

Retraction

Retracted: A Study of Different Existing Methods for the Stock Selection in the Field of Quantitative Investment

Wireless Communications and Mobile Computing

Received 25 July 2023; Accepted 25 July 2023; Published 26 July 2023

Copyright © 2023 Wireless Communications and Mobile Computing. 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This article has been retracted by Hindawi following an investigation undertaken by the publisher [1].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] P. Li and J. Xu, "A Study of Different Existing Methods for the Stock Selection in the Field of Quantitative Investment," *Wireless Communications and Mobile Computing*, vol. 2022, Article ID 2695099, 17 pages, 2022.

Review Article

A Study of Different Existing Methods for the Stock Selection in the Field of Quantitative Investment

Pengfei Li  and Jungang Xu 

School of Computer Science & Technology, University of Chinese Academy of Sciences, Beijing, China

Correspondence should be addressed to Pengfei Li; lipengfei@ucas.ac.cn

Received 22 February 2022; Accepted 8 March 2022; Published 19 March 2022

Academic Editor: Mohammad Farukh Hashmi

Copyright © 2022 Pengfei Li and Jungang Xu. 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.

Due to the computer's powerful data computing ability, more and more traditional investors pay attention to computer-based quantitative investment. Moreover, computer-based quantitative investment has gradually replaced some traditional investment methods which rely on subjective judgments. As a typical quantitative trading strategy, stock selection has attracted a lot of attention. And many researchers have put forward various methods and published a large number of research results. In particular, with the development of artificial intelligence technology, an increasing number of researchers try to apply different machine learning and deep learning methods to this field to obtain more stable and efficient stock selection models. Even though there is a growing interest in developing methods for stock selection, there is a lack of review papers that are solely focused on different types of methods for stock selection. Hence, our motivation in this paper is to provide a comprehensive literature review on different types of methods for stock selection in the field of quantitative investment. Firstly, we introduce the basic concept of stock selection. Secondly, according to the classification of traditional methods and machine learning methods, we introduce the widely used stock selection methods in detail. Then, we give a statistical analysis about the relevant literatures in this field. Finally, the stock selection methods are summarized. The main contribution of this paper is we analyse various quantitative analysis methods from the perspective of stock selection for the first time. And it has some guidance for researchers who engaged in quantitative trading or interested in quantitative investment, and they will benefit from it.

1. Introduction

The application of computer technology in the investment market has attracted more and more attention. How to use various artificial intelligence methods to improve the efficiency and stability of returns in the investment market has become a very hot research field. In particular, with the continuous development of artificial intelligence technology, quantitative investment has attracted more and more attention and research. Quantitative investment is to transform human transaction logic into transaction models, test the transaction models' effectiveness with the help of data, and then use them to guide the transaction in real world. Quantitative investment relies on statistical and measurement methods to establish appropriate investment strategies and obtain investment returns through computer automation

(semiautomatic) trading, which is the product of statistics, computer, and finance.

Since the 1990s, some quantitative investment researchers have expressed the complex transaction logic through mathematical logic formula and analysed the investment market through mathematical methods such as statistics and probability. They hope they can find laws and form stable and reusable investment strategies that can be used in actual transactions. In particular, with the development of artificial intelligence and the successful application in the fields of text, speech, and image in recent years, more and more researchers expect to use artificial intelligence methods to explain some complex market phenomena, so as to build more stable and effective quantitative strategies. When we review the stock selection literatures, we found that the stock selection methods have been used for a long

time, and they are changing with the development of the environment. We think the main reason is the continuous development and application of different technologies. We have consulted a large number of literatures related to quantitative investment and found that they contain many studies on stock selection methods. We also find that there are a lot of research reviews on financial time series, but there is no detailed review on stock selection methods. Therefore, this paper is the first review of the stock selection methods. This paper is aimed at providing a study of the current stock selection methods in the view of academic and industry. Moreover, it is expected guide the researchers in actual scientific research and application according to their actual needs. Our basic motivation in this paper is to find answers to the following research questions:

- (i) What models can be used to stock selection?
- (ii) Can traditional stock selection methods be used in the current stock market?
- (iii) Compared with traditional stock selection methods, what is the effect of machine learning methods?

In this paper, we focus on the study of stock selection methods. We do not make the specific distinction for the specific application market and do not conduct other quantitative trading strategies such as quantitative timing that is occasionally mentioned in many literatures. Simultaneously, we mainly provide the literature research on stock selection methods, and other literatures on economics, psychology, and sociology will not be included in this paper.

We surveyed and reviewed not only journals and conferences but also master's and doctoral dissertations, book chapters, arXiv papers, and some online literature. Both English studies and Chinese studies are within the scope of our research.

In this study, we focus on the specific methods, without considering the countries, markets, and groups of stock selection methods in the literature. There is also much literature on the current situation of stock selection methods, but only for the methods proposed in the literature itself; there is lack of review of multiple methods. Therefore, we hope that through the introduction of various stock selection methods, researchers and practitioners can get a more intuitive comparative reference, to better understand how to develop their own stock selection strategy according to specific needs.

The rest of the paper is structured as follows. Section 2 briefly introduces the basic concepts of stock selection and classifies the stock selection methods from the perspective of dataset. In Section 3, we focus on the stock selection methods involved in the literature from the perspectives of traditional methods and machine learning methods. In Section 4, we make a statistical analysis of all the literature. Finally, we analyse and summarize the development trend of stock selection methods.

2. Stock Selection

Stock selection is a typical quantitative investment strategy, and the key of stock selection is to dig out the driving factors

behind the stock price and analyse the internal links between these factors and stock price. Stock selection is different from stock timing, which is another quantitative investment strategy. It does not make any predictions on market trend, but obtains the trend of stocks relative to the market through analysis of market trend and the own factors of the stock and meets the excess return. The stock selection method is very complex, and a good stock selection method should consider its profitability and risk control ability to ensure the effectiveness.

The author of [1] defined stock selection as "the process of using data-based methods to establish a model to select stocks with better performance in the future and obtain returns." According to the data sources of the stock selection strategy, it can be divided into two categories: Fundamental Stock Selection and Market Stock Selection. Each class has several methods, as shown in Figure 1.

Stock selection based on fundamental analysis is a set of stock selection methods that use the existing public information to determine buy or sell stock by analysing whether the stock value and stock price are reasonable. The data involved in fundamental analysis mainly include macroeconomic data, microeconomic data, and financial data. Stock selection methods based on fundamental analysis mainly include multifactor model, style rotation model, and industry rotation model. Stock selection based on market analysis is also called stock selection based on price data or stock selection based on technical analysis. It mainly analyses all price-related data in the stock market. This kind of data is generally in the trading instruction book of the stock market. Compared with the fundamental data, the price data required has the characteristics of faster update frequency and larger data volume. Due to the strong time series characteristics of stock data, the data noise and uncertainty are large, and the repeatability is random, which makes the stock selection more difficult. Stock selection methods based on market analysis mainly include capital flow model, trend tracking model, consistent expectation model, momentum reversal model, and chip stock selection model.

Stock selection is usually multifaceted. If we only rely on fundamental data or market data to select stocks, there may be relatively large problems, which will result in the return rate of the selected stocks not reaching the expectation. Therefore, they are usually mixed in the actual application. Because this paper mainly studies stock selection methods, we do not focus on distinguishing between Fundamental Stock Selection and Market Stock Selection.

3. Stock Selection Methods

The traditional qualitative investment and quantitative investment are essentially based on the market efficient hypothesis or noneffective theory. It can build excess return portfolio by analysing the underlying factors that affect stock prices. Quantitative investment relies more on computer analysis instead of traditional subjective analysis. Quantitative strategies are formulated by the analysis of massive historical data or market experience. These strategies can avoid emotional interference and avoid

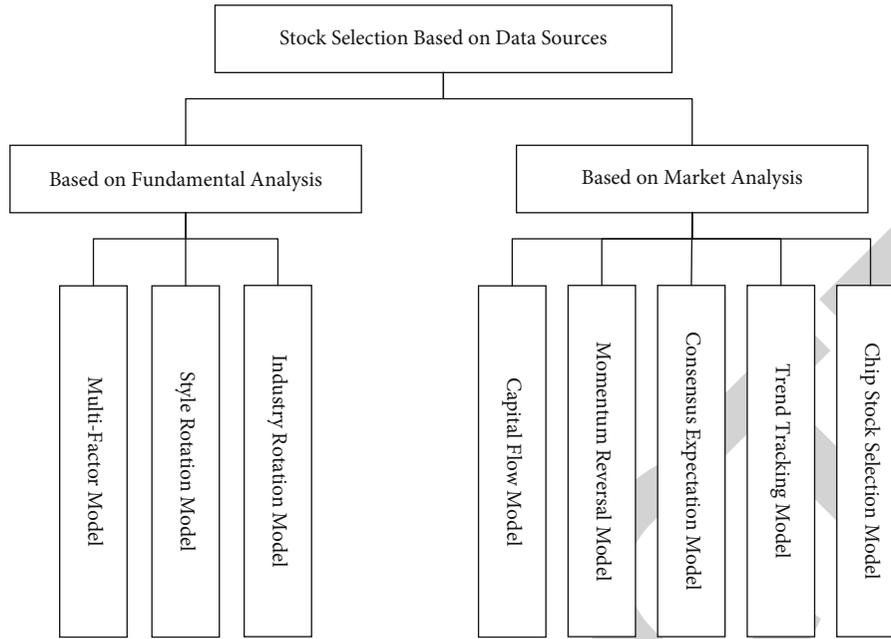


FIGURE 1: Stock selection based on different data sources.

repeating previous wrong decisions in the face of the same market conditions.

