Expert Systems With Applications 200 (2022) 116941

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

An adaptive feature selection schema using improved technical indicators for predicting stock price movements

Gang Ji [a], Jingmin Yu [b], Kai Hu [a,*], Jie Xie [a], Xunsheng Ji [a]

[a] Key Laboratory of Advanced Process Control for Light Industry, Ministry of Education, Jiangnan University, Wuxi, PR China
[b] Wuhan Geomatics Institute, Wuhan, PR China

ARTICLE INFO

Keywords:
Stock price movement direction forecasting
Wavelet denoising
Feature selection
Time window

ABSTRACT

Accurate stock market forecasts can bring high returns for investors. There have been a growing number of studies employing machine learning technology to perform stock prediction tasks with the development of machine learning and artificial intelligence technologies. However, accurately predicting stock price trends still is an elusive goal, not only because the stock market is affected by policies, market environment, market sentiment, etc., but also because stock price data is inherently complex, noisy, and nonlinear. Many technical indicators have been used as input features to stock prediction models, but the quality of technical indicators has always been a neglected issue, thus the application of feature engineering in stock prediction tasks needs to be further expanded. Using 18 technical indicators as the original features, this paper presents improved technical indicators based on wavelet denoising and a novel two-stage adaptive feature selection method. Finally, the random forest model is used as the stock prediction model. Experiments show that in contrast to the original technical indicators, the improved technical indicators significantly enhance the performance of the model (e.g., F1 scores increased by 34.48% on the SSE Composite Index (SSEC) data set, 41.56% on the Hang Seng Index (HSI) data set, 34.48% on the Dow Jones Industrial Average (DJI) data set, 32.75% on the Standard & Poor's 500 Index (S&P 500) data set). The experimental results verify the importance of the quality of technical indicators in the task of stock prediction. Meanwhile, the results also demonstrate the effectiveness of the feature selection method, which can achieve higher prediction accuracy with fewer features. In addition, we established multiple data sets according to the size-varied time windows to study the influence of the size-varied time windows. The results show that properly increasing the size of the time window can exert a positive impact on the model. Finally, by utilizing our two-stage adaptive feature selection method, we remove redundant features, and achieve excellent results on data sets from four different stock markets (e.g., F1 scores reached 0.754 on the SSEC data set, 0.794 on the HSI data set, 0.789 on the DJI data set, 0.821 on the S&P 500 data set). Overall, this study experimentally verifies that improving feature quality can positively impact model performance, and that choosing an appropriate combination of input features can not only improve model performance, but reduces the negative impact of the curse of dimensionality as well.

1. Introduction

Nowadays, the stock market has become one of the main fields for investment. Accurate prediction of the stock market movement can reduce risks and can bring abundant and excessive returns. Therefore, it is necessary to study the forecasting methods for stock market movements. According to the Efficient Market Hypothesis (EMH) proposed by Fama in 1970 (Fama, 1970), stock prices reflect all information in the stock market. Furthermore, Fama argued that stock prices follow random walks (Fama, 1965). These two theories indicate that stock price changes are independent, and a series of stock price changes have no memory. Thus, the historical information cannot be used to predict future stock market prices. However, a large number of recent studies have proposed a different view. Research results have shown that financial markets are predictable to a certain extent (Bollerslev et al., 2014; Ferreira and Santa-Clara, 2011; Phan et al., 2015). Shah et al. (2019) reviewed many studies about stock forecasting. These studies show that stock market forecasting can outperform the overall market.

* Corresponding author.
E-mail addresses: lijjyuackle@gmail.com (G. Ji), yjingmin@126.com (J. Yu), hukai_wlw@jiangnan.edu.cn (K. Hu), xiej2018@jiangnan.edu.cn (J. Xie), jixunsheng@163.com (X. Ji).

https://doi.org/10.1016/j.eswa.2022.116941
Received 29 May 2021; Received in revised form 16 March 2022; Accepted 17 March 2022
Available online 19 March 2022
0957-4174/

These forecasting methods are often inspired by two traditional methods for stock market analysis, fundamental and technical analysis.

Fundamental analysis focuses on the intrinsic value of the company and other factors that will affect the stock price, such as policies, macroeconomics, etc. Technical analysis takes stock price fluctuation behavior as the central research target, using historical price information, trading volume information, and statistical-based technical indicators to predict stock price fluctuations. Compared to technical indicators, fundamental indicators have a lower update frequency, because companies publish their achievements quarterly, semi-annually, or even annually. The low update frequency makes fundamental-analysis-based models less efficient in short-term forecasting research. In contrast; indicators based on technical analysis have always been a hot topic, especially in short-term stock price forecasting and price movement direction forecasting research. Designers and users of technical indicators do not accept the EMH which assumes that the current price reflects all knowledge. They are convinced that news, policy changes, and other factors are not fully reflected in stock prices. Therefore, an effective way to forecast the stock market is by analyzing the price movement model, namely the technical indicators. Much research has been carried out based on the technical-analysis-based models and has achieved high accuracy. Some of them are used in stock market trading, bringing about high returns (Dinesh et al., 2021; Long et al., 2020; Naik and Mohan, 2019). A review of stock technical analysis research of 55 years (Farias Nazário et al., 2017) confirmed the validity of the technical analysis. Therefore, using technical indicators to build a prediction model is a widely-recognized and promising research direction.

However, using the technical analysis properly is not an easy task, for it requires operation experience and a profound understanding of the technical indicators. The key to establishing technical analysis-based indicators is to properly express historical price and volume information. Since the emergence of the stock market, financial analysts have developed a large number of technical indicators to reflect the information hidden in the historical movements of the stock, such as Moving average convergence divergence (MACD), Relative Strength Index (RSI), Williams Overbought/Oversold Index (W&R), and etc. These indicators play a certain role in analyzing the stock price movement, but the performance when using the indicators solely in a statistical way is not ideal. For example, MACD is an indicator used to judge the trend. The user usually supposes that when a golden cross appears in the MACD trend plots, it is a signal of stock price rise, while a death cross signifies a stock price decline. But the actual trend may be the opposite. Other technical indicators also encounter the same situation. The time-series stock price data typically includes a great deal of noise because of the effect of various factors such as news, policies, and market sentiment. These factors often lead to the inefficiency of technical indicators in practical applications.

To accurately predict the stock market to obtain high returns, researchers have proposed many forecasting models (Jiang, 2021). Most traditional time-series forecasting methods are based on statistical theory, such as the exponential smoothing model (ESM), autoregressive integrated moving average (ARIMA), and their extensions. As reviewed in Shah et al. (2019), methods based on statistical theory have high interpretability from a statistical point of view. Unfortunately, the statistical method assumes that the stock price data is linear, stable, and follows a fixed distribution. But research shows that stock price data has strong nonlinear characteristics (Bartiromo, 2004; Jiang et al., 2007). This contradiction limits the performance of statistical models in stock forecasting.

Machine learning, including neural networks, deep learning, ensemble learning methods, etc., may be a new choice. Since its emergence, machine learning has been broadly used in the classification and prediction tasks of nonlinear data (Masini et al., 2020). The excellent performance of machine learning on nonlinear data makes it popular in stock forecasting tasks (Ballings et al., 2015; Patel et al., 2015b). Like other machine learning tasks, stock prediction also requires the selection of a suitable combination of features from a given set as the input to a prediction model. The quality of the features in the set is also one of the keys to the performance of the prediction model. Therefore, in this article, we propose a method for improving technical indicators based on wavelet denoising and we design a feature selection method that is more suitable for selecting the optimal feature subset from feature sets with different feature magnitudes. In addition, we also considered the size-varied time windows to study the impact of historical features on the model. We conduct experiments on four real data sets of different stock markets to prove the effectiveness of our method and discuss the results.

The following text is organized as follows: In Section 2, all related work, including technical indicators, are introduced. In Section 3, the materials and data sets used in this study are described. In Section 4, we provide details for the concrete methodologies used in our proposed model, including the classifiers, the wavelet denoising techniques, and the feature importance. Section 5 gives details for the adaptive feature selection schema. Section 6 lists the results of the experiments and demonstrations of the effectiveness of the proposed methods. Section 7 draws some conclusions and discusses the future research directions.

