construction [42]. In general, the more the number of decision trees in the random forest, the higher the accuracy of the constructed model, but it will reduce the training response speed of the model. In addition, the number of split points per decision tree is always smaller than the number of feature points in the random forest T . In addition, as in A and B , the training response speed of the constructed model is slowed down by the increase in

the number of split points in the decision tree, and the diversity of the random forest is also reduced.

In summary, the lesser the variability in the set of disaggregated features of each decision tree in the random forest, the lesser the variability between each decision tree, and the diversity of individual decision trees decreases, which in turn affects the accuracy of the random forest model prediction. Therefore, in the process of constructing the model, it is extremely important to select the value of A . Usually, 20% A , $\log(2T)$, and \sqrt{T} are used. The resulting process of random forest model construction can be obtained as follows.

Find the optimal parameters of the random forest algorithm in the sample data while preprocessing the data for normalization, missing values, and outliers.

The selected stocks are binarized 0/1, and the study treats the stock returns as definite class data, marking the stock returns present in Table 10 as 1 and the rest of the stock returns in the corresponding positions (the bottom 2 monthly return rows) as 0 and eliminating the intermediate data.

Optimize the selected data by means of a crossover network, *i.e.*, evaluation and analysis of the effect of each set of optimized parameters.

Assign values to A and B to form the corresponding combinations, and use the fivefold cross-validation method

TABLE 10: Top 2 industry returns at different stages.

Industry	Recovery period	Industry	Overheating period	Industry	Stagflationary period	Industry	Recession
Film & TV	2.77%	Film & TV	2.84%	Medicine	1.97%	Information	-0.21%
Telecommunications	2.25%	Energy	2.61%	Film & TV	1.83%	Film & TV	-0.25%

TABLE 11: Mean AUC test results for different parameter combinations.

Number of splitting points (A -value)	Decision tree trees (B)	Average AUC	Number of splitting points (A -value)	Decision tree trees (B)	Average AUC
2	10	0.557	4	10	0.562
2	20	0.560	4	20	0.561
2	30	0.563	4	30	0.562
2	40	0.560	4	40	0.566
2	50	0.561	4	50	0.566
2	60	0.568	4	60	0.563
2	70	0.567	4	70	0.566
3	10	0.557	5	10	0.558
3	20	0.561	5	20	0.564
3	30	0.569	5	30	0.560
3	40	0.561	5	40	0.569
3	50	0.560	5	50	0.571
3	60	0.577	5	60	0.567
3	70	0.575	5	70	0.565

to test the average AUC of each group of samples, and the test results are used as the criterion for judging the effectiveness of each group of optimization parameters. Among them, the average AUC is a common dichotomous evaluation tool, which is generally used to judge the degree of merit of a certain prediction model. If the $AUC > 0.7$, the prediction model has a better effect, and the effect of the prediction model and the average AUC vary proportionally.

3.5.3. Solving the Model. According to the above process of establishing the random forest model, the corresponding operations were performed to combine A and B in two sets, and different 21 sets of parameter combinations were obtained. In addition, the number of decision tree split points A was set to $[2, 3, 4]$ when the crossgrid search was performed, and the trees B of the decision tree in the random forest were $[10, 20, 30, 40, 50, 60, 70]$. The average AUC of each set of parameter combinations is tested and calculated using the random forest algorithm, and the results are shown in Table 11.

Therefore, combining the data in the table, we can get that when the number of splitting feature points of the decision tree A is 3 and the number of decision trees B in the “forest” is 60, the classification effect of the random forest model is the best and the average AUC is the largest, so $A = 3$ and $B = 60$ are chosen as the optimal operating parameters in the random forest.

3.6. Constructing the Optimal Stock Portfolio. According to the application of the above, random forest algorithm process in the stock selection strategy can be used to classify the 20 stocks in Table 1 for stock selection, where the optimal number of split points for each decision tree, i.e., each study interval, is 3. In addition, combined with the division of different economic cycles in Table 4, the optimal portfolio of stock selection for the investment strategy in the current year’s study interval can be derived, as shown in Table 12.