For the past 30 years, stock selection has been a hot topic in stock market investment research. With the successful application of deep learning model in various fields, stock selection has a new idea, and correspondingly, there is much new related research. With the increasing amount of stock-related data and the increasing dimensions, the traditional stock selection methods have been greatly limited in terms of performance and efficiency. Machine learning methods have inherent advantages in dealing with such multidimensional and large-scale data. Our survey found that stock selection methods based on machine learning have attracted more and more attention and application.

We find that the existing stock selection methods mainly include analytic hierarchy process (AHP), fuzzy analysis, data envelopment analysis (DEA), genetic algorithm, clustering, support vector machine (SVM), neural network (NN), random forest, ensemble learning, convolutional neural network (CNN), recurrent neural network (RNN), long short-term memory (LSTM), and reinforcement learning (RL). We divide stock selection methods into traditional methods and artificial intelligence methods. And the artificial intelligence methods are divided into machine learning and deep learning. We try our best to introduce the methods how and why to be used instead of the implementation details.

3.1. Traditional Methods

3.1.1. Multiattribute Decision-Making (MADM). Multiattribute decision-making, also known as finite-scheme multiobjective decision-making, refers to selecting the optimal

alternatives or ranking the alternatives considering multiple attributes. It is an important part of modern decision-making science.

Analytic hierarchy process (AHP) is one of the multiattribute decision-making methods formally proposed by American operational research scientist Saaty [2] in 1971. It is a systematic and hierarchical analytical method combining qualitative and quantitative analysis. The structure of the AHP is simple by combining the qualitative method with the quantitative method. It needs less quantitative data. It is very effective when quantitative data is missing and qualitative data is rich. However, the characteristics of AHP also lead to its own defects, which lead to the lack of convincing decision-making results. When there are too many indicators, the amount of data to be counted is very large. The weight is difficult to determine, and the eigenvalue and eigenvector solution is more complex. Table 1 tabulates the stock selection papers using AHP.

The authors of [3] applied the AHP to establish hierarchical structure and judgment matrix, which was combined with qualitative evaluation and quantitative evaluation, evaluated and ranked the stocks in investment stock projects, and then made reasonable selection decisions. Based on [3], the authors of [4] added a fuzzy complementary judgment matrix. Through the consistency test of the fuzzy judgment matrix, the shortcomings of an analytic hierarchy process model were overcome. Moreover, the decision-making could reflect the consistency of people's thinking and judgment and improve its reliability and effectiveness.

There are many other MADM methods, such as analytic network process (ANP), Elimination et Choix Traduisant l'aréalité (ELECTRE), multiattribute value theory (MAVT), data envelopment analysis (DEA), technique for order

TABLE 1: Stock selection using AHP.

Art.	Methods	Dataset	Period	Country
[3]	Analytical hierarchy process	Preference, inside information, risk, national economy, economic policy, emergency, stock price, company status, stock trend	One week	China
[4]	Fuzzy analytical hierarchy process	Preference, inside information, risk, national economy, economic policy, emergency, stock price, company status, stock trend	One week	China

TABLE 2: Stock selection using other MADM methods.

Art.	Methods	Dataset	Period	Country
[5]	VIPOR-DANP	Glamor stocks in semiconductor industry	2009.05–2012.12	Taiwan, China
[6]	Lehmann multiple decision	Price, returns	2014	Russia
[7]	Gordon model	Taiwan market index	2006	Taiwan, China
[8]	TOPSIS-DANP	Glamor stocks in semiconductor industry	2007	Taiwan, China
[9]	Grey relational analysis	Nine different companies conducting activities in ten lines of business	2011–2012	Turkey
[10]	DEA cross-efficiency evaluation	Korea market index	2002–2011	Korea
[11]	OWA, DEA	Tokyo stock exchange	2011	Japan
[12]	Gangless cross-evaluation in DEA	Tehran stock market	2011	Iran
[13]	PIF-TOPSIS	Seven companies	2017	India

preference by similarity to an ideal solution (TOPSIS), and grey relation analysis (GRA). These methods have also been used in quantitative stock selection, as shown in Table 2,

The authors of [5] proposed a glamor stock selection model based on VIKOR-DANP, which reduced the limitation of the traditional regression model. One stock with better return than the other four stocks was selected from the five given stocks. In [6], the stock selection problem was studied as a multidecision problem. The authors thought that the quality of the multidecision process was measured by the conditional risk. Therefore, based on the Lehmann multiple decision theory, the optimal multidecision statistical process of stock selection was constructed. The results showed that it could be used for return, volume, and risk. The authors of [7] studied the influencing factors and relative weights of dividend, discount rate, and dividend growth rate based on Gordon model and multicriteria decision-making method. And with the help of expert judgment, they combined the theoretical structure with the expert opinion, which provided theoretical and practical support for stock selection. The authors of [8] proposed a TOPSIS-TANP model by improving the limitations of the original G-Score model and combining the standard weight with expert evaluation. The method was based on a financial model, and the results of the experiment showed that it could get high returns by ranking the selected stocks. The authors of [9] used the GRA method to conduct stock selection for 9 different companies in 10 industries of Borsa Istanbul. The model was built and tested through the beta coefficient, returns, standard deviation, and variation coefficient. Finally, the

relationship between these coefficients and stock price from the perspective of reporting risk was correlated, thus proving the model's validity. DEA was used in [10–12]. The DEA model was used to determine the efficiency score of stock companies and then selected the stocks with high efficiency. The authors of [13] proposed probabilistic intuitionistic fuzzy TOPSIS (PIF-TOPSIS) method to transform fuzzy sets into intuitionistic fuzzy sets and used probabilistic intuitionistic fuzzy elements to evaluate attributes. The actual data was used to rank the enterprises, which proved that the PIF-TOPSIS method outperformed the other multicriteria decision-making models.

3.1.2. Principal Component Analysis (PCA). PCA is a statistical analysis method that reduces multiple variables to a few principal components by dimension reduction technology. These principal components can reflect most of the original variables, usually expressed as some linear combinations of the original variables. In short, PCA is to try to combine many of the original indexes with certain relevance and present new and independent comprehensive indicators to replace the original ones.

The PCA method can eliminate the correlation between the evaluation indexes and reduce the index selection workload. Moreover, the main components in PCA are arranged in order of variance, which allows the components with large variance be used and some components with small variance be discarded. However, the explanation of the principal component in PCA is relatively vague, not as clear and exact as the original variable, which will pay the corresponding

cost in the dimension reduction. Moreover, when there are positive and negative factors, the significance of the comprehensive evaluation function is not clear. PCA is mainly used in combination with other stock selection methods. Table 3 tabulates the stock selection papers using PCA.

The authors of [14–17] used PCA to process the original data to extract low-dimensional and efficient feature information. And then, they used different stock selection methods such as GA and SVM to select stocks. The authors of [18] used EKPCA to map, process, and identify the data of CSI 300 constituent stocks in high-dimensional space and then extracted the factor characteristics and determined the required model factors. They built a multifactor stock selection model by converting the sample data by eigenvector and kernel function and stock return regression. In [18], a regression equation was established for the independent variables using the efficient kernel principal component analysis (EKPCA) algorithm to predict the returns. 50 influencing factors including fundamentals, technical indicators, and investor sentiment indicators were selected. The basic mode was determined using the EKPCA algorithm, and the efficient kernel principal component was extracted in the high-dimensional feature space. Compared with the classical KPCA algorithm, this algorithm had a higher characteristic collection efficiency, and the selected stock portfolio had a market benchmark level.

3.1.3. Mean Regression. Mean regression means that the stock price always fluctuates around its value centre in the long run. Specifically, the stock price will fluctuate to its value centre, whether it is higher than or lower than its value centre in the short term. Table 4 tabulates the stock selection using mean regression.

The authors of [20–22] used mean regression to carry out stock selection. In [23], quantitative regression method was used for stock selection. The authors of [24] used the generalized auto regressive conditional heteroskedasticity (GARCH) method to focus on emerging markets and build better stock selection model.

3.1.4. Fuzzy Method. Fuzzy method is a new control method based on fuzzy mathematics theory. Fuzzy method transforms the qualitative evaluation into quantitative evaluation, that is, using fuzzy mathematics to make an overall evaluation of things or objects restricted by many factors. It has the characteristics of clear results and strong systematicness. It can better solve the fuzzy and difficult to quantify problems and is suitable for the solution of various uncertain problems. It is usually used in stock selection by combining other methods. The related literature of fuzzy method in stock selection is shown in Table 5.

The authors of [25–28] used the fuzzy method or fuzzy theory to carry out stock selection. The results of the method confirmed the effectiveness of the fuzzy method in stock selection.

3.1.5. Candlestick. Candlestick is a traditional method in stock field. The related literature of candlestick in stock selection is shown in Table 6.

TABLE 3: Stock selection using PCA.

Art.	Methods	Dataset	Period	Country
[14]	FCA, GA	SSECI	1997–2006	China
[15]	PCA-SVM	Shanghai market	2009–2010	China
[16]	PCA	ASX200	2000–2014	Australia
[17]	PCA	Ghana stock exchange	2004–2015	Ghana
[18]	KPCA	China stock market	2017	China
[19]	EKPCA	Shanghai market	2013–2018	China

3.1.6. Statistical Method. We find that there are many other statistical methods used in stock selection. The related literature is shown in Table 7.