2. Related work

Many researchers consider using technical indicators as input features to predict stock price or price movement direction, and have achieved accurate results in short-term forecasting. For example, Zhang et al. (2018) used technical indicators as input features in their stock price trend prediction system. The results showed that their system has certain advantages in accuracy and average return per transaction. Au Haq et al. proposed an extended set of forty-four technical indicators from daily stock data of eighty-eight stocks. These technical indicators were used as input features to the model after their feature selection technology (Haq et al., 2021). The results showed that their model works well. As Bustos and Pomares-Quimbaya, (2020) reviewed the studies on stock market movement prediction and concluded that ensemble learning is becoming more and more popular. The ensemble learning models show high predictive ability and in some work, outperform artificial neural networks and support vector machines. Nevertheless, the performance of the ensemble learning model largely depends on the quantity, quality, and combination of input features (Guan et al., 2014).

Therefore, enhancing the quality of the input features is an important research target. Denoising is a direct research solution, and several studies also show that the denoising process can significantly improve the performance of the stock price prediction. In a study by Lu (2010), an integrated independent component analysis (ICA)-based denoising scheme with the neural network was proposed for stock price prediction. Experimental results showed that the performance of the forecasting model was improved by using denoised data as the input of the neural network model. Similarly, in the study of Li and Tam, (2017), they proposed a novel model to combine real-time wavelet denoising functions with the LSTM to predict the East Asian stock indexes. The empirical results revealed that their proposed prediction model performance displays significant improvements compared to those of the original LSTM model without utilizing the wavelet denoising function. Yan and Ouyang (2018) used wavelet analysis to remove noise in financial time-series data and improved the model generalization ability. Among the denoising methods we've investigated, wavelet analysis has been shown to be an effective denoising scheme (Rhif et al., 2019a).

However, in stock prediction work, the denoising process is often used to process the data before participating in the training of machine learning (Song et al., 2021; Wu et al., 2021), little attention is paid to the noise in the data before the calculation of technical indicators. In this paper, we argue that the preprocessing of the technical indicators may also affect the prediction task. Because if we take the stock price

Table 1
The part of SSEC, HSI, DJI, S&P 500 data sets.

Data	Date	Open	High	Low	Close	Volume
SSEC	2021–03-01	3531.48	3552.57	3511.99	3551.40	31,548,752,600
	2021–03-02	3566.85	3566.85	3485.36	3508.60	33,983,048,600
	2021–03-03	3500.15	3577.62	3498.72	3576.90	34,765,684,600
	2021–03-04	3546.64	3552.20	3487.38	3503.49	39,361,612,000
	2021–03-05	3463.31	3523.57	3456.67	3501.99	35,640,922,300
HSI	2021–03-01	29457.89	29550.75	29195.97	29452.57	2,629,062,100
	2021–03-02	29708.39	29765.96	28957.31	29095.86	2,895,849,600
	2021–03-03	29249.43	29912.00	29183.56	29880.42	3,228,618,000
	2021–03-04	29525.48	29597.16	29102.10	29236.79	2,957,909,000
	2021–03-05	28667.14	29397.27	28513.13	29098.29	3,996,713,300
DJI	2021–03-01	29457.89	29550.75	29195.97	29452.57	2,629,062,100
	2021–03-02	29708.39	29765.96	28957.31	29095.86	2,895,849,600
	2021–03-03	29249.43	29912.00	29183.59	29880.42	3,228,618,000
	2021–03-04	29525.48	29597.16	29102.10	29236.79	2,957,909,000
	2021–03-05	28667.14	29397.27	28513.13	29098.29	3,996,713,300
S&P 500	2021–03-01	3842.51	3914.50	3842.51	3901.82	5,071,540,000
	2021–03-02	3903.64	3906.41	3868.57	3870.29	5,493,690,000
	2021–03-03	3863.99	3874.47	3818.86	3819.72	6,150,790,000
	2021–03-04	3818.53	3843.67	3723.34	3768.47	7,142,240,000
	2021–03-05	3793.58	3851.69	3730.19	3841.94	6,842,570,000

Table 2
Technical indicators in this study.

Indicator	Description	Indicator	Description
CP	Close Price	ROC	Rate Of Change
MACD	Moving Average Convergence/Divergence	VAR	Variance
RSI	Relative Strength Index	DEMA	Double Exponential Moving Average
FASTK	Stochastic Fast K	ATR	Average True Range
FASTD	Stochastic Fast D	BETA	Beta
ULTISC	Ultimate Oscillator	ADX	Average Directional Movement Index
PRICE_C	Price Change	CCI	Commodity Channel Index
TSF	Time Series Forecast	OBV	On Balance Volume
VAR	Variance	WR	Williams %R

information as a set of time series with high noise and high volatility, the technical indicators calculated from the original data may also be mixed with noise, affecting the model performance. Thus, in this study, we introduce the wavelet analysis and test a variety of wavelet basis functions to process stock price data. Then we use the denoised stock price information to calculate technical indicators. In this way, technical indicators become more effective.

Besides the improved indicators as the features, another critical step in data preprocessing is to find proper feature combinations. Although improved indicators have been established as features, different combinations of features still greatly influence the prediction results. There may be redundant or irrelevant features in the feature set. Removing those less relevant features can reduce the amount of calculation, thus improving model performance. An effective way to solve this problem is feature selection, which has been proven to be an effective way to enhance model performance in many studies (Chandrashekar and Sahin, 2014). In the study of Kou et al. (2021), they proposed a two-stage multi-objective feature selection method that optimizes the number of features and model classification performance. In another study, a modified differential evolution (DE) algorithm was used to perform feature selection for cardiovascular disease. The results showed that the accuracy of the proposed hybrid model is 83%, which is higher than that of some other existing models (Vivekanandan et al., 2017).

Among the feature selection methods, feature importance as a quantitative indicator seems to be a promising one. It is computed for evaluating the extent of features' influence on the model performance, thus helping to choose essential ones. In the study of Verma et al. (2020), a feature importance method was utilized to select the 15 most salient features in prediction. However, these feature selection methods are only applicable to a given feature set. This article attempts to create feature sets of different time dimensions according to the size-varied time windows, and then selects the best feature subset of each feature set. Therefore, the adaptive feature selection schema is needed.

3. Data description and preprocessing

In this study, we use the data set from four stock markets, SSEC (SSE Composite Index) of China stock market, HSI (Hang Seng Index) of Hong Kong stock market, S&P 500 (Standard & Poor's 500 Index), and DJI (Dow Jones Industrial Average) of US stock market. These data sets represent four types of markets. The Chinese market is an immature market. The Hong Kong market is a semi-mature market. The American market is a mature market, and one of the earliest stock markets. Each data set includes the open price, high price, low price, close price, and volume from January 4, 2005, to March 8, 2021. Part of the four data sets is shown in Table 1.

Many studies have shown that technical indicators are very effective features for stock forecasting (Basak et al., 2019; Haq et al., 2021; Patel et al., 2015a). We chose 18 indicators in our study, which are shown in Table 2.

Normalization is unnecessary in ensemble learning methods because the scale of the data does not affect the final result. In addition, stock forecasting is a classification task in this study, the target to be predicted is calculated as follows:

$$target_{after\_i} = \begin{cases} 1, (close_i - close)/close > 0 \\ -1, (close_i - close)/close < 0 \end{cases} \tag{1}$$

where $target_{after\_i}$ is the label after i days, $close$ is the close price, $close_i$ is the close price on the i th day.

In the experiment, in chronological order, the first 80% of the data was the training set, and the last 20% of the data was the test set. For example, in the SSEC data set, there were 3863 data points in total after calculation of technical indicators and data cleaning. Among those the first 3090 data points were used for training and the remaining 773 data points were used for testing. The same was true for the other three data sets. We set up size-varied time windows, and extracted the information of each day contained in the time window as a lagged variable. All the information in the time window forms a data point. The features contained in each data point are given in the following formula:

$$DayPoint = \{Indicator_1, Indicator_2, ..., Indicator_n\}, \quad n = Indicatornumber \tag{2}$$

Table 3
Performance index of dichotomy.

Indicators	Formula	Meaning
Accuracy	$A = \frac{TP+TN}{T+F}$	Correctly predict the proportion of samples in all samples
Precision	$P = \frac{TP}{TP+FP}$	The percentage of a sample that is predicted to be positive is also positive
Recall	$R = \frac{TP}{TP+FN}$	The percentage of all positive samples that were correctly predicted
F1 score	$\frac{2}{F1} = \frac{1}{P} + \frac{1}{R}$	Harmonic mean of Precision and recall

TP: If an instance is a positive class and is predicted to be a positive class, it is True Positive.
FP: If an instance is a positive class but is predicted to be a Negative class, it is False Negative.
TN: If an instance is a negative class but is predicted to be a positive class, it is False Positive.
FN: If an instance is Negative, but is predicted to be Negative, it is True Negative.