4. Strategy Portfolio Validity Test

Investors will consider comprehensively yield and risk, and their recognition abilities and matching consciousness for yield and risk play important effects on successful investments [36]. The optimal allocation of stocks in each period based on the sector rotation perspective is obtained by using the random forest model for stock selection in a given stock, but its risk and performance have not been further verified.

4.1. Strategy Risk Evaluation Indicators

4.1.1. Beta. Beta is used to describe the systemic risk of the investor’s investment products, which reflects the sensitivity of the corresponding investment strategy to changes in the general stock market. Generally speaking, the mathematical meaning of Beta is as follows: for every 1% change in the general stock market, the percentage change in Beta of the investment strategy in the same direction, as follows:

$$\beta = \frac{\text{cov}(p_n, B_n)}{\sigma_B^2}, \quad (4)$$

where p_n is the daily return of the strategy, B_n is the daily return of the benchmark strategy, the denominator of the equation is the variance of the daily return of the benchmark strategy, and the numerator is the covariance of the daily return of the strategy and the benchmark.

4.1.2. Alpha Value. For investors, an investment activity is bound to face both systematic and unsystematic risks, of which Alpha represents the unsystematic risk. It is calculated as follows:

$$\alpha = P_r - r_f - \beta(B_r - r_f), \quad (5)$$

where P_r denotes the annualized return of the constructed strategy, r_f denotes the market risk-free return, and B_r denotes the benchmark annualized return.

TABLE 12: Portfolio selection by investment strategy under different cycles.

Position change date	Stock 1	Stock 2	Stock 3
Jan. 31	Kaiying Network	Perfect World	Chinese Media
Feb. 27	SuperMap Software	Zhongke Chuangda	Mason Culture
Mar.31	Songcheng Performing Arts	Skyworth Digital	Tianyin Holdings
Apr. 30	Triple 7 Entertainment	Xinwanda	Crystal Photonics
May 31	Rongtai Health	Jebsen shares	Perfect World
June 30	Common Progress Shares	Anjie Technology	Chinese Media
July 31	SuperMap Software	Zhongke Chuangda	Shunnet Technology
Aug. 31	Skyworth Digital	Shunnet Technology	Anjie Technology
Sept. 30	Skyworth Digital	Xinwanda	Goer Corporation
Oct. 31	SuperMap Software	Chinese Media	Triple 7 Entertainment
Nov. 30	Rongtai Health	Perfect World	Mason Culture
Dec. 31	Triple 7 Entertainment	Jebsen shares	Rongtai Health

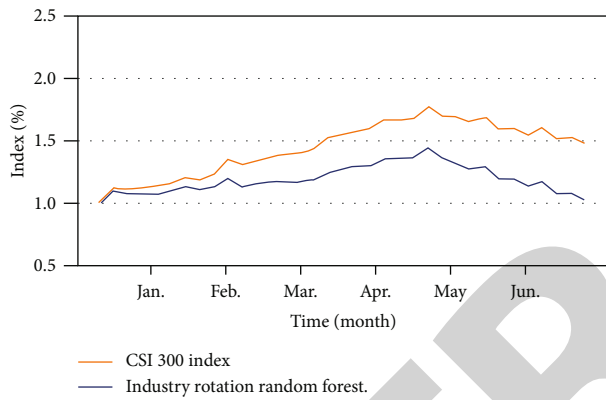


FIGURE 4: Random forest investment strategy versus market benchmark.

4.1.3. *Earning Volatility*. This ratio is used to measure the riskiness of the measured output, and the higher the fluctuation value, the higher the risk strategy, as follows:

$$\sigma_p = \sqrt{\frac{250}{n-1} \sum_{t=1}^n (P_t - \bar{P}_t)^2}, \bar{P}_t = \frac{1}{n} \sum_{t=1}^n P_t, \quad (6)$$

where P_t denotes the annualized strategy return, r_f denotes the market risk-free return, and I equation denotes the strategy return volatility.

4.1.4. *Max. Retracement Rate*. This metric is used to represent the worst-case scenario that can occur in building a strategy, as follows:

$$\text{MaxDrawDown}_t = \max \left(1 - \frac{P_j}{P_i} \right), \quad (7)$$

where MaxDrawDown_t denotes the maximum retracement at day t , P_j and P_i denote the total assets of the strategy at days i and j , and $t > j > i$. Also, the lower the maximum retracement, the better the strategy.