In [31], rule induction was used to select stocks. The authors of [32–34, 36] used traditional statistical methods to select stocks. The authors of [35] employed the naïve Bayes method to classify stocks and realize stock selection. And the authors of [37] proposed a complete architecture of an intraday trading management system based on time series data pattern mining algorithm. It realized stock selection and stock portfolio construction. The authors of [38] studied the quarterly shareholder data of listed energy companies in China, constructed the common shareholder network of China’s energy listed companies, and made the stock selection based on this network.

3.2. Machine Learning

3.2.1. Evolutionary Algorithm. Evolutionary algorithm, also known as evolutionary algorithms (EAS), is not a specific algorithm, but an “algorithm cluster.” The evolutionary algorithm’s inspiration is inspired by the evolutionary operation of nature, which generally includes basic operations such as gene coding, population initialization, crossover mutation operator, and operation and reservation mechanism. Compared with traditional optimization algorithms such as calculus and exhaustive methods, evolutionary computation is a mature global optimization method with high robustness and wide applicability. It has the characteristics of self-organization, self-adaptive, and self-learning. It can effectively deal with complex problems that traditional optimization algorithms are difficult to solve without limiting the nature of the problem. Evolutionary algorithms include genetic algorithm, genetic programming, evolutionary strategy, and evolutionary programming. These methods are mainly combined with other stock selection methods to optimize the parameters of other stock selection methods and improve stock selection efficiency.

With the development of the evolutionary algorithm, there are many mutation algorithms, such as ant colony algorithm and population algorithm. We found that a large number of literature on the stock selection achieved very good results, as shown in Table 8.

The authors of [39] provided a stock selection model using genetic programming. The decision variables used in the model are financial factors determined by experts. It proved the advantages of genetic programming in the existing complex portfolio management. And it could develop a

TABLE 4: Stock selection using mean regression.

Art.	Methods	Dataset	Period	Country
[20]	Mean regression	EMDB	1982.01–1999.05	USA
[21]	Mean regression	MSCI EM countries	2003–2016	MSCI EM countries
[22]	Mean regression	China stock market	2008–2017	China
[23]	Quantile regression	China stock market	2007–2017	China
[24]	GARCH	China stock market	1993–2001	China

TABLE 5: The literature of FM in stock selection.

Art.	Methods	Dataset	Period	Country
[25]	Fuzzy ranking and weighting algorithm	ISE	2013	Turkey
[26]	Fuzzy quantitative analysis and fuzzy multicriteria decision	Thailand stock market	2012–2016	Thailand
[27]	Fuzzy approach	NSE	2011.04–2017.03	USA
[28]	Fuzzy quantitative analysis	Thailand stock market	2012–2016	Thailand

TABLE 6: Stock selection using candlestick.

Art.	Methods	Dataset	Period	Country
[29]	Japanese candlestick charting	S&P 500	5 days	Japan
[30]	Wrapper ANFIS-ICA and Japanese candlestick	New York stock exchange	2000–2009	USA

TABLE 7: Stock selection using statistical methods.

Art.	Methods	Dataset	Period	Country
[31]	Rule induction	S&P 500	1987.10–1993.06	USA
[32]	Statistics	IDC	1984.06–1999.07	USA
[33]	Copula model	Germany stock market	2005–2009	Germany
[34]	FFC	Shanghai stock market	2002–2016	China
[35]	Naïve Bayes classifier	Shanghai stock market	2002–2004	China
[36]	Statistics	TSE	1983–1996	Japan
[37]	Intraday pattern	S&P 500	2008.01–2012.08	USA
[38]	Statistics	China stock market	2012–2018	China

multifactor model to rank the stocks in S&P 500 and select the high-yield stocks. In [40], the fuzzy model and genetic algorithm were used for stock selection. The genetic algorithm was used to optimize model's parameters, and the input variables of the stock scoring model were selected and a stock scoring mechanism based on basic variables was established. The stocks ranking at the top were the stocks that could be selected after the stocks were scored. Experimental results showed that this method's returns on investment were significantly better than the benchmark returns. The authors of [41] compared stock selection models of investor sentiment index based on regression and genetic algorithm. In the evolutionary model, genetic algorithm was used to optimize the model parameters and select input variables' features. Experimental results showed that the model based on the combination of regression and genetic algorithm was obviously better than the benchmark model and the model based on regression. A hybrid genetic support vector machine regression model was proposed in [42]. SVM was used to generate the forecast returns of a

group of stocks, and the top stocks were regarded as the components of the portfolio. Genetic algorithm was used for feature selection and model parameter optimization. It had proved that feature selection can reveal which features play a more important role in the proposed model and also showed that, in this specific application, the contribution of feature selection to effective stock selection was more significant than that of optimizing model parameters. In [43], the multiobjective genetic algorithm (MOGA) was used to optimize the model parameters and feature selection, and then, the stocks were graded, sorted, and selected to form a portfolio. On this basis, MOGA was improved to help choose stocks with financial knowledge. With the help of relevant knowledge of the investment field, the evaluation criteria were refined, making the improved MOGA method obtain the higher return portfolio. In [44], a new stock selection model with discrete and continuous variables, namely, feature selection and weight optimization model, was proposed. Based on the sigmoid DE algorithm, the stock samples of China's A-share market were studied. It was proved that

TABLE 8: Stock selection using evolutionary algorithm.

Art.	Methods	Dataset	Period	Country
[39]	Genetic programming	S&P 500, monthly financial data and stock return	1990.01–2005.12	USA
[40]	Fuzzy models and genetic algorithm	Taiwan stock market	1995–2009	Taiwan, China
[41]	Regression and GA	Taiwan stock market	2009–2010	Taiwan, China
[42]	GA and SVM	Taiwan stock market	1996–2010	Taiwan, China
[43]	Improved MOGA	Taiwan stock market	1992–2012	Taiwan, China
[44]	Sigmoid-based mixed discrete-continuous differential evolution algorithm	Shanghai stock market	2005–2012	China
[45]	Dempster-Shafer evidence theory	Bombay stock exchange	2008–2012	Bombay
[46]	Cluster analysis and AHP and TLBO	BSE	2015.02–2016.01	India
[47]	K-means and GA	China stock market	2010.05–2015.07	China

the DE algorithm based on sigmoid outperforms the benchmark in the typical linear characteristic selection model and other process optimization algorithms. The authors of [45] proposed a new portfolio construction model, which combined investors' views on stocks, early performance of stocks, and market uncertainty and adopts Dempster-Shafer evidence using ant colony algorithm to optimize the performance of the portfolio, combined with fuzzy return and risk to deal with the fuzziness of stock performance, significantly reducing the development time and cost of other existing models due to repeated expert interaction. The authors of [46] proposed a hybrid method for portfolio selection, including investor topology, clustering analysis, AHP, and teaching-learning based optimization (TLBO). TLBO was a population-based algorithm that required two common control parameters to control its size and the generation. The model sorted the assets according to the investment purpose, information source, portfolio, and other factors. The dividend selection, net profit, and portfolio were completed by TLBO, and the stock portfolio higher than the benchmark return was selected. The authors of [47] combined genetic algorithm, K -means, and multifactor stock selection to realize a set of stock selection system which can automatically optimize the combination of factors.

3.2.2. Clustering Analysis. Clustering analysis is to group data objects according to the information that describes the objects and their relationships found in the data. The aim is that objects in groups are similar (relevant) to each other, while objects in different groups are different (unrelated). The greater group's similarity, the greater the gap between groups, the better the clustering effect. Clustering analysis belongs to unsupervised learning, and its difference from the classification is that the classification required by clustering is unknown. Cluster analysis is an exploratory analysis, which is automatically classified from sample data. Different methods used in clustering analysis often lead to different

conclusions. The result of clustering analysis depends on the method and algorithm to measure distance. Therefore, different researchers do cluster analysis on the same data group, and the results are not consistent.

Cluster analysis is simple and intuitive. For exploratory research, it can provide multiple possible solutions. The final solution needs the subjective judgment and subsequent analysis of researchers. The stock selection selects the high-income stock combination among multiple stocks, which is completely in line with the clustering characteristics. We find that there are many successful researches on clustering analysis in stock selection methods, and the relevant literature is shown in Table 9.

The authors of [48] used the clustering algorithm and time-series outlier analysis, used the PAM clustering algorithm to constrain the initial set of stocks, and used outlier analysis to define two independent, active trading strategies. These results were compared with the passive strategy of fully investing in Standard & Poor's 500 large-cap stock index. As a result, the stock investment constructed by combining the clustering algorithm and time-series outlier analysis portfolio was better than pure passive index strategy. In [49], a clustering method was proposed to group the stocks in the spot market according to the risk-return criterion. The paper selected the return, risk, P/E ratio, book price ratio, sales price ratio, number of stocks sold ratio, dividend return, hierarchical clustering algorithm for clustering analysis, and the Ward method to sort, forming the final stock portfolio. The authors of [50] made cluster analysis on the constituent stocks of Shanghai Stock Exchange 180 index, comprehensively considering the profitability, capital expansion ability, asset management ability, growth type, and debt-paying ability of listed companies. Then, the component stocks of Shanghai Stock Exchange 180 index were divided into eight categories. Using conditional stock selection and cluster analysis, the success rate and return rate were tested. It was proved that the stock portfolio yield

TABLE 9: Stock selection using clustering analysis.