Table 4
The performance of different wavelet basis functions on four data sets.

	Sym4	Db4	Coif4	Haar
SSEC	**65.14**	65.86	65.28	83.77
HSI	401.76	398.16	**388.45**	498.34
DJI	**228.41**	244.06	245.26	289.40
S&P 500	**25.94**	27.10	27.75	32.68

$$DataPoint = \{DayPoint_i\}, i = 1, 2, 3, ..., d, d = TimeWindow \qquad (3)$$

Where *DayPoint* is all features of one day, *DataPoint* is a data point. The time window *d* is varied as 3, 5, 10, 15, 30, 45 and 60 days.

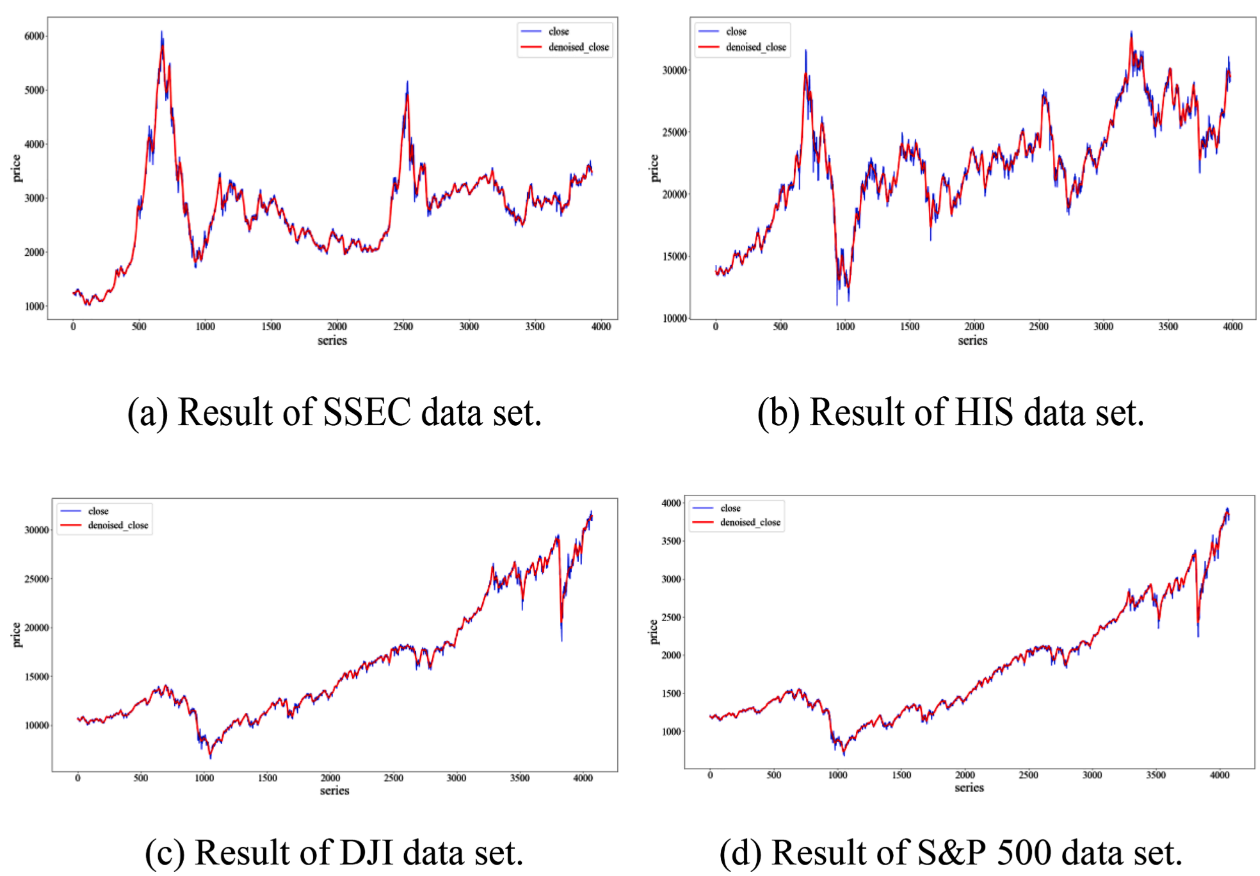
4. Methodology

4.1. Wavelet transform (WT)

WT has been widely used in image, speech, and signal processing. Many studies have shown that wavelet analysis is an effective denoising approach (Bruce et al., 2006; Kompella et al., 2016; Martínez and Gilabert, 2009; Masset, 2015; Patel et al., 2015a). One of the main

Table 5
Performance of model for the data before and after denoising.

	Indicators	Before Denoise	After Denoise
SSEC	Accuracy	0.583	**0.732**
	Precision	0.583	**0.734**
	Recall	0.583	**0.731**
	F1 score	0.580	**0.731**
HSI	Accuracy	0.575	**0.780**
	Precision	0.555	**0.780**
	Recall	0.573	**0.780**
	F1 score	0.551	**0.780**
DJI	Accuracy	0.568	**0.745**
	Precision	0.555	**0.741**
	Recall	0.568	**0.746**
	F1 score	0.551	**0.741**
S&P 500	Accuracy	0.580	**0.766**
	Precision	0.571	**0.761**
	Recall	0.580	**0.766**
	F1 score	0.574	**0.762**



(a) Result of SSEC data set.
(b) Result of HIS data set.
(c) Result of DJI data set.
(d) Result of S&P 500 data set.

Fig. 1. The results of wavelet denoising using the optimal wavelet basis function on the close price of four data sets.

Table 6
The model performance of different forecast target in the SSEC data set.

forecast target	Accuracy	Precision	Recall	F1 score
0	0.643	0.641	0.641	0.641
1	0.700	0.701	0.701	0.701
2	0.714	0.712	0.712	0.712
3	0.733	0.739	0.734	0.735
4	0.731	0.729	0.731	0.731
5	0.728	0.738	0.729	0.729
6	0.735	0.744	0.734	0.736
7	0.736	0.741	0.736	0.736
8	0.735	0.740	0.736	0.736
9	0.736	0.742	0.737	0.739

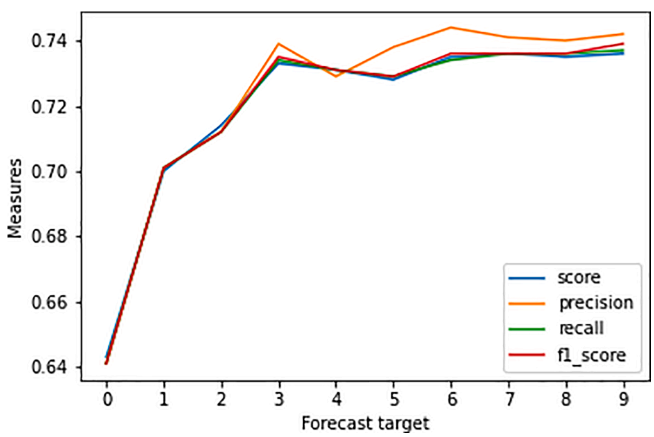


Fig. 2. The result of the model performance changing with the forecast target in SSEC data set.

Table 7
The number of features before and after feature selection of size-varied time windows in four data sets.

Time window	Feature number	Feature number after feature selection 1			
		SSEC	HSI	DJI	S&P 500
0	18	18	18	18	18
3	72	30	43	35	42
5	108	42	37	39	30
10	198	25	31	34	32
15	288	28	35	55	51
30	558	42	44	51	55
45	828	38	32	69	63
60	1098	53	44	65	49

advantages of WT is that it decomposes a signal according to the frequency and represents the signal in the time domain. As for the wavelet transformation, both the time and the frequency information from the signal are retained. It is thus a more powerful transformation for time–frequency analysis (Rhif et al., 2019b).

WT can decompose signals on different time scales. The purpose of WT is to translate and shrink the wavelet basis function. The wavelet basis function is defined as:

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi(\frac{t-b}{a}), a > 0, b \in R. \tag{4}$$

where a is a scale parameter and b is a time center parameter. When a and b change continuously, the wavelet transform process is continuous wavelet transform (CWT). For a signal $f(t)$, CWT is defined as:

$$W_f(a,b) = \int_{-\infty}^{\infty} f(t)\psi^*_{a,b}(t)dt = \frac{1}{\sqrt{a}}\int_{-\infty}^{\infty}\psi^*_{a,b}(\frac{t-b}{a})f(t)dt \tag{5}$$

Table 8
SSEC data set feature subset and the corresponding model performance.