4.2. *Strategy Performance Evaluation Indicators*. For a portfolio of investment strategies, in addition to its riskiness, which needs to be tested and effectively avoided, its long-term results, or performance, also need to be tested and evaluated, mainly in the following areas ([7, 8]).

4.2.1. *Annualized Expected Rate of Return*. This indicator refers to the expected rate of return for the effective quantitative investment period of the selected investment strategy within one year, as follows:

$$P_r = \left(\frac{P_{\text{end}}}{P_{\text{start}}} \right)^{(250/n)} - 1, \quad (8)$$

where P_{end} denotes the final total assets held by the investor, P_{start} denotes the initial total assets invested by the investor, and n is the number of trading days.

4.2.2. *Benchmark Annualized Rate of Return*. The role of this indicator is only to provide a reference value for the broad market return after the model is solved for the corresponding test, as follows:

$$B_r = \left(\frac{B_{\text{end}}}{B_{\text{start}}} \right)^{(250/n)} - 1, \quad (9)$$

where B_{end} denotes the final trade of the reference base, B_{start} denotes the initial trade value of the reference base, and n is the number of trading days.

4.2.3. *Sharpe Ratio*. The Sharpe ratio reflects the additional return that a unit of risk can bring when an investor invests using a portfolio of investment strategies, while the strategy is taking on risk. Therefore, the interrelationship between the risk-expected return of the strategy and the risk-expected return in the actual situation can be simultaneously analyzed in a comprehensive manner, as follows:

$$\text{SharpRatio} = \frac{P_r - r_f}{\sigma_p}, \quad (10)$$

TABLE 13: Performance evaluation of random forest stock selection strategies.

Item	Annualized yield	Annualized volatility	Sharpe ratio	Maximum retracement	Information ratio
Random forest based on industry rotation	28.13%	21.19%	0.91	-38.04%	0.53
CSI 300 index	17.01%	26.11%	0.62	-41.03%	

where P_r is the annualized rate of return generated by the model strategy, r_f is the market risk-free rate of return, and the numerator is the return volatility of the model strategy.

4.3. Reasonableness Test of the Strategy Plan. Based on the conclusions obtained in the above study of the sector index rotation strategy, while its corresponding optimization parameters ($A = 60$, $B = 3$) were determined by the random forest algorithm, the resulting optimization parameters need to be back-tested and analyzed before determining whether the investment strategy portfolio is feasible. In the models with different back-testing time periods, the factor data of the past six months in the same year are selected to make basic forecasts of the sample returns, while obtaining the corresponding probability magnitudes. After that, an equal number of stocks are additionally selected from the stock pool and sorted by return size, while the equal-weighted allocation is processed to construct the back-test portfolio [2]. In addition, considering the research object, the CSI 300 index industry sectors are selected for benchmark comparison, and the results are shown in Figure 4.

Based on the results of the model test in Figure 4, the analysis was performed using the above selected indicators, and the results are shown in Table 13.

Table 13 shows the performance of the portfolio of stock selection strategies constructed according to the random forest model compared with the CSI 300 index. According to the above table, the maximum retracement rate of the stock selection strategy portfolio under the random forest model is -38.04%, which is slightly lower than the -41.03% of the CSI 300 index, so it can be seen that the investment strategy constructed by the random forest model is more effective for the CSI 300 index. In addition, the portfolio of the random forest model has a Sharpe ratio of 0.91 and an annualized return of 28.13%, and its annualized volatility is only 21.19%, which is lower than the 26.11% of the CSI 300 index, so the portfolio return of the stock selection strategy has a more obvious improvement compared with the general portfolio and is also more stable.

5. Research Conclusion

The risk analysis of the investment strategy portfolio built by the random forest algorithm and the corresponding performance evaluation above show that from the perspective of the maximum retracement rate, the maximum retracement rate of the benchmark stock market is -41.03%, while the maximum retracement rate of the investment strategy portfolio built by the random forest based on sector rotation is

-38.04%, which is higher than the value of the benchmark market ratio, indicating that the strategy portfolio was constructed under the conditions, the probability of a possible worst-case scenario is smaller than that of the base market, and thus, it can be shown that the strategy portfolio effectively hedges a certain amount of investment risk [21].