Art.	Methods	Dataset	Period	Country
[48]	Clustering algorithm and time series outlier analysis	S&P 500	1998.02–2001.12	USA
[49]	Hierarchical clustering algorithm and Ward methodology	Economica database	1997–2001	North and South America
[50]	K -means	Constituent stocks on China Shanghai 180	2008.06	China
[51]	Clustering algorithm	Thai stock market	2011.01–2015.09	Thai
[52]	Clustering algorithm	LQ45	100 days	Indonesia
[53]	Semisupervised K -means kernel function clustering algorithm	China stock market	2004–2017	China

selected by cluster analysis method was higher than that of stocks without cluster analysis. Reference [51] proposed a combination of long term (200-day index moving average) and short term (50-day index moving average), trend (long-term and short-term trend using the percentage difference between daily closing price and moving average time of index), and momentum (long-term and short-term momentum indicators adopt 125-day and 20-day price change rate, respectively) to identify the most likely better than the market index. The best combination of these indexes was determined by cluster analysis. The experiment proved that the method was superior to the market. Reference [52] assisted the Black Litterman portfolio by clustering given views. The method focused on calculating the Value at Risk (VaR) to illustrate the risk measurement of Black Litterman portfolio as a reference portfolio. In this paper, they used the black man clustering method to maximize portfolio selection. The experimental results verified the reliability of the method. Reference [53] improved the semisupervised K -means algorithm for the multifactor stock selection model. The modified Gaussian kernel function was added to construct the improved Gaussian kernel function based on introducing the labeled data to the gravity image of the unlabeled data. Constructing an improved Gaussian kernel function and a semisupervised K -means clustering method based on gravitational image factors greatly improved the K -means model's ability to handle high-dimensional, linearly inseparable problems without increasing the complexity of the algorithm. The excess returns of stocks selected by the improved model were significantly higher than those of the traditional model.

3.2.3. Support Vector Machine (SVM). SVM is a generalized linear classifier that classifies data according to supervised learning. Its decision boundary is the maximum margin hyperplane to solve the learning samples. The basic idea of SVM is to solve the separation hyperplane which can divide the training dataset correctly and has the largest geometric interval. For a linearly separable dataset, there are infinitely hyperplanes, but the separating hyperplane with the largest geometric interval is unique. For the case of linear inseparable, it is necessary to map the two-dimensional linear indivisible samples into the high-dimensional space so that the

sample points are linearly separable in the high-dimensional space. For linearly inseparable samples in the finite-dimensional vector space, we map them to the higher dimensional vector space and then learn the SVM by maximizing the interval.

SVM is supported by strict mathematical theories and has strong interpretability. Moreover, it does not rely on statistical methods, thus simplifying the usual classification and regression problems. The kernel method can be used to deal with nonlinear classification and regression problems. A small number of support vectors determine the final decision function, and the complexity of calculation depends on the number of support vectors rather than the dimension of sample space. However, SVM training time is generally long, especially when the number of training samples is large. Moreover, the kernel method needs to store the kernel matrix, which leads to the high space complexity of the algorithm. Therefore, SVM is more suitable for the task of the small batch of samples. Table 10 tabulates the stock selection papers using SVM.

In [54], a hybrid feature selection method based on F -score and supported sequential forward search (F -score) was proposed to find the optimal parameters of the SVM kernel function. This method combined the advantages of the filtering method and the packaging method. It selected the optimal feature subset from the original feature set and used the five-time cross-validation network search technology to find the optimal parameter value of SVM kernel function. The SVM stock selection forecast was realized through this method, and its accuracy was better than the traditional BP network. The authors of [55] used the volume weighted support vector machine (VW-SVM) to select short-term stocks and then constructed stock portfolio. Fisher method was used for feature selection, and the delay of the technical index was introduced to expand the training vector. The model was better than the traditional SVM model and had better stock selection returns in the experimental results. The authors of [56] combined PCA and BrainStorming Optimization (BSO) method to establish a hybrid V-SVR model to predict stock prices and made the stock selection. PCA was used to select the input variables of v-SVR from 20 technical indexes, and BSO was used to search the optimal parameters of v-SVR. In [57], a multifactor stock

TABLE 10: Stock selection using SVM.

Art.	Methods	Dataset	Period	Country
[54]	SVM and F_SSFS	NASDAQ index	One year	USA
[55]	VW-SVM	Poland stock market	2003.01–2013.10	Poland
[56]	V-SVR	CSI300 and SZSE	2011.04–2011.12	China
[57]	AdaBoost	China stock market	2011.01–2016.03	China
[58]	AdaBoost-SVM	CSI300	12 months	China
[59]	K-SVM	China stock market	1 month	China
[60]	KPCA-GA-SVR	China stock market	2003–2011	China
[61]	HA-SVM	Shanghai stock market	2009–2010	China
[62]	FD-SVM	China stock market	2015–2016	China
[63]	AdaBoost-SVM	CSI300	12 months	China
[64]	GBDT-SVM	China stock market	2013–2017	China
[65]	RF-QGA-SVR	China stock market	2013–2014	China

selection model based on AdaBoost was established by using AdaBoost integrated custom weekly classifier model. AdaBoost was not a classifier, but it can be used to enhance the classifier algorithm. The authors used the decision tree classifier to implement the AdaBoost-DT stock selection model. It proved that the model had stronger profitability and risk than the traditional multifactor stock selection model. References [58, 63] were the same as [57], which also applied the AdaBoost enhancement method. However, these two papers used SVM as classifier and establishes AdaBoost-SVM stock selection model. The model with AdaBoost had stronger profitability and smaller return fluctuations than the traditional SVM. The authors of [58] tested SVM models with linear kernel, polynomial kernel, Gaussian kernel, and sigmoid kernel and constructed different K-SVM stock selection models on this basis. The results showed that K-SVM had a strong prediction ability of stock selection returns. The models were constructed by different kernel functions, and they had different prediction abilities, so they needed to be selected according to the original input. Reference [60] constructed an improved support vector regression (SVR) model based on KPCA and genetic algorithm. The stock selection ability of the model was measured in the short and medium-term through fundamental data and transaction data. The authors of [61] used the heuristic algorithm (HA) to process many raw financial data and extracted low-dimensional and effective feature information from it. While preserving the feature information of original data, it also improved the training accuracy and training efficiency. Comparing the HA-SVM with the PCA-SVM model, it proved that the HA-SVM model has higher return rate than the PCA-SVM model. Reference [62] proposed an integrated model of fuzzy dynamic SVM (FD-SVM) for stock selection. The results showed that the average return was higher than the average return in the industry. Compared with the traditional SVM, the accuracy of the model was significantly improved. The authors of [64] proposed a GBDT-SVM multilevel stock selection model based on many factors. The authors used machine learning technology to optimize factor selection and dynamic adjustment of factor weight, to improve the ability of the multifactor model to

obtain stock excess return. The research in China's A-share market showed that the improved model has higher prediction accuracy and higher yield. Reference [65] used RF-QGA-SVR as a stock selection model, used SVR for regression analysis, used RF for feature screening of stock feature variables, used QGA penalty factors to optimize the kernel function and slack variables of SVM, and finally used SVR to obtain the stock return rate and construct a stock selection portfolio by engaging in ranked stocks.

3.2.4. Artificial Neural Network (ANN). ANN is a nonlinear, adaptive information processing system composed of many interconnected processing units. It is proposed based on the research results of modern neuroscience. Moreover, it attempts to process information by simulating neural network processing and memorizing information of the human brain. ANN can be divided into multilayer and single-layer, each layer contains some neurons. A directed arc connects each neuron with variable weight. The network can process information and simulate the relationship between input and output by repeatedly learning and training the known information and gradually adjusting and changing neurons' connection weight. It does not need to know the exact relationship between the input and output and does not need many parameters. It only needs to know the nonconstant factors that cause the output change, namely, the nonconstant parameters. According to the connection's topological structure, ANN model can be divided into two types: forward network and feedback network.

Compared with other traditional methods, ANN has obvious advantages: the problems in the ANN can contain many instances represented by "attribute-value"; the artificial neural network used for the problem with the output of the objective function can be the vector of discrete value, real-value, or several real values or discrete-valued attributes; the learning method of ANN is robust to the noise observations in the training data, the training samples may contain errors, and these errors will not affect the final output; ANNs are usually used where the learning objective function may need to be quickly evaluated; ANNs can withstand a long training time, depending on factors such as the number of

weights in the network, the number of training samples considered, and the setting of various learning algorithm parameters. Table 11 tabulates the stock selection papers using ANN.

The authors of [66] used multilayer feedforward neural network (MFNN) to predict the stock excess returns based on technical and fundamental factors and constructed a hedging portfolio of long and short positions with the same capitalization. The return was higher than that of S&P 500 index. The authors of [67–69, 72–74] used ANNs to select stocks in different stock markets and constructed stock portfolio. The returns of the stock portfolio were higher than ordinary indexes. The authors of [71, 75] utilized the back propagation neural network (BPNN) to conduct stock selection in Taiwan stock market and Shenzhen stock market. Reference [75] added the principal component analysis for index selection, and its accuracy and return were higher than the BPNN network. Reference [70] compared the multilayer perceptron (MLP), the adaptive neurofuzzy inference system (ANFIS), and the general growth and construction radial basis function (GGAP-RBF) and proposed how to use the ROC to choose the stocks systematically.