	Feature number in the feature subsets	Accuracy	Precision	Recall	F1 score
0	18	0.725	0.729	0.725	0.724
1	17	0.724	0.725	0.725	0.725
2	16	0.725	0.729	0.725	0.724
3	15	0.725	0.730	0.725	0.725
4	14	0.727	0.730	0.729	0.729
5	13	0.730	0.730	0.730	0.730
6	12	0.728	0.729	0.726	0.726
7	11	0.729	0.732	0.730	0.730
8	10	**0.733**	**0.732**	**0.732**	**0.732**
9	9	0.728	0.729	0.726	0.726
10	8	0.729	0.732	0.730	0.730
11	7	0.730	0.730	0.730	0.730
12	6	0.730	0.730	0.730	0.730
13	5	0.727	0.728	0.726	0.726
14	4	0.719	0.719	0.718	0.718
15	3	0.707	0.706	0.706	0.706
16	2	0.683	0.683	0.683	0.683
17	1	0.627	0.630	0.630	0.630

Table 9
HSI data set feature subset and the corresponding model performance.

	Feature number in the feature subsets	Accuracy	Precision	Recall	F1 score
0	18	0.763	0.764	0.764	0.764
1	17	0.766	0.766	0.766	0.766
2	16	0.769	0.769	0.769	0.769
3	15	0.766	0.768	0.768	0.768
4	14	0.772	0.771	0.771	0.771
5	13	0.775	0.774	0.774	0.774
6	12	0.780	0.781	0.781	0.781
7	11	0.784	0.784	0.784	0.784
8	10	0.779	0.780	0.780	0.780
9	9	0.783	0.782	0.782	0.782
10	8	0.777	0.776	0.776	0.776
11	7	0.778	0.779	0.779	0.779
12	6	**0.786**	**0.785**	**0.785**	**0.785**
13	5	0.769	0.770	0.770	0.770
14	4	0.767	0.768	0.768	0.768
15	3	0.751	0.751	0.751	0.751
16	2	0.728	0.729	0.729	0.729
17	1	0.661	0.660	0.660	0.660

Table 10
DJI data set feature subset and the corresponding model performance.

	Feature number in the feature subsets	Accuracy	Precision	Recall	F1 score
0	18	0.721	0.729	0.722	0.703
1	17	0.717	0.734	0.716	0.696
2	16	0.730	0.738	0.729	0.715
3	15	0.741	0.745	0.741	0.732
4	14	0.735	0.745	0.736	0.720
5	13	0.728	0.740	0.727	0.710
6	12	0.728	0.739	0.728	0.712
7	11	0.728	0.739	0.729	0.714
8	10	0.724	0.733	0.726	0.710
9	9	0.737	0.738	0.738	0.728
10	8	**0.756**	**0.754**	**0.754**	**0.752**
11	7	0.747	0.746	0.748	0.745
12	6	0.740	0.740	0.741	0.740
13	5	0.742	0.740	0.741	0.736
14	4	0.731	0.730	0.731	0.730
15	3	0.736	0.735	0.736	0.735
16	2	0.695	0.694	0.694	0.694
17	1	0.606	0.610	0.610	0.610

Table 11
S&P 500 data set feature subset and the corresponding model performance.

	Feature number in the feature subsets	Accuracy	Precision	Recall	F1 score
0	18	0.740	0.736	0.743	0.725
1	17	0.739	0.738	0.739	0.722
2	16	0.739	0.738	0.740	0.726
3	15	0.743	0.739	0.742	0.729
4	14	0.745	0.741	0.745	0.732
5	13	0.742	0.739	0.744	0.728
6	12	0.747	0.745	0.749	0.732
7	11	0.750	0.749	0.749	0.738
8	10	0.759	0.755	0.759	0.75
9	9	0.756	0.754	0.756	0.748
10	8	0.755	0.752	0.755	0.745
11	7	0.749	0.746	0.750	0.735
12	6	0.753	0.752	0.752	0.740
13	5	**0.773**	**0.770**	**0.774**	**0.768**
14	4	0.767	0.762	0.766	0.764
15	3	0.766	0.765	0.768	0.765
16	2	0.727	0.726	0.728	0.726
17	1	0.682	0.690	0.680	0.680

where $W_f(a,b)$ is the wavelet transform coefficient, $\psi^*_{a,b}$ is the conjugate function of the wavelet basis function. By continuously changing the scale parameter a and time center parameter b, the function $W_f(a,b)$ can select the different positions of the signal and analyze the change of signal at different scales. However, this increases the complexity of calculation, which leads to the difficulty of application and implementation. Discrete wavelet transforms (DWT) is an effective substitute for CWT.

The DWT discretizes the proportional scale a and time center parameter b of the CWT according to the power of two, and particularly suitable for sampling values. The DWT decomposes the signal into multiple orthogonal wavelet sets, and each set being a time series describing the coefficients of the signal in the corresponding frequency band changing with time. Suppose that after discretization $a = a_0^j, b = ka_0^j b_0, k, j \in Z$,

then the $WT_f(j,k)$ of DWT is defined as:

$$\psi_{j,k}(t) = a^{-\frac{j}{2}}\psi(a^{-j}t - kb_0) \tag{6}$$

$$WT_f(j,k) = \int f(t)\psi^*_{j,k}(t)dt \tag{7}$$

Where the $\psi_{j,k}(t)$ is wavelet transform function, $\psi^*_{j,k}$ is the conjugate function of $\psi_{j,k}(t)$. As shown in the above formula, the basis of DWT is the Mallat algorithm. It is a signal decomposition method, which obtains high-frequency or low-frequency signals of the sequence by scaling and contracting the proportional parameter a. Specifically, according to the unnatural decomposition level N, the input signal is decomposed into low-frequency and high-frequency signals. The low-frequency component of each layer is decomposed to get the low-frequency signal and high-frequency signal again. N scales' low frequency and high-frequency component information is finally obtained until the number of decomposition layers reaches N. For stock price data, the low-frequency component reflects the overall trend of the sequence, while the high-frequency component contains random noise information. In order to realize the denoising function, the constraint or zeroing of high-frequency components can effectively eliminate noise to smooth the sequence.

4.2. Random forest

Random forest is a machine learning algorithm based on the decision tree. In the decision tree classification algorithm, the leaf nodes are split recursively until a leaf node containing only a single classification appears. Random forest is an extended variant of bagging. Its idea is to establish multiple decision trees, generate multiple training sets by bootstrap methods, and create a decision tree for each training set. Only some features are randomly selected rather than all features. Random forest is a special bagging method because it randomly samples the data set and features simultaneously. In the classification task, the final decision result is the voting result of all decision trees, while in the regression task, the final decision result is the mean of all decision tree results. Many researchers have used tree-based algorithms to predict stock prices or stock trends with success. In summary, the algorithm steps of random forest are as follows:

1 Select N samples from the sample set as a training set by bootstrap.
2 Generate a decision tree with the sample set obtained by sampling. In each node generated, N features are randomly selected and not repeated, and the sample set is divided by using these N features to find the best partition feature (Gini coefficient, gain rate or information gain can be used to judge).
3 Repeat steps 1 to 2 for K times, where K is the number of decision trees in the random forest.
4 The random forest obtained from training is used to predict the test samples, and the result of prediction is determined by the voting method.