In addition, from the perspective of performance evaluation [43], the annualized return of the random forest stock picking strategy model based on sector rotation is much higher than that of the benchmark index, and its annualized volatility is significantly lower than that of the benchmark market; in addition, the Sharpe ratio of the random forest stock picking strategy model is 0.91, while that of the benchmark market is 0.62, which is a significant difference between the two. It can be seen that the improved random forest stock picking strategy model under the sector rotation theory has improved the performance and investment utility for the basic market [44]; therefore, the investment strategy portfolio has strong investment feasibility and investment value.

In addition to this, the random forest algorithm plays more of a computational role in the study of the entire process of building a random forest stock selection model. Unlike the traditional overfactor quantitative stock selection and single-factor stock selection, this method is based on the sector rotation theory, which takes more into account the sector index connection between different stocks and classifies the macroeconomic period of the overall market, and uses machine algorithms to process the multifactor stock selection configuration to enhance the investability and efficiency of the overall investment strategy portfolio.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Authors' Contributions

Jia-Ming Zhu was responsible for the methodology, conceptualization, supervision, and leadership. Yu-Gan Geng was responsible for the conceptualization, visualization, software, validation, and data analysis. Wen-Bo Li was responsible for writing the manuscript, verification, and investigation. Xia Li was involved in data collation verification and method

design. Qi-Zhi He contributed to the study conception and design, supervision, review, and editing. All authors read and approved the final manuscript.

Acknowledgments

This study was funded by the Teaching and Research Fund Project of the Anhui University of Finance and Economics (acxkjs2021005, acyljc2021002, and acjyyb2020011), Humanities and Social Science Fund of Ministry of Education of China (20YJA790021), General Project of Anhui Natural Science Foundation (1908085MG232), National Innovation and Entrepreneurship Training Program for Undergraduates of China (202110378006), Anhui University Major Project of Humanities and Social Science Research (SK2020ZD006), and Major Project of Philosophy and Social Science Planning of Zhejiang Province (22YJRC07ZD).