3.2.5. Random Forest. Random forest is a classifier which contains multiple decision trees, and its output category is determined by the mode of the categories output by individual trees. Random forest integrates multiple decision trees through ensemble learning. Its basic unit is the decision tree, and its essence is an ensemble learning method. Random forest has a good accuracy rate, and it can effectively run on large datasets. Moreover, it can process input samples with high-dimensional characteristics and does not need to reduce the dimension. In the classification problem, the random forest can evaluate the importance of each feature. In the random forest generation process, an unbiased estimation of internal generation error can be obtained, and good results can be obtained for default values.

Random forest can solve classification and regression problems. A large number of studies have proved that it has good performance in classification and regression tasks. Table 12 tabulates the stock selection papers using random forest.

The authors of [76, 78] used random forest to select stocks by financial data, industry data, stock data, and other data. The experimental results showed that its return was higher than the average return. The authors of [77] used the minimum spanning tree (MST) method to investigate the relationship between stocks and explored the stock network structure. Moreover, it used the Kruskal algorithm, PRIM algorithm, and weight matrix method to solve the stock network's minimum spanning tree. Furthermore, it modeled the stock factors by grouping technology and selects stocks. A random forest support vector machine (RFSVM) model was proposed in [79]. The model used the random forest to process the variables of the original data and used random forest to reduce the dimension, which improved the reliability and effectiveness of decision-making. Comparing with other models, PCA-SVM was confirmed that the model had a higher stock selection accuracy.

3.2.6. Deep Learning. As the most important branch of machine learning, deep learning has developed rapidly in recent years. Moreover, it has attracted extensive attention both domestic and abroad. As one of the hottest trends in machine learning and artificial intelligence research, deep learning achieves the purpose of interpreting external data by establishing and simulating the human brain's hierarchical structure to extract features from low-level to high-level external input data. In recent years, as a new branch of machine learning, deep learning has achieved great success in many fields, including image processing, computer vision, speech recognition, machine translation, art, medical imaging, medical information processing, robot control, biology, natural language processing (NLP), and network security.

Deep learning can be regarded as a neural network structure with multiple hidden layers. It can form more abstract high-level representation attributes or features by combining low-level features to discover the distributed feature representation of data. Deep learning is a new technology in the research of machine learning algorithm. Its motivation is to build and simulate the neural network of human brain for analysis and learning. Deep learning can learn a kind of deep nonlinear network structure to represent the input data, realize the complex function approximation, and show the strong ability to learn the essential characteristics of datasets from a few sample sets. Typical deep learning models include deep feedforward neural network, deep belief network, convolution neural network, and recurrent neural network (including LSTM and GRU). Table 13 tabulates the stock selection papers using deep learning.

A multi-index feature selection method based on multi-channel convolutional neural network (MI-CNN) structure was proposed in [80]. This method used the maximum information coefficient feature selection (MICFS) method to select the candidate indicators to ensure the correlation with the stock trend and reduce the redundancy among different indicators, thus constructing a high-yield stock portfolio. References [81–83, 86–88, 92] were all based on the LSTM model for stock selection. The Deep Stock Rank (DSR) model proposed by reference [81] was based on the LSTM model, which can predict the future revenue ranking of stocks and then make the stock selection. Reference [83] selected stocks based on the sequence of yield prediction and studied LSTM and GRU prediction models. The two models' prediction accuracy was better than the traditional forward neural network model, and the GRU model was slightly better than the LSTM model. The authors of [84, 89, 91] proposed the reinforcement learning model based on reward correction for stock selection. These methods used different models to predict the future stock trends and corrected them according to the reward function of reinforcement learning to gradually approach the long-term average level and obtained higher stock selection returns. In [85], a multifactor stock selection model was constructed based on the adaptive recurrent neural network algorithm (RNN-ACT). Compared with RNN, this model has less calculation time, lower cost, and higher accuracy. Compared with LSTM, the model has a significant improvement in accuracy and convergence speed. The authors of

TABLE 11: Stock selection using ANN.

Art.	Methods	Dataset	Period	Country
[66]	MFNN	S&P 500	1989–1995	USA
[67]	ANN	Canada stock market	1989–1994	Canada
[68]	ANN	CLSA	1993–1996	Singapore
[69]	NN	S&P 500 and DJIA	1975–1996	USA
[70]	MLP, ANFIS, and GGAP-RBF	DJIA	1995–2004	USA
[71]	BPNN	Taiwan stock market	2000–2007	Taiwan, China
[72]	ANN	S&P 500	1991–2010	USA
[73]	ANN	S&P 500	2012.05–2017.03	USA
[74]	FNN	DJIA	2013–2017	USA
[75]	FCA and BPNN	Shenzhen stock market	2014–2016	China

TABLE 12: Stock selection using random forest.

Art.	Methods	Dataset	Period	Country
[76]	RF	CSI 500	2013.02–2017.08	China
[77]	MST	China stock market	1997–2004	China
[78]	RF	China stock market	2007–2015	China
[79]	RF-SVM	China stock market	2015	China

TABLE 13: Stock selection using deep learning.

Art.	Methods	Dataset	Period	Country
[80]	MI-CNN	S&P 500	2008–2017	USA
[81]	DSR	China stock market	2006–2017	China
[82]	LSTM	CSI 300	2007–2017	China
[83]	LSTM and GRU	S&P 500	1960–2018	USA
[84]	RL	China stock market	2016–2017	China
[85]	RNN-ACT	CSI 300	2010–2017	China
[86]	LSTM	China stock market	2010–2018	China
[87]	LSTM	CSI 300	2020–2021	China
[88]	LSTM	India stock market	2004–2019	India
[89]	RL	300 China stocks	2016	China
[90]	DNN	CSI 500	2017–2020	China
[91]	RL	S&P 500	1998–2018	USA
[92]	CNN and LSTM	CSI 300	2014–2018	China
[93]	CNN	S&P 500	2010–2017	USA

[93] proposed a CNN stock selection model, and this model can select profitable stocks with low risk and forecasting close prices for a given horizon.

3.2.7. Other Machine Learning Methods. We also found some other stock selection methods. There is little literature on these methods. Therefore, they are hard to generalize into the above categories. Table 14 tabulates the stock selection papers using other machine learning methods.

The authors of [94] used prototype ranking (PR) for stock selection. The authors of [95] used the K -nearest neighbour for stock selection. The authors of [96] applied ListNet and RankNet to carry out long and short stock selection strategies, and the results were better than the benchmark. Reference [97] established a nonlinear integer programming model and proposed a heuristic method to solve this problem, making the stock selection more reasonable. Reference [98] carried out the compound model of global stocks and used the data mining method and the multifactor model to construct the stock selection portfolios. The authors of [99] compared the Holm decline process and the Hochberg rise process of different loss functions, taking the conditional risk as the selection threshold function and selecting stocks with sharp ratio. In [100], a linear programming model based on the ordered weighted average (OWA) operator was proposed to identify the optimal stock portfolio without reordering. Based on cash flow and financial indicators, [101] used association rules to select stocks, making the results of stock selection more reliable and practical. In [102], the authors used the block-bootstrap method to eliminate the funds with weak stock selection ability and selected the stock funds with significant ability. In [103], 11 factors were selected using the multifactor model, and the validity test was conducted by the regression method. The effective factors were selected to construct a multiple regression stock selection models. Reference [104] proposed a hybrid stock selection model combined with stock forecasts, which effectively captures the complex stock market's future characteristics and effectively constructs stock portfolio. The authors of [105] used Piotroski financial index scoring method based on P/B stock characteristics to conduct stock selection in China's stock market.

TABLE 14: Stock selection using other machine learning methods.

Art.	Methods	Dataset	Period	Country
[94]	Prototype ranking (PR)	New York stock market	1962–2004	USA
[95]	KNN	China stock market	1999–2016	China
[96]	ListNet and RankNet	TRNA	2003–2014	USA
[97]	Nonlinear integer programming	S&P 500	1 year	USA
[98]	Data mining	Fact set	1997–2011	USA
[99]	Holm and Hochberg	New York stock market	2010–2014	USA
[100]	OWA	Tehran stock exchange market	2011	Iran
[101]	Association rules	SET	2014–2017	Thailand
[102]	Block-bootstrap	China stock market	2004–2008	China
[103]	Regression	China stock market	2014–2016	China
[104]	Highly weighted predicted factors	China stock market	2006–2016	China
[105]	Piotroski	China stock market	2000–2012	China

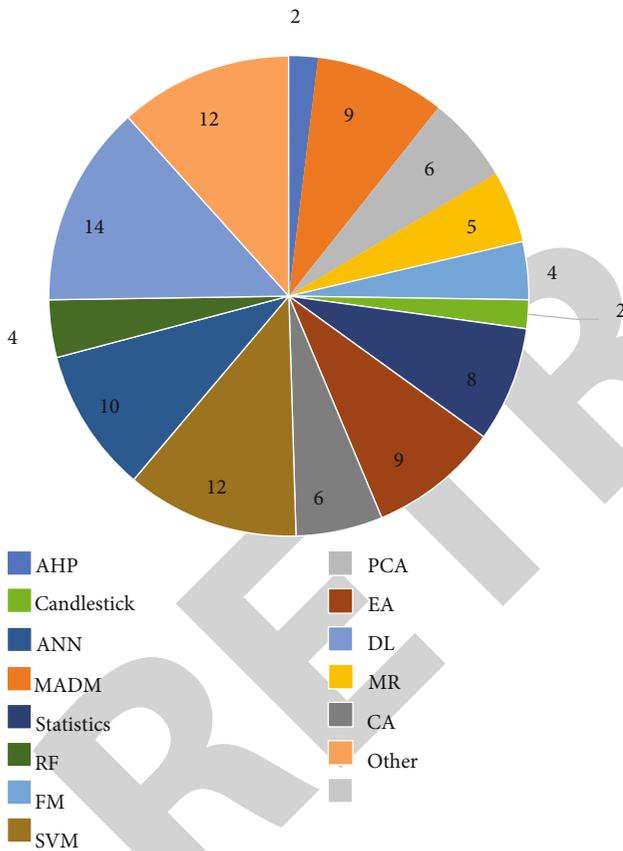


FIGURE 2: The stock selection methods.