4.3. Feature importance

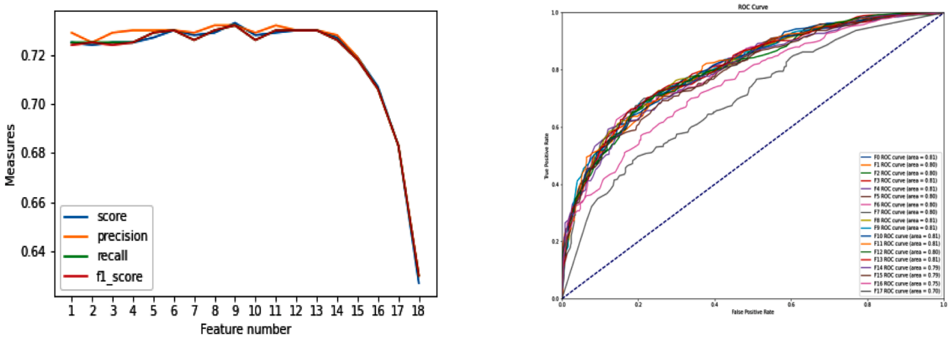
After training a model, in addition to being interested in the prediction results of the model, we usually want to know which features in the model are more important and which features have the greatest impacts on the forecasting. In the field of stock forecasting, higher forecasting accuracy means higher returns, and feature selection has been proven to be an effective method to improve forecasting performance. When performing stock prediction tasks, investors can choose technical indicators participating in the training according to the importance of their characteristics, delete technical indicators having a negative impact on the prediction model, and retain technical indicators having a positive impact on the prediction model. Feature importance describes the relative importance of each feature in the feature set. When training models with random forest algorithms, there are usually three ways to get feature importance scores: Gini importance, permutation importance, and SHAP importance.

Gini importance is to measure the importance of the feature by calculating the amount of impurity reduction in the node through the Gini index. The amount of impurity reduction in each feature in the random forest can be averaged, and the features are ranked accordingly. Each feature is randomly ranked based on the permutation importance, and the change of model performance is calculated. The feature that affects the performance the most is the most important one. SHAP importance is derived from the marginal benefits of individuals in cooperative games. That is, the importance of a feature is calculated by calculating the contribution of a single feature in the feature set in the model. Gini importance and SHAP importance mainly focus on solving the problem of interpretability of the model, while the permutation importance focuses more on the impact of features on the model.

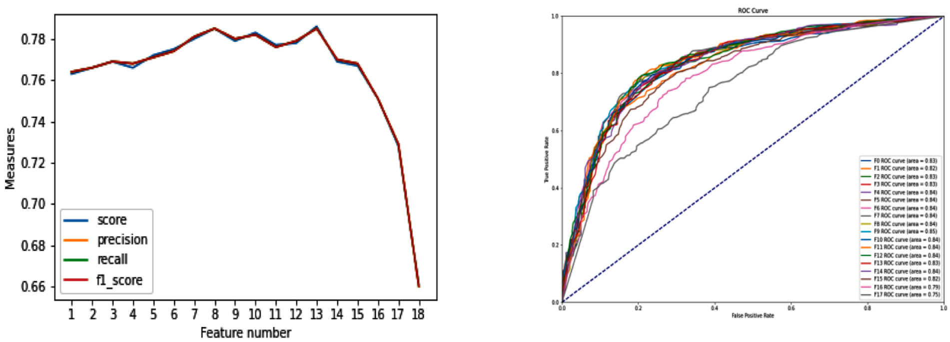
The purpose of this research on the importance of features is to study the impact of features on the performance of the model by selectively removing features, thereby improving the performance of the model. Therefore, permutation importance is the optimal option for describing the influence of the features on the model performance.

5. Adaptive feature selection method

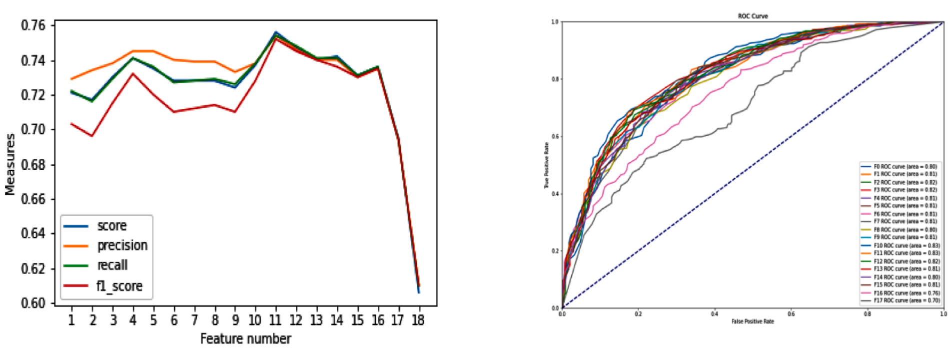
All features in the feature set can be sorted by calculating the permutation importance (PI). If feature selection is performed by deleting features one by one, the feature selection process can be time-consuming because of the large number of historical features. Moreover, the

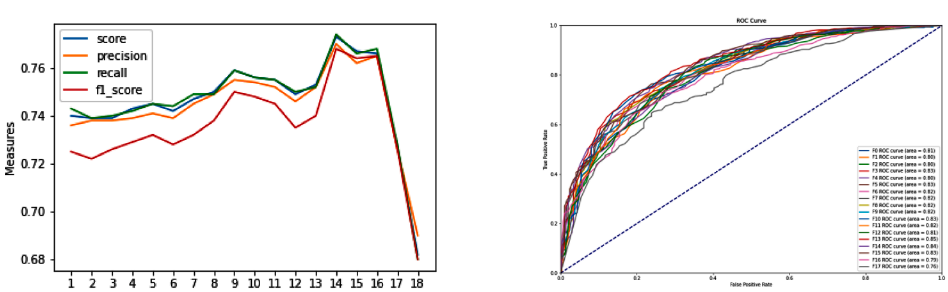


(a) Performance curve of SSEC data set with feature number.



(b) Performance curve of HSI data set with feature number.





(d) Performance curve of S&P 500 data set with feature number.

Fig. 3. Performance curve with feature number.

Table 12
Performance of model and feature number of the best feature subset of size-varied time windows in four data sets.

DATA	Time window	Accuracy	Precision	Recall	F1 score
SSEC	0	0.733	0.732	0.732	0.732
	3	0.745	0.744	0.744	0.744
	5	0.750	0.752	0.751	0.751
	10	0.753	0.755	0.752	0.752
	15	**0.753**	**0.756**	**0.754**	**0.754**
	30	0.751	0.753	0.753	0.753
	45	0.743	0.743	0.743	0.743
	60	0.742	0.741	0.741	0.741
HSI	0	0.786	0.785	0.785	0.785
	3	0.790	0.790	0.790	0.790
	5	0.791	0.791	0.791	0.791
	10	0.788	0.789	0.789	0.789
	15	0.791	0.791	0.791	0.791
	30	0.789	0.789	0.789	0.789
	45	**0.793**	**0.794**	**0.794**	**0.794**
	60	0.790	0.790	0.790	0.790
DJI	0	0.756	0.754	0.754	0.752
	3	0.767	0.768	0.769	0.764
	5	0.784	0.783	0.784	0.783
	10	0.784	0.784	0.788	0.784
	15	**0.791**	**0.789**	**0.790**	**0.789**
	30	0.773	0.771	0.774	0.771
	45	0.784	0.784	0.784	0.784
	60	0.781	0.779	0.780	0.780
S&P 500	0	0.773	0.770	0.774	0.768
	3	0.779	0.778	0.777	0.772
	5	0.801	0.805	0.809	0.805
	10	0.818	0.815	0.816	0.814
	15	0.821	0.820	0.820	0.820
	30	**0.822**	**0.821**	**0.821**	**0.821**
	45	0.821	0.821	0.821	0.821
	60	0.819	0.819	0.819	0.819

ranking reliability of the permutation importance values of features obtained by single training is low. Therefore, we propose an improved feature selection method called the adaptive feature selection method, which has two stages. The implementation steps are as follows:

First stage:

a) The original feature set is used as the input training random forest model, and the PI value of each feature in the feature set is calculated by using the trained model and sorted according to the PI value from small to large, getting the sorted feature set $F = \{f1, f2, \ldots\ldots fn\}$.

b) Repeat a) process n times to get n sorted feature sets.

c) Count the number of times k that each feature ranked in the top $R \times 100\%$ of the total ranking, where R is calculated by the formula as follow:

$$R = \frac{0.15e}{e^{0.015/e}} \times n_{Timewindow} + \frac{0.2e}{e^{0.35}} \times n_{Timewindow} - 0.001 \times n_{Timewindow} \tag{8}$$

Where the $n_{Timewindow}$ is the number of time window.

d) Traverse the total feature set F, in which the features that meet the conditions of $k > K$ form a new feature set $F^{'}$, which is much less than F.

Second stage:

a) The new feature set $F^{'}$ is used as the input training random forest model, and the PI value of each feature in the feature set is calculated by using the trained model (Calculate multiple times to take the average). Sort by PI value from small to large to get the sorted feature set $F_s^{'}$.
b) Take out the feature with the lowest PI value from $F_s^{'}$ to get a new feature set $F_n^{'}$. Replace $F^{'}$ with $F_n^{'}$.
c) Repeat procedures a) and b) until the feature set is empty, and get the set T of the feature subset and the accuracy set S corresponding to the feature subset.
d) The feature subset corresponding to the highest accuracy in set S is the best feature subset. End of feature selection.