References

- [1] J.-M. Zhu, W.-Y. Xia, J.-J. Sun, J. B. Liu, and F. H. Yu, "The spread pattern on Ebola and the control schemes," *International Journal of Innovative Computing and Applications*, vol. 9, no. 2, pp. 77–89, 2018.
- [2] J. T. Guan, *Research on random forest multifactorial stock selection strategy based on industry rotation*, Shanghai Normal University, 2020.
- [3] Q. Z. He, H. Tong, and J. B. Liu, "How does inequality affect the residents' subjective well-being: inequality of opportunity and inequality of effort," *Frontiers in Psychology*, vol. 13, article 843854, 2022.
- [4] P. Andrikopoulos, J. Clunie, and A. Siganos, "Short-selling constraints and 'quantitative' investment strategies," *The European Journal of Finance*, vol. 19, no. 1, pp. 19–35, 2013.
- [5] S. K. Shu and L. Li, "Multifactorial quantitative stock selection strategy, computer engineering and application," vol. 57, no. 1, pp. 110–117, 2021.
- [6] X. N. Han, Q. T. Wang, and X. Y. Zhu, "Research on the development of China's quantitative investment strategy and its countermeasures," in *Institute of Management Science and Industrial Engineering, Proceedings of 2019 International Conference On Humanities, Management Engineering and Education Technology*, pp. 45–52, Qingdao, Shandong, China, 2019.
- [7] L. Wang, "Stock selection strategy based on deep forest," *Economic Research Guide*, vol. 27, no. 413, pp. 78–79, 2020.
- [8] Y. Z. Wang, *Empirical study of multifactorial quantitative stock selection based on the Boosting algorithm*, Shandong University, 2020.
- [9] Y. B. Zhu, *Design of a multifactorial stock selection scheme based on the XGBoost and Light GBM algorithm*, Nanjing University, 2020.
- [10] S.-M. Zhang, W.-L. Zhan, H. Hu, Y.-S. Liu, and J.-M. Zhu, "Research on ethanol coupling to prepare C₄ olefins based on BP neural network and cluster analysis," *Journal of Chemistry*, vol. 2022, Article ID 5324336, 10 pages, 2022.
- [11] X. Zhang, Z. Ding, J. Hang, and Q. He, "How do stock price indices absorb the COVID-19 pandemic shocks?," *The North American Journal of Economics and Finance*, vol. 60, no. 4, article 101672, 2022.
- [12] M. Z. Ouyang, "Multi-factorial quantitative stock selection strategy based on GRU neural network," *Zhongnan University of Economics and Law*, vol. 2020, 2020.
- [13] Y. H. Zhao and S. Q. Fan, "Stock forecasting analysis based on deep learning and quantitative investment algorithms with multiple indicators," in *Wuhan Zhicheng Times Cultural Development Co. Proceedings of the 3rd International Conference on Computer Engineering, Information Science & Application Technology (ICCIA 2019)*, pp. 573–576, Chongqing, China, 2019.
- [14] Z. N. Luo, "Stacking quantitative stock selection strategy research based on integrated tree model," *China Prices*, vol. 2021, no. 2, 2021.
- [15] L. Li, S. Z. Xu, Y. H. Liu, and X. R. Yang, "Predicting the rise and fall of stock prices based on the modified BP-Ada Boost," *Journal of Physics: Conference Series*, vol. 1518, article 012060, 2020.
- [16] L. F. Gruber and M. West, "Bayesian online variable selection and scalable multivariate volatility forecasting in simultaneous graphical dynamic linear models," *Econometrics and Statistics*, vol. 3, pp. 3–22, 2017.
- [17] L. Cáceres, J. I. Merino, and N. Díaz Díaz, "A computational intelligence approach to predict energy demand using random forest in a Cloudera cluster," *Applied Sciences*, vol. 11, no. 18, pp. 8635–8635, 2021.
- [18] Y. Lin and R. D. Ye, "Can machine learning unlock the continuous alpha? Empirical study based on China A-share market," *Open Journal of Business and Management*, vol. 9, no. s5, pp. 2358–2369, 2021.
- [19] C. Kathuria, D. Mehrotra, and N. K. Misra, "A novel random forest approach to predict phase transition," *International Journal of System Assurance Engineering and Management*, vol. 13, no. 1, pp. 494–503, 2022.
- [20] Y. Qi and X. L. Li, "Research on fund stock selection strategy based on the perspective of industry concentration," *Finance and Finance*, vol. 2019, no. 1, pp. 17–21, 2019.
- [21] Y. F. Wen, R. X. Wu, Z. H. Zhou et al., "A data-driven method of traffic emissions mapping with land use random forest models," *Applied Energy*, vol. 305, article 117916, 2022.
- [22] Z. C. Guo, T. Wang, and S. L. Liu, "Comparison of the back-propagation network and the random forest algorithm based on sampling distribution effects consideration for estimating nonphotosynthetic vegetation cover," *International Journal of Applied Earth Observation and Geoinformation*, vol. 104, article 102573, 2021.
- [23] C. C. Chen, P. Zhang, Y. Liu, and J. Liu, "Financial quantitative investment using convolutional neural network and deep learning technology," *Neurocomputing*, vol. 390, no. 5, pp. 384–390, 2020.
- [24] M. Q. Chen, Z. Zhang, J. Shen, Z. Deng, J. He, and S. Huang, "A quantitative investment model based on random forest and sentiment analysis," *Journal of Physics: Conference Series*, vol. 1575, no. 1, pp. 128–133, 2020.
- [25] Y. Q. Mo, *An empirical study of random forest stock selection strategies*, Capital University of Economics and Trade, 2019.
- [26] Y. J. Zhang, G. Chu, and D. H. Shen, "The role of investor attention in predicting stock prices: The long short-term memory networks perspective," *Finance Research Letters*, vol. 38, no. 2, article 101484, 2021.
- [27] C. Alakuş, D. Larocque, S. Jacquemont et al., "Conditional canonical correlation estimation based on covariates with