4. Statistics and Analysis

After reviewing the 103 studies on stock selection method, we are now ready to provide some general statistical data on the current research status and analyse the future development trend on this basis.

According to the different methods of stock selection in the previous chapter, we classified the literature, as shown in Figure 2.

Figure 2 shows the statistics of different stock selection methods in 103 studies we investigated. We found that with the continuous development of machine learning and deep learning in recent years, more and more literatures use deep learning methods to research stock selection, which solved some of the pain points in stock selection and achieves better research results.

Figure 3 shows the research on different stock markets in the 103 studies we investigated. So far, 53 studies related to stock selection in the China stock market have been published, followed by 25 related articles in the US market, and 4 related articles in Europe. Moreover, there are 5 related articles in Japan and 17 related articles in other markets. Most of the studies we investigated were published after 2000. As shown in Figure 3, with the growing maturity of China’s stock market and the continuous expansion of market scale, more and more researchers have achieved very good research results in China’s stock market.

Figure 4 shows the publication date of the studies we investigated. We can see that the number of stock selection methods is increasing in the past few years. In particular, combined with previous Figures 1 and 2, the increase of studies is closely related to the continuous development of machine learning and deep learning technology in recent years.

After the comprehensive statistics and analysis of 103 studies, we believe that the machine learning methods will be researched and applied by more and more researchers and will have more significant performance improvement in stock selection and quantitative trading. For the increasingly complex market environment, the advantage of machine learning will be further amplified, and the shortcomings of traditional methods in stock selection will be supplemented and improved by machine learning methods. We believe that the future stock selection method will be a comprehensive method that complements the traditional method and the machine learning method. Researchers should not only learn new computer methods but also fully understand the traditional methods. Based on fully analysing market data and behavior, use corresponding methods for research and application.

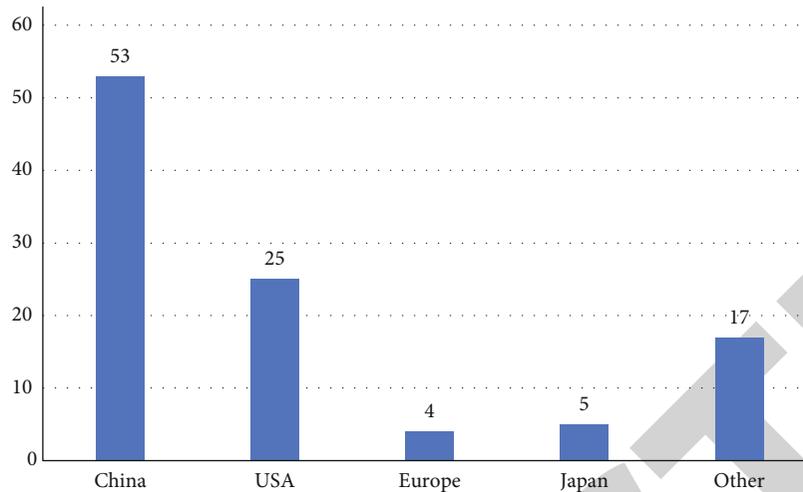


FIGURE 3: The histogram of publication market.

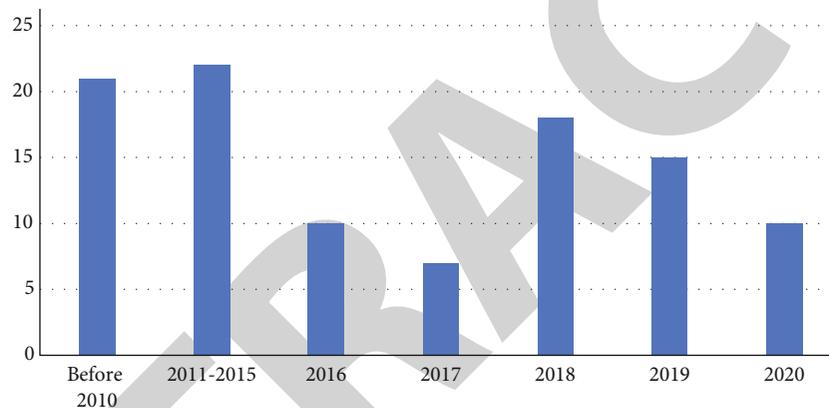


FIGURE 4: The histogram of publication count in years.

We choose to return to our original research question through our research on the literature of the survey. Our answers are as follows:

(i) What models can be used to quantify stock selection?

Answer: both traditional statistical methods and classification methods can build stock selection models, which need to be screened according to different data sources and basic strategies. In particular, deep learning methods have received more and more attention and have been researched further and applied widely in recent years.

(ii) Can traditional stock selection methods be used in the current stock market?

Answer: the traditional stock selection method can still be used in the current stock market. The behavior of the stock market has not changed in essence. The traditional stock selection method has been fully verified in theory and practice. No matter how the stock market changes, its basic principle has not changed fundamentally. As long as the data and strategies used in stock selection are consistent

with the corresponding market. The traditional stock selection methods can still be used in the current stock market.

(iii) Compared with traditional stock selection methods, how effective is machine learning represented by deep learning?

Answer: the results of most studies about machine learning and deep learning show that in the stock selection method, machine learning and deep learning make the stock selection result more reasonable and efficient in terms of data dimensions and factor selection compared with traditional stock selection methods.

5. Conclusion

Over the years, as an important part of the quantitative investment strategy, stock selection has also changed rapidly with the continuous development of computer technology. Recently, with the development of machine learning, quantitative investment methods have been promoted, and many new studies have appeared. This paper hopes to review

existing studies and provide a more comprehensive description of the current research status for stock selection methods. Our research results show that although the research history of the stock selection method is very long, stock selection methods also have new research directions and ideas with the development of machine learning. After reviewing the studies on stock selection, we think there are several options for the development of stock selection in the future.

Firstly, we think that the advantages of RNN are becoming more and more obvious in time series data prediction, especially the successful application of LSTM and GRU in many fields. It will promote the increasing application and innovation of stock selection.

Secondly, the advantages of CNN in processing image have been confirmed. Stock data can be more directly converted into 2D images, whether in the form of curve or candlestick chart. So we really believe that how to use CNN to process stock image will also be a relatively innovative research direction.

In addition, some stock selection methods are based on text data. How the text data represented by news affects stock selection is also a very valuable research direction.

Finally, we think that with the increasing number and dimension of stock data, single method, whether traditional method or machine learning method, cannot meet the higher requirements for stock selection. Therefore, the use of ensemble learning should help to build a more stable and efficient stock selection models.

Stock selection is a very interesting and valuable research field. Our future work will focus on the application of deep learning methods in stock selection and find the best way to integrate multiple methods into stock selection.

Data Availability

No data are required.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] R. K. Narang, *Inside the Black Box: A Simple Guide to Quantitative and High-Frequency Trading*, John Wiley & Sons, Ltd, 2013.
- [2] T. L. Saaty, *The Analytic Hierarchy Process: Planning, Priority Setting, Resource Allocation*, McGraw-Hill, 1980.
- [3] C. Yang and C. Zhang, "Application of attribute hierarchical model in decision making of stock selection," *College Mathematics*, vol. 22, no. 5, pp. 27–30, 2006.
- [4] X. He and W. Liu, "Decision making of stock selection based on fuzzy analytic hierarchy process," *Journal of Chang Chun University*, vol. 21, no. 6, pp. 43–46, 2011.
- [5] K. Y. Shen, M. R. Yan, and G. H. Tzeng, "Combining VIKOR-DANP model for glamor stock selection and stock performance improvement," *Knowledge-Based Systems*, vol. 58, no. 1, pp. 86–97, 2014.
- [6] P. A. Koldanov and G. A. Bautin, *Multiple Decision Problem for Stock Selection in Market Network [J]*, Springer International Publishing Switzerland, 2014.
- [7] W.-S. Lee, G.-H. Tzeng, J.-L. Guan, K.-T. Chien, and J.-M. Huang, "Combined MCDM techniques for exploring stock selection based on Gordon model," *Expert Systems with Applications*, vol. 36, no. 3, pp. 6421–6430, 2009.
- [8] K. Y. Shen, V. Y. C. Chen, and G. H. Tzeng, "An integrated MCDM model for glamour stock selection," in *Joint International Conference on Soft Computing & Intelligent Systems* IEEE.
- [9] M. F. Bayramoglu and C. Hamzacebi, "Stock selection based on fundamental analysis approach by Grey relational analysis: a case of Turkey," *International Journal of Economic and Finance*, vol. 8, no. 7, pp. 178–184, 2016.
- [10] O. Kwang Wuk, S. Lim, and J. Zhu, "Use of DEA cross-efficiency evaluation in portfolio selection: an application to Korean stock market," *European Journal of Operational Research*, vol. 236, no. 1, pp. 361–368, 2014.
- [11] G. R. Amin and M. Hajjani, "Application of optimistic and pessimistic OWA and DEA methods in stock selection," *International Journal of Intelligent Systems*, vol. 31, no. 12, pp. 1220–1233, 2016.
- [12] G. R. Amin and A. Oukil, "Gangless cross-evaluation in DEA: an application to stock selection," *RAIRO-Operations Research*, vol. 53, no. 2, pp. 645–655, 2019.
- [13] D. K. Joshi and S. Kumar, "An interactive approach to probabilistic intuitionistic fuzzy multi-criteria decision making in stock selection problem," *Journal of Reliability and Statistical Studies*, vol. 11, no. 2, pp. 25–36, 2018.
- [14] Y. He, K. Fataliyev, and L. Wang, "Feature selection for stock market analysis," in *ICONIP*, pp. 737–744, 2013.
- [15] H. Yu, R. Chen, and G. Zhang, "A SVM stock selection model within PCA," *Procedia computer science*, vol. 31, pp. 406–412, 2014.
- [16] L. Yang, W. Rea, and A. Rea, "Stock selection with principal component analysis," *Working papers in Economics*, Stock selection with principal component analysis, 2015.
- [17] A. John, F. T. Oduro, and J. Ackora-Prah, "Selection of stocks on the Ghana stock exchange using principal component analysis," *International Journal of Theoretical and Applied Mathematics*, vol. 2, no. 2, pp. 100–109, 2016.
- [18] M. Zhou and L. Yin, "Quantitative stock selection strategies based on kernel principal component analysis," *Journal of Financial Risk Management*, vol. 9, no. 1, pp. 23–43, 2020.
- [19] Y. Huang, H. Fang, L. Dai, and T. Chen, "Research on multi-factor stock selection model based on EKPCA algorithm," *Finance*, vol. 9, no. 4, pp. 327–340, 2019.
- [20] J. van der Hart, E. Slagter, and D. van Dijk, "Stock selection strategies in emerging markets," *Journal of Empirical Finance*, vol. 10, no. 1-2, pp. 105–132, 2003.
- [21] W. Fan, "Study on stock selection strategy based on SPSS regression mean quantization," *Business and Management Research*, vol. 5, no. 3, pp. 26–29, 2016.
- [22] Z. Kang, *Quantitative Stock Selection Strategy Based on Alpha Mean Regression*, Shanghai Normal University, Shanghai, 2018.
- [23] Z. Liang, "Multi-factor strategy based on quantile regression," *Journal of Southwest University (Natural Science Edition)*, vol. 41, no. 1, pp. 89–96, 2019.