6. Experiment and analysis

6.1. Model performance

In this study, the stock movement direction forecasting task was treated as a dichotomy problem. The common dichotomy model evaluation indexes were used to evaluate the performance of the model, including accuracy, precision, recall rate, and F1 score. Their meanings are shown in Table 3.

Another useful indicator is the Receiver Operating Characteristic (ROC) curve. Each point on the ROC curve reflects the sensitivity to the same signal stimulus.

6.2. Experimental result of denoising

Firstly, four basic functions, Sym4, Db4, Coif4, and Haar, are used to conduct denoising experiments on data from different sources, and MES values were used to evaluate the denoising results. The performances of different wavelet basis functions on four data sets are shown in Table 4. The table shows that Sym4 exerted the best effect on the three data sets SSEC, DJI, S&P 500, while COIF4 had the best effect on HSI. The wavelet basis function with the best effect in each data set was selected for denoising in the following experiment. The trend before and after wavelet denoising is shown in Fig. 1.

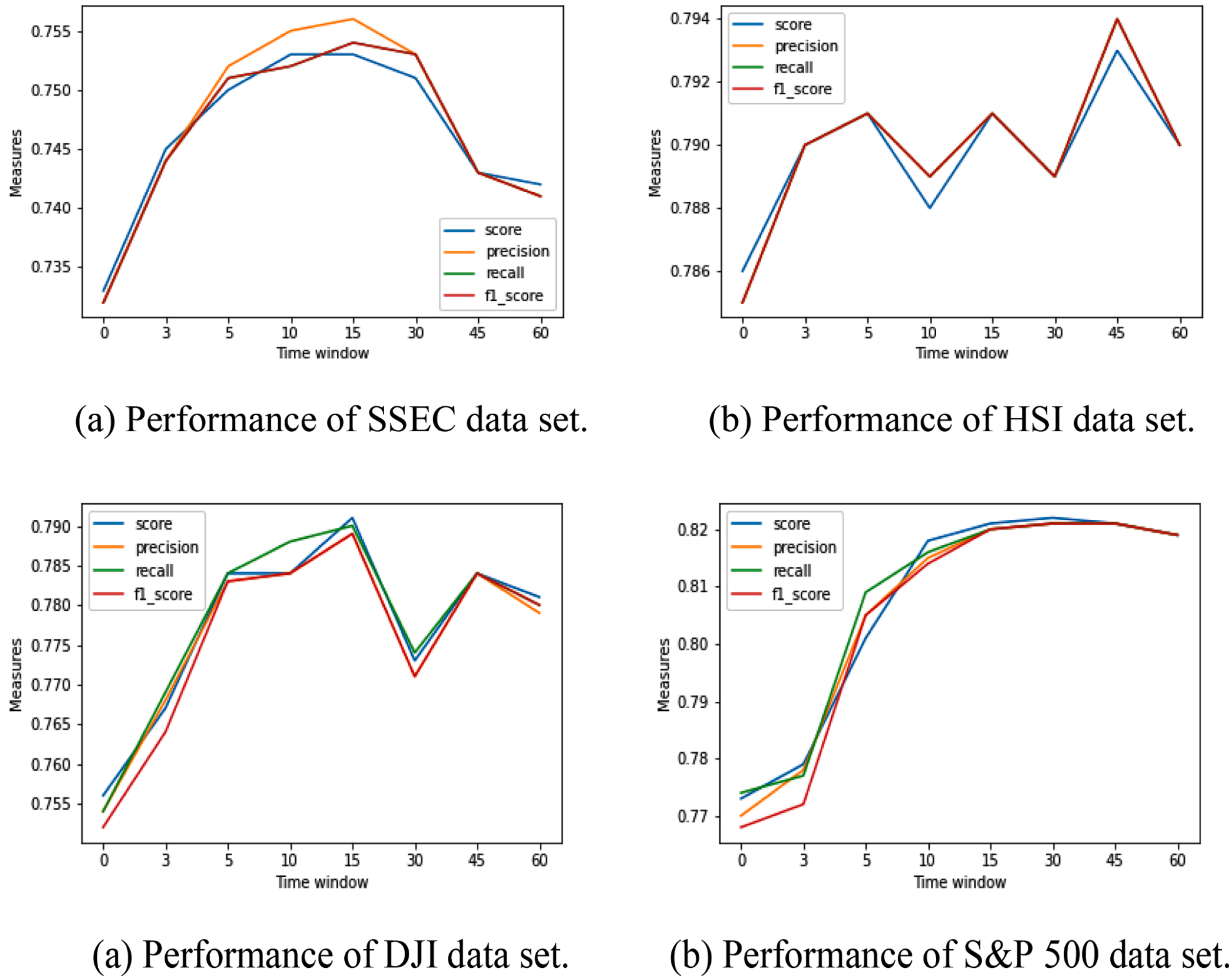
Based on the conclusions of the previous experiment, we used wavelet analysis to process the original data. We calculated the technical indicators with the stock closing prices before and after wavelet denoising, and obtained the technical indicators before and after the improvement. Furthermore, the impact of improved technical indicators on the performance of the forecasting model has been investigated. We took the data from the improved and the unimproved technical indicators to predict the fluctuations after three days. The results are shown in Table 5.

As can be seen from the results in the table, compared with the SSEC data set, the accuracy and precision of the model after wavelet denoising were improved by 25.56% and 25.90%, respectively, and the recall rate and F1 score were improved by 25.39% and 34.48%. On the HSI data set, the accuracy is improved by 34.95%, the precision is improved by 40.54%, the recall rate and the F1 score were improved by 36.12% and 41.56%, respectively, Meanwhile on the DJI data set, the accuracy was improved by 31.16%, the precision is improved by 31.51%, the recall rate was improved by 31.34%, and the F1 score was improved by 34.48%. On the S&P 500 data set, the accuracy and precision were increased by 32.07% and 33.27%, respectively. The recall rate was improved by 32.07%, and the F1 score is improved by 32.75%. Overall, these results indicate that using the wavelet analysis method to denoise stock closing price data, and obtaining improved technical indicators as features can greatly improve the performance of the model. Furthermore, this method was valid for stock data from different sources, among which the HSI from Hong Kong showed the best effect.

6.3. Experimental result of forecast target

The forecast target of all experiments in this article is the movement direction of stock price after three days. We attempt to reduce the impact of emergencies on stock prices, because research shows that emergencies will have a greater impact on stock prices in the first three days (Mittal and Goel, 2012). Taking the SSEC data set as an example, we conducted a confirmation experiment and explained the results. We used the improved indicators of the day to predict the direction of stock movements for 0, 1, 2, 3, 4, 5, 6, 7, 8, and 9 days from the current time. Considering the randomness of a single experiment caused by various factors like the initial weights settings of the model, we conducted multiple experiments and took the average. The results are shown in Table 6 and Fig. 2.

As can be seen from the results in the table and the figure, taking the



(a) Performance of SSEC data set. (b) Performance of HSI data set.

(a) Performance of DJI data set. (b) Performance of S&P 500 data set.

Fig. 4. Performance curve of four data sets with time windows.

F1 score as an example, the growth rate of the first three data was relatively large, the average of which is 0.031, while the growth rate of the fourth dropped significantly, with an average of 0.001. When the forecast target date was greater than four, as the forecast target date was delayed, the risk increased greatly. This implies that holding the stock for a long time may cause huge losses. In addition, the delayed forecast of the target date may also encounter greater volatility occurring during this period, causing greater risk. Therefore, even if the delayed forecast target may slightly increase the F1 score, the subsequent improvement was insignificant relative to the introduced risk. In summary, the forecast target of this article was determined to be the direction of stock price movement after three days.

6.4. *Experimental result of feature selection*

In this subsection, we discuss the feature sets built from the size-varied time windows, and the design of a feature selection method with two stages. We selected the best feature subset of each feature set to study the relationship between the performance of the prediction model and the size of the time window. When the time window was set to 3, 5, 10, 15, 30, 45, 60, the number of features increased rapidly as the size of the time window increased. This led to many redundant features in the feature set established according to the time window, which significantly increases the computation workload.