- random forests,” *Bioinformatics*, vol. 37, no. 17, pp. 2714–2721, 2021.
- [28] S. D. Astuti, M. H. Tamimi, A. A. S. Pradhana et al., “Gas sensor array to classify the chicken meat with *E. coli* contaminant by using random forest and support vector machine,” *Biosensors and Bioelectronics: X*, vol. 9, no. 12, p. 100083, 2021.
- [29] S. Angadi and V. S. Reddy, “Multimodal sentiment analysis using reliefF feature selection and random forest classifier,” *International Journal of Computers and Applications*, vol. 43, no. 9, pp. 931–939, 2021.
- [30] J. W. Bai, Y. M. Li, J. W. Li et al., “Multinomial random forests: fill the gap between theoretical consistency and empirical soundness,” *Pattern Recognition*, vol. 1903, article 04003, 2019.
- [31] C. Q. Li, Y. W. Lin, and Q. S. Xu, “An enhanced random forest with canonical partial least squares for classification,” *Communications in Statistics-Theory and Methods*, vol. 50, no. 18, pp. 4324–4334, 2021.
- [32] F. Xu, L. Y. Mo, H. Chen, and J. M. Zhu, “Genetic algorithm to optimize the design of high temperature protective clothing based on BP neural network,” *Frontiers in Physics*, vol. 39, 2021.
- [33] F. Motamedi, H. Pérez Sánchez, A. Mehridehnavi, and etc, “Accelerating big data analysis through LASSO-random forest algorithm in QSAR studies,” *Bioinformatics*, vol. 38, no. 2, pp. 469–475, 2022.
- [34] X.-W. Cai, Y.-Q. Bao, and M.-F. Hu, “Simulation and prediction of fungal community evolution based on RBF neural network,” *Computational and Mathematical Methods in Medicine*, vol. 2021, Article ID 7918192, 13 pages, 2021.
- [35] H. Y. Liu and A. Q. Xie, “Development of quantitative investment model based on selecting stocks with AI,” in *Institute of Management Science and Industrial Engineering. Proceedings of 2019 8th International Conference on Advanced Materials and Computer Science (ICAMCS 2019)*, pp. 205–209, Chongqing, China, 2019.
- [36] Q. Z. He, P. T. Xia, B. Li, and J. B. Liu, “Evaluating investors’ recognition abilities for risk and profit in online loan markets using nonlinear models and financial big data,” *Journal of Function Spaces*, vol. 2021, Article ID 5178970, 15 pages, 2021.
- [37] J. W. He, Z. Y. Wei, and X. B. Zhu, “Research on the application of machine learning in quantitative investment,” in *2020 2nd International Conference on Economic Development and Management Science*, pp. 73–79, Dalian, China, 2020.
- [38] L. Wang, *Multifactorial quantitative stock selection strategy based on deep forest*, Shanghai University of Engineering, 2019.
- [39] J. Y. Tang, A. Henderson, and P. Gardne, “Exploring AdaBoost and random forests machine learning approaches for infrared pathology on unbalanced data sets,” *The Analyst*, vol. 146, no. 19, pp. 5880–5891, 2021.
- [40] Y. Jiang, J. L. Gradus, T. L. Lash, and M. P. Fox, “Addressing measurement error in random forests using quantitative bias analysis,” *American Journal of Epidemiology*, vol. 190, no. 9, pp. 1830–1840, 2021.
- [41] L. Stephanie, L. Genevieve, and S. Tibor, “286Collider-stratification bias when estimating variable importance using random forests,” *International Journal of Epidemiology*, vol. 50, Supplement 1, 2021.
- [42] J. M. Nguyen, P. Jézéquel, P. Gillois et al., “Random forest of perfect trees: concept, performance, applications and perspectives,” *Bioinformatics*, vol. 37, no. 15, pp. 2165–2174, 2021.
- [43] S. Z. Chen, R. Yang, and M. Y. Zhong, “Graph-based semi-supervised random forest for rotating machinery gearbox fault diagnosis,” *Control Engineering Practice*, vol. 117, article 104952, 2021.
- [44] S. Georgios, V. V. Robert, M. A. E. J. Daisy, and P. John, “Advanced data fusion: random forest proximities and pseudo-sample principle towards increased prediction accuracy and variable interpretation,” *Analytica Chimica Acta*, vol. 1183, article 339001, 2021.