- [24] J. Guerard and A. Saxena, "A case study of forecasted earnings acceleration and stock selection in global and emerging stock markets," *Frontiers in Applied Mathematics and Statistics*, vol. 4, pp. 1–9, 2018.
- [25] F. Tiryaki and M. Ahlatcioglu, "Fuzzy stock selection using a new fuzzy ranking and weighting algorithm," *Applied Mathematics & Computation*, vol. 170, no. 1, pp. 144–157, 2005.
- [26] S. Yodmun and W. Witayakiattilerd, "Stock selection into portfolio by fuzzy quantitative analysis and fuzzy multicriteria decision making," *Adv. Oper. Res.*, vol. 2016, pp. 1–14, 2016.
- [27] I. Patni and N. Gupta, "Construction of optimal portfolio and selection of stock using fuzzy approach," *Fuzzy Approach*, vol. 7, no. 3.30, pp. 118–121, 2018.
- [28] W. Witayakiattilerd, "Fuzzy quantitative analysis method for stock selection into portfolio," *Chiang Mai Journal of Science*, vol. 46, no. 4, pp. 799–811, 2019.
- [29] M. J. Horton, "Stars, crows, and doji: the use of candlesticks in stock selection," *Quarterly Review of Economics & Finance*, vol. 49, no. 2, pp. 283–294, 2009.
- [30] S. Barak, J. H. Dahooie, and T. Tichy, "Wrapper ANFIS-ICA method to do stock market timing and feature selection on the basis of Japanese candlestick," *Expert Systems with Applications*, vol. 42, no. 23, pp. 9221–9235, 2015.
- [31] G. H. John, P. Miller, and R. Kerber, "Stock selection using rule induction," *IEEE Expert*, vol. 11, no. 5, pp. 52–58, 1996.
- [32] L. André, R. van Dijk, and T. Kloek, "Stock selection, style rotation, and risk," *Journal of Empirical Finance*, vol. 9, no. 1, pp. 1–34, 2002.
- [33] J. Salim, "Copula model selection of stock return time series using information complexity," *Journal of Business Theory and Practice*, vol. 6, no. 4, pp. 294–307, 2018.
- [34] W. Wu, J. Chen, L. Xu, Q. He, and M. L. Tindall, "A statistical learning approach for stock selection in the Chinese stock market," *Innovations*, vol. 5, no. 1, 2019.
- [35] Q. Yingneng and H. Yunfa, "Stock selection using naïve Bayes classifier," *Computer Application and Software*, vol. 24, no. 6, pp. 90–92, 2007.
- [36] M. Suzuki, "PSR—an efficient stock-selection tool?," *International Journal of Forecasting*, vol. 14, no. 2, pp. 245–254, 1998.
- [37] N. Goumatianos, I. Christou, and P. Lindgren, "Stock selection system: building long/short portfolios using intraday patterns," *Procedia Economics and Finance*, vol. 5, pp. 298–307, 2013.
- [38] B. Sun, H. Li, P. An, and Z. Wang, "Dynamic energy stock selection based on shareholders' coholding network," *A Statistical Mechanics and Its Applications*, vol. 542, p. 122243, 2020.
- [39] Y. L. Becker and U. M. O'Reilly, "Genetic programming for quantitative stock selection," in *Genetic and Evolutionary Computation Conference, GEC Summit 2009, Proceedings*, Shanghai, China, 2009.
- [40] C.-F. Huang, C.-H. Chang, B. R. Chang, and D.-W. Cheng, "A study of a hybrid evolutionary fuzzy model for stock selection," in *2011 IEEE international conference on fuzzy systems*, Taipei, Taiwan, 2011.
- [41] C. F. Huang, T. N. Hsieh, B. R. Chang, and C.-H. Chang, "A comparative study of regression and evolution-based stock selection models for investor sentiment," in *Third International Conference on Innovations in Bio-inspired Computing & Applications* IEEE Computer Society.
- [42] C. F. Huang, "A hybrid stock selection model using genetic algorithms and support vector regression," *Applied Soft Computing Journal*, vol. 12, no. 2, pp. 807–818, 2012.
- [43] S. S. Chen, C. F. Huang, and T. P. Hong, *An Improved Multi-Objective Genetic Model for Stock Selection with Domain Knowledge*, Springer International Publishing, 2014.
- [44] L. Yu, L. Hu, and L. Tang, "Stock selection with a novel sigmoid-based mixed discrete-continuous differential evolution algorithm," *IEEE Transactions on Knowledge & Data Engineering*, vol. 28, no. 7, pp. 1891–1904, 2016.
- [45] G. S. M. Thakur, R. Bhattacharyya, and S. Sarkar, "Stock portfolio selection using Dempster-Shafer evidence theory," *Journal of King Saud University-Computer and Information Sciences*, vol. 30, no. 2, pp. 223–235, 2018.
- [46] D. Panwar, M. Jha, and N. Srivastava, "Stock selection and portfolio optimization using teaching-learning-based algorithm," *Journal of Advanced Research in Dynamical and Control Systems*, vol. 10, no. 9, pp. 433–443, 2019.
- [47] L. Ji, *Design and Implementation of Multifactor Strategy Optimization System Based on Genetic Algorithm*, South China University of Technology, 2018.
- [48] S. Craighead and B. Klemesrud, "Stock selection based on cluster and outlier analysis," in *IMA Conference.*, 2002.
- [49] N. Da Costa, J. Cunha, and S. Da Silva, "Stock selection based on cluster analysis," *Finance*, vol. 13, no. 1, pp. 1–9, 2005.
- [50] R. Wang, "Stock selection based on data clustering method," in *2011 Seventh International Conference on Computational Intelligence and Security* IEEE.
- [51] R. Peachavanish, "Stock selection and trading based on cluster analysis of trend and momentum indicators," in *Proceedings of the International Multi Conference of Engineers and Computer Scientists*, 2016.
- [52] R. Subekti, E. R. Sari, R. Kusumawati, H. O. Pintari, and P. Renggani, "Value at risk in the Black Litterman portfolio with stock selection through cluster analysis," *Journal of Physics Conference Series*, vol. 1320, no. 1, p. 012004, 2019.
- [53] L. Wenxing and L. Junqi, "Improvement of semi-supervised kernel clustering algorithm based on multi-factor stock selection," *Statistics & Information Forum*, vol. 33, no. 3, pp. 30–36, 2018.
- [54] M. C. Lee, "Using support vector machine with a hybrid feature selection method to the stock trend prediction," *Expert Systems with Applications*, vol. 36, no. 8, pp. 10896–10904, 2009.
- [55] K. Zikowski, "Using volume weighted support vector machines with walk forward testing and feature selection for the purpose of creating stock trading strategy," *Expert Systems with Applications*, vol. 42, no. 4, pp. 1797–1805, 2015.
- [56] W. Jianzhou, R. H. Qang, C. Wang, and L. Shen, "Improved v-support vector regression model based on variable selection and brain storm optimization for stock price forecasting," *Applied Soft Computing*, vol. 49, pp. 164–178, 2016.
- [57] R. Zhang and T. Cao, "Multi-factor stock selection model based on Adaboost," *Business and Economic Research*, vol. 8, no. 4, pp. 119–129, 2018.
- [58] Z. Ru, L. Zi-Ang, C. Shaozhen, C. M. Zhao, and M. J. Yuan, "Adaboost-SVM multi-factor stock selection model based on Adaboost enhancement," *International Journal of Statistics & Probability*, vol. 7, no. 5, pp. 9–18, 2018.
- [59] R. Zhang, Z. A. Lin, S. Chen, Z. Lin, and X. Liang, "Multi-factor stock selection model based on kernel support vector machine," *Journal of Mathematics Research*, vol. 10, no. 5, pp. 9–129, 2018.