The first stage of the feature selection method proposed in this paper was used to process the feature sets built by size-varied time windows. The results are shown in Table 7. As can be seen from the results in the table, the number of features in each feature set was reduced to the appropriate size and the features that are critical to the model were retained by the first stage of the feature selection method. Our method successfully reduces the number of features in feature sets of different orders of magnitude to the same order of magnitude, greatly saving computational resources. In addition, this method is a rough feature selection method, so the features in the selected feature subset are not fixed. The results in the table are from multiple experiments and retain the features with a high probability of occurrence. This will not affect the selection of the best feature subset because the retained features are more important than the deleted features, and the best feature subset is always selected from the more important features.

In the second stage of feature selection, the feature subset selected in the first stage is used as an input feature to select the best feature subsets. This feature subset was used for a training session, and the PI value of each feature were calculated according to the trained model, sorted from small to large. The feature with the smallest PI value was taken out without putting it back to form a new feature subset, and used for training. The performance index of the corresponding model was calculated. This operation was repeated until the feature set was empty. The results are shown in Table 8, Table 9, Table 10, and Table 11. The performance curve with the number of features is shown in Fig. 3. In addition, we use ROC curves shown in Fig. 3 to assist in illustrating the effect of feature selection.

It can be seen from the results in the table and figure that our method is effective in selecting the optimal feature set. Model performance was improved over no feature selection, but the effect was different on the

Table 13
The size of the time window and the corresponding features in the best feature subset.

DATA	Time window	Features in the best feature subset
SSEC	0	price_c, DEMA, VAR, CCI, fastd, rsi14, ULTISC, ROC, WR, fastk
	3	macd_3, WR_2, VAR, CCI, fastd_1, ROC, ULTISC, WR, fastk
	5	fastd_4, ROC_5, rsi14_3, macd_2, DEMA_3, fastk_5, WR_1, WR_3, ULTISC_1, CCI_5, fastd_5, CCI, ULTISC, ROC, WR, fastk
	10	WR_3, rsi14_4, VAR_5, fastd_7, WR_5, CCI_5, ROC, ULTISC_7, CCI, ULTISC, WR, fastk
	15	ROC_5, ROC_4, fastd_7, ULTISC_7, ROC, ULTISC, CCI, WR, fastk
	30	VAR_7, WR_22, WR_10, ULTISC_10, ROC, CCI, ULTISC_9, ULTISC, WR, fastk
	45	ROC_11, ROC, WR_13, fastd_15, ADX_34, WR_6, CCI_1, CCI, CCI_5, WR, fastk
	60	WR_9, BETA_38, ULTISC_36, VAR_2, CCI_1, fastk_22, WR_1, VAR_59, ULTISC, ULTISC_10, CCI, ROC, WR, fastk
HSI	0	Close, BETA, VAR, ADX, macd, ROC, price_c, CCI, ULTISC, WR, fastk
	3	Close, ROC_3, ROC_2, ADX_1, ROC_1, fastd_3, BETA_1, price_c, CCI, ROC, WR, ULTISC, fastk
	5	BETA_2, ROC_4, fastd_2, VAR_5, price_c, ROC, WR, ULTISC, fastk
	10	rsi14_5, ULTISC_10, Close_8, price_c_2, ROC, price_c, ULTISC_7, CCI, WR, ULTISC, fastk,
	15	CCI_1, WR_15, ROC_2, ULTISC_7, ULTISC_9, ADX_10, CCI_9, ATR_10, price_c, ULTISC_11, CCI, ROC, ULTISC, WR, fastk
	30	ADX_7, BETA_29, ULTISC_5, ROC, CCI, ULTISC, WR, fastk
	45	CCI_37, ROC_1, fastd_3, CCI_43, ULTISC_12, ULTISC_22, ROC, WR, ULTISC, fastk
	60	fastd_37, CCI, CCI_43, ROC_47, VAR_40, ROC_39, fastk_27, ROC, ULTISC_21, ROC_3, fastk_3, ULTISC, WR, fastk
DJI	0	ADX, price_c, ULTISC, fastd, VAR, CCI, fastk, WR
	3	fastk_3, fastd, rsi14_2, fastk_1, macd_3, price_c, rsi14_3, BETA, BETA_1, ADX, WR_1, ULTISC_2, WR_3, fastk_2, CCI, ROC, VAR, ULTISC, WR, fast
	5	fastk_1, ULTISC_3, volume_4, ULTISC, WR_3, volume_3, fastk_2, volume_1, VAR_5, VAR_3, ULTISC_4, CCI_1, WR_4, BETA_4, fastd, WR_5, ROC, price_c, ULTISC_5, VAR, CCI, WR, fastk
	10	ROC_3, VAR_5, ROC_1, volume_4, price_c_5, WR_6, WR_5, VAR_1, fastd, WR_3, VAR_8, macd_2, rsi14_1, WR_8, CCI_1, price_c_6, fastd_5, rsi14_4, ULTISC, CCI, WR, fastk
	15	fastd, WR_1, ROC_11, price_c_13, fastd_5, WR_2, fastk_3, CCI_11, ULTISC, WR_4, ROC_15, CCI_13, CCI_7, fastd_6, ROC_9, ULTISC_14, fastd_15, CCI, WR, fastk
	30	rsi14_3, rsi14_10, volume_17, macd_22, fastk_20, ADX_3, fastd_4, ULTISC, fastk_17, CCI_7, fastk_3, CCI, fastk, WR
	45	ROC_9, price_c_6, VAR_41, WR_7, rsi14_5, ULTISC_4, WR_17, fastd_33, VAR_26, ULTISC, fastk_2, ROC_10, fastd, rsi14_9, CCI_1, price_c, VAR, ROC, WR, fastk
	60	rsi14, WR_54, fastd_27, ULTISC_41, fastd, CCI_13, ULTISC_36, ULTISC_12, WR_4, fastk_2, ULTISC, ULTISC_6, price_c_6, fastd_3, WR_3, ROC, WR, fastk
S&P 500	0	price_c, VAR, ADX, ROC, CCI, ULTISC, fastk, WR,
	3	ULTISC_2, VAR, WR_1, CCI, price_c, ULTISC, ROC, fastk, WR
	5	VAR, WR_5, WR_4, ROC, ULTISC_2, CCI, fastk, WR
	10	ATR_10, VAR_9, BETA_9, volume_8, BETA_4, ROC_9, VAR_1, volume_2, WR_4, CCI_1, volume_5, VAR, ULTISC_1, CCI_7, fastk_6, macd, fastk_1, WR_6, fastd, CCI_9, ROC_4, ROC, CCI, WR, fastk
	15	volume_1, volume_9, price_c_13, fastk_2, VAR_10, rsi14_4, ROC_1, fastk_1, fastd_2, BETA_8, CCI_1, CCI_8, price_c, VAR, ROC_4, ULTISC, fastd, CCI_9, CCI, ROC, ULTISC_4, CCI_10, WR, fastk
	30	VAR_28, ULTISC, fastk_27, CCI_16, ADX, rsi14_2, volume, VAR_22, VAR, ROC_27, rsi14_7, fastk_1, VAR_4, fastd, ROC_3, ROC, CCI_10, CCI, WR, fastk
	45	rsi14_32, ADX_29, fastk_20, volume_26, ULTISC_3, WR_3, fastk_18, CCI_39, ULTISC_4, VAR_5, fastk_1, ADX_3, fastk_2, CCI_18, ROC_1, BETA_29, CCI_11, rsi14_4, CCI_10, fastd, ROC, CCI, WR, fastk
	60	rsi14_3, ULTISC_3, WR_5, ULTISC, fastk_2, ROC, CCI fastk, WR

four data sets. Specifically, taking the F1 score as an example, there was a 1.0% improvement on the SSEC data set, a 2.7% improvement on the HSI data set, a 7.0% improvement on the DJI data set, and a 5.9% improvement on the S&P 500 data set. We can see from the resulting figures that the performance curves of the SSEC and HSI data sets show insignificant changes, while the performance curves of the DJI and S&P 500 data sets exhibit relatively obvious changes. Such differences suggest that these statistically-based technical indicators are more consequential to mature markets. In summary, our method performs better in mature markets. In addition, we can also obtain the most important features by analyzing the features in the best feature subset, which will enhance the interpretability of the work in this article. We will discuss this in the next experimental results.