- [60] S. U. Zhi and F. U. Xiaoyuan, "Kernel principal component genetic algorithm and improved SVR stock selection model," *Statistical Research*, vol. 30, no. 5, pp. 55–62, 2013.
- [61] C. Rongda and Y. Huanhuan, "Stock selection model based on support vector machine within heuristic algorithm," *Systems Engineering*, vol. 32, no. 2, pp. 40–48, 2014.
- [62] G. Anjing, "Research on stock selection based on the method of dynamic fuzzy integration support vector machine," *Harbin Institute of Technology*, vol. 2, 2017.
- [63] W. Lun, "Multi-factor stock selection model based on Ada-boost-SVM," *Economic Research Guide*, vol. 10, pp. 107–108, 2019.
- [64] M. Qingyan, "Optimizing multi-factor stock selection system using GBDT-SVM multi-level model," *Statistics and Application*, vol. 8, no. 1, pp. 184–192, 2019.
- [65] L. Tang and Q. Lin, "Stock selection based on a hybrid quantitative method," *Open Journal of Statistics*, vol. 6, no. 2, pp. 346–362, 2016.
- [66] A. E. Levin, "Stock selection via nonlinear multi-factor models," *Advances in Neural Information Processing Systems*, vol. 8, 1996.
- [67] J. Ghosn and Y. Bengio, "Multi-task learning for stock selection," *Advances in Neural Information Processing Systems*, vol. 9, 1997.
- [68] T. S. Quah and B. Srinivasan, "Improving returns on stock investment through neural network selection," *Expert Systems with Applications*, vol. 17, no. 4, pp. 295–301, 2006.
- [69] G. Stanley and S. R. Stansell, "Can value-based stock selection criteria yield superior risk-adjusted returns: an application of neural networks," *International Review of Financial Analysis*, vol. 12, no. 1, pp. 83–97, 2003.
- [70] T.-S. Quah, "Using neural network for DJIA stock selection," *Engineering Letters*, vol. 15, 2007.
- [71] C. F. Tsai and Y. C. Hsiao, "Combining multiple feature selection methods for stock prediction: union, intersection, and multi-intersection approaches," *Decision Support Systems*, vol. 50, no. 1, pp. 258–269, 2011.
- [72] Y. C. Liu and I. Yeh, "Using mixture design and neural networks to build stock selection decision support systems," *Neural Computing & Applications*, vol. 28, no. 3, pp. 521–535, 2017.
- [73] G. H. Kim and S. H. Kim, "Variable selection for artificial neural networks with applications for stock price prediction," *Applied Artificial Intelligence*, vol. 33, no. 1, pp. 54–67, 2019.
- [74] Y. Huang, L. F. Capretz, and D. Ho, "Neural network models for stock selection based on fundamental analysis," in *2019 IEEE Canadian Conference of Electrical and Computer Engineering* IEEE.
- [75] L. Taiqiang, *An Empirical Study on Comprehensive Stock Selection Based on Principal Component Analysis and BP Neural Network*, Ji'nan University, 2017.
- [76] Z. Tan, Z. Yan, and G. Zhu, "Stock selection with random forest: an exploitation of excess return in the Chinese stock market," *Heliyon*, vol. 5, no. 8, 2019.
- [77] X. Zheng, *Multi-Factor Model Based on the Minimal Spanning Tree*, Hebei University of Technology, 2016.
- [78] L. Qi and Y. Junqi, "Application on random forest algorithm in multi-factor stock selection," *Manager Journal*, vol. 2, p. 243, 2017.
- [79] J. Xiujuan, "Quantitative stock selection based on support vector machine of random forest," *Journal of Regional Financial Research*, vol. 1, pp. 27–30, 2019.
- [80] H. Yang, Y. Zhu, and Q. Huang, "A multi-indicator feature selection for CNN-driven stock index prediction," *Springer Nature Switzerland AG*, vol. 11305, pp. 35–46, 2018.
- [81] X. Zhang and Y. Tan, "Deep stock ranker: a LSTM neural network model for stock selection," *Data Mining and Big Data*, vol. 10943, 2018.
- [82] R. Zhang, C. Huang, W. Zhang, and S. Chen, "Multi factor stock selection model based on LSTM," *International Journal of Economics and Finance*, vol. 10, no. 8, p. 36, 2018.
- [83] J. Sun, "A stock selection method based on earning yield forecast using sequence prediction models," 2019, <http://arxiv.org/abs/1905.04842>.
- [84] H. Chen, C. Zhang, and Y. Li, "A reinforcement learning model based on reward correction for quantitative stock selection," *IOP Conference Series Materials ence and Engineering*, vol. 768, no. 7, p. 072036, 2020.
- [85] Z. Zhiyuan, *Research and Application of Multi-Factor Stock Selection Model Based on RNN-ACT Algorithm*, Kunming University of Technology, 2018.
- [86] X. Yuan, J. Yuan, T. Jiang, and Q. U. Ain, "Integrated long-term stock selection models based on feature selection and machine learning algorithms for China stock market," *IEEE Access*, vol. 8, pp. 22672–22685, 2020.
- [87] W. Xiao and Y. Tang, "Quantitative stock selection model based on long-short term memory (LSTM) neural network," *Proceedings of Business and Economic Studies*, vol. 4, no. 3, pp. 19–24, 2021.
- [88] S. Mahlawat, U. Prabhakar, N. Goyal, P. Parth, and V. Ramamohan, "A long short-term memory approach towards stock selection and portfolio optimization," 2020, Available at SSRN 3673874.
- [89] Y. F. Shen, T. Liu, W. K. Liu, R. Q. Xu, Z. Li, and J. Wang, "Deep reinforcement learning for stock recommendation," *Journal of Physics: Conference Series*, vol. 250, no. 1, p. 012012, 2021.
- [90] Q. Sun, "Neural network algorithm strategy based on multi-factor stock selection," in *2020 International Conference on Applications and Techniques in Cyber Intelligence*, 2020.
- [91] T. Huotari, J. Savolainen, and M. Collan, "Deep reinforcement learning agent for S&P 500 stock selection," *Axioms*, vol. 9, no. 4, p. 130, 2020.
- [92] J. Yang, Y. Li, X. Chen, J. Cao, and K. Jiang, "Deep learning for stock selection based on high frequency price-volume data," 2020, <http://arxiv.org/abs/1911.02502>.
- [93] Y. Touzani, D. Khadija, and F. Khoukhi, "Stock price forecasting: new model for stocks selection and price forecasting based on convolutional neural network," *Sustainable Development*, vol. 1105, 2020.
- [94] R. J. Yan and C. X. Ling, "Machine learning for stock selection," in *Industrial and Government Track Short Paper*, pp. 1038–1042, Springer, 2007.
- [95] H. Sun, W. Rong, J. Zhang, Q. Liang, and Z. Xiong, "Stacked denoising autoencoder based stock market trend prediction via K-nearest neighbour data selection," in *International Conference on Neural Information Processing*, 2017.
- [96] Q. Song, A. Liu, and S. Y. Yang, "Stock portfolio selection using learning-to-rank algorithms with news sentiment," *Neurocomputing*, vol. 264, pp. 20–28, 2017.

- [97] K. Gnanendran, J. K. Ho, and R. P. Sundarraj, "Stock selection heuristics for interdependent items," *European Journal of Operational Research*, vol. 145, no. 3, pp. 585–605, 2003.
- [98] J. Guerard, H. Markowitz, and G. Xu, "Earnings forecasting in a global stock selection model and efficient portfolio construction and management," *SSRN Electronic Journal*, vol. 31, 2014.
- [99] A. P. Koldanov, V. A. Kalyagin, and P. M. Pardalos, *Step Down and Step Up Statistical Procedures for Stock Selection with Sharp Ratio*, International Workshop on Machine Learning. Springer International Publishing, 2015.
- [100] H. Mohaddeseh and G. R. Amin, "Modelling stock selection using ordered weighted averaging operator," *International Journal of Intelligent Systems*, vol. 33, no. 11, pp. 2283–2292, 2018.
- [101] A. Wijitcharoen and P. Padungweang, "Stock selection using association rules on cash flow and accrual financial indicators," in *IAIT*, 2018.
- [102] X. Ning, L. Zhixin, and C. Wenyi, "Stock selection ability of mutual fund based on block-bootstrap technique," *Systems Engineering*, vol. 27, no. 3, pp. 46–52, 2009.
- [103] X. Jingzhao, "Analysis on quantitative investing strategy based on multi-factor model," *Exploration of Financial Theory*, vol. 3, pp. 30–38, 2017.
- [104] F. Yang, Z. Chen, J. Li, and L. Tang, "A novel hybrid stock selection method with stock prediction," *Applied Soft Computing Journal*, vol. 80, pp. 820–831, 2019.
- [105] Y. Xin, Z. Feng, C. Yuji, and Z. Zhuang, "An empirical study on Piotroski stock selection strategy based on value investment," *Times Finance*, vol. 8, pp. 20–22, 2012.