In the following experiments, we use improved technical indicators as features, and the method proposed in this article to obtain the best feature subsets of each feature set built by size-varied time windows. The size of the time window was set to 3, 5, 10, 15, 30, 45, and 60, respectively. The forecast target was the direction of the stock price movement after three days. The best performance for each time window and the corresponding best feature subsets are shown in Table 12. The relationship between model performance and time window is shown in Fig. 4.

In order to enhance the reliability of the experimental results, all experiments carried out many times, and the results averaged. As can be seen from the results in the table, compared with using only the current day's data, the highest F1 score on the SSEC data set increased to 0.754, with 3.00% improvement, and the highest F1 score on the HSI data set increased to 0.794, with 1.14% improvement, the highest F1 score on the DJI data set increased to 0.789, with 4.92% improvement, and the highest F1 score on the S&P 500 data set increased to 0.821, with 6.90% improvement. This indicates that in the prediction task discussed in this article, adding past technical indicators as features can better the model performance. Furthermore, observing the curve of the performance change with the size of the time window, we can infer that as the size of the time window increases, the model performance first improves slightly and then stabilizes, or decreases slightly. This implies that the longer the interval, the smaller the positive impact of historical technical indicators. We continue to investigate Table 13 to verify this finding.

Table 13 shows the best feature subsets corresponding to the data sets built by size-varied time windows after feature selection. We can see from the table that although the features in each best feature subset are different, some features are always left in the best feature subset. They are 'WR', 'fastk', 'CCI' 'ULTISC', and 'ROC', which are the most important features for the model. As the size of the time window increased, some of the technical indicators in the first 15 days had a positive impact on the model performance. However, the earlier technical indicators had almost no effect on the model performance, so they were deleted in the feature selection process. Taking the best feature subset of 60-days-sized time window as an example, technical indicators based on data of the first 15 days in the 60-day time window were deleted by the feature selection process. In addition, although some historical technical indicators were left in the best feature subset, they did not have a positive effect from the perspective of model performance. For example, in the experiment with the S&P 500 data set, when the time window size was 45, there are more features in the best feature subset than the best feature subset when the time window size was 60, even though the model performances show a slight difference. This also confirms one of our previous conclusions that the longer the interval, the smaller the impact of historical technical indicators. Nevertheless, this exposes the limitations of the feature selection proposed in this article, which is, the best feature subset selected by the method in this article is not fixed, but a locally optimal solution within a certain range. This is because the importance of feature used in this article is permutation importance, which is a method with inconsistent results.

To sum up, when the size of the time window is within a certain range, the performance of the model can be improved through feature selection. As the time window continues to expand, model performance will reach its peak, and start to decrease afterward. The performances of the experimental results on the four data sets are consistent.

7. Conclusion

Stock price movement forecasting has always been an important research topic as accurate prediction can bring investors high returns. In this work, the price movement problem is converted to a binary classification problem of rising or falling stock prices. A method for improving technical indicators based on wavelet denoising and a method for feature selection based on improved indicators are proposed.

To be more specific, two critical issues have been addressed in this research. The first is the noise in the price data. In the field of stock market forecasting, most of the existing denoising research focuses on the data directly involved in the machine learning training stage, but little attention has been paid to the noise in the data before the calculation of technical indicators. To attend to this situation in this paper, the improved technical indicators are generated based on the denoised price data set, which has greatly improved the performance of the model. We test the improved technical indicators obtained on four different stock markets. The results demonstrate that the improved technical indicators could effectively improve the model performance.

The second issue is feature selection. This paper designs an adaptive feature selection method based on the permutation importance, obtaining the best feature subset from feature sets containing features of different magnitudes. We tested our method on four real data sets. The experimental results suggest that the method can obtain the optimal local solution of the best feature subset, and it is applicable to data sets with different characteristics and size-varied time windows. In addition, we compare the effects of size-varied time windows on the model performance and demonstrate through experiments that proper time window sizes exert a positive impact on the model performance.

Overall, the method in this paper remarkably improves the prediction accuracy. This method can be applied to the trend prediction of stocks to bring higher returns. The technical indicator improvement method and feature selection method proposed in this paper can also be used to further extend traditional investment methods. Moreover, it can also be applied to various other fields where machine learning techniques and data science are used.

There are some research directions worth trying in the future. In this paper, we only used 18 common technical indicators as the input features, while more information can be introduced. For example, macroeconomic variables such as monetary policy, exchange rate, unemployment rate, etc., and fundamental indicators such as market value, price-earnings ratio, profit growth rate, and other factors can greatly affect the stock in the long run. These features can be introduced to model long-term trading in the stock market. In contrast, news across the market may affect the stock market sentiment in the short term. Employing market sentiment as one of the input features may also positively impact the performance of the model.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work is supported by The National Natural Science Foundation of China (no. 71904064) and The Natural Science Foundation of Jiangsu Province (no. BK20190580) partially supports this research. The research is also supported by the 111 Project and the Fundamental Research Funds for the Central Universities (Grant No: JUSRP11922).

References

Ballings, M., Van den Poel, D., Hespeels, N., & Gryp, R. (2015). Evaluating multiple classifiers for stock price direction prediction. *Expert Systems with Applications, 42* (20), 7046–7056. https://doi.org/10.1016/j.eswa.2015.05.013.

Bartiromo, R. (2004). Non linear behaviour of stock market volatility. Retrieved October 01, 2004, from https://ui.adsabs.harvard.edu/abs/2004cond.mat.10768B cond-mat/0410768.

Basak, S., Kar, S., Saha, S., Khaidem, L., & Dey, S. R. (2019). Predicting the direction of stock market prices using tree-based classifiers. *The North American Journal of Economics and Finance, 47*, 552–567. https://doi.org/10.1016/j.najef.2018.06.013.

Bollerslev, T., Marrone, J., Xu, L., & Zhou, H. (2014). Stock Return Predictability and Variance Risk Premia: Statistical Inference and International Evidence. *The Journal of Financial and Quantitative Analysis, 49*(3), 633–661. http://www.jstor.org/stable/43303962.

Bruce, L. M., Mathur, A., & Byrd, J. J. D. (2006). Denoising and Wavelet-Based Feature Extraction of MODIS Multi-Temporal Vegetation Signatures. *GIScience & Remote Sensing, 43*(1), 67–77. https://doi.org/10.2747/1548-1603.43.1.67

Bustos, O., & Pomares-Quimbaya, A. (2020). Stock market movement forecast: A Systematic review. *Expert Systems with Applications, 156*, Article 113464. https://doi.org/10.1016/j.eswa.2020.113464.

Chandrashekar, G., & Sahin, F. (2014). A survey on feature selection methods. *Computers & Electrical Engineering, 40*(1), 16–28. https://doi.org/10.1016/j.compeleceng.2013.11.024.

Dinesh, S., N. R, R., Anusha, S. P., & S, R. (2021, 2-4 April 2021). Prediction of Trends in Stock Market using Moving Averages and Machine Learning. 2021 6th International Conference for Convergence in Technology (I2CT).

Fama, E. F. (1965). Random Walks in Stock Market Prices. *Financial Analysts Journal, 21* (5), 55–59. http://www.jstor.org/stable/4469865.

Farias Nazário, R. T., e Silva, J. L., Sobreiro, V. A., & Kimura, H. (2017). A literature review of technical analysis on stock markets. *The Quarterly Review of Economics and Finance*, *66*, 115-126. https://doi.org/https://doi.org/10.1016/j.qref.2017.01.014.

Fama, E. F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance, 25*(2), 383–417. https://doi.org/10.2307/2325486

Ferreira, M. A., & Santa-Clara, P. (2011). Forecasting stock market returns: The sum of the parts is more than the whole. *Journal of Financial Economics, 100*(3), 514–537. https://doi.org/10.1016/j.jfineco.2011.02.003.

Guan, D., Yuan, W., Lee, Y.-K., Najeebullah, K., & Rasel, M. K. (2014). A Review of Ensemble Learning Based Feature Selection. *IETE Technical Review, 31*(3), 190–198. https://doi.org/10.1080/02564602.2014.906859

Haq, A. U., Zeb, A., Lei, Z., & Zhang, D. (2021). Forecasting daily stock trend using multi-filter feature selection and deep learning. *Expert Systems with Applications, 168*, Article 114444. https://doi.org/10.1016/j.eswa.2020.114444.

Jiang, J., Ma, K., & Cai, X. (2007). Non-linear characteristics and long-range correlations in Asian stock markets. *Physica A: Statistical Mechanics and its Applications, 378*(2), 399–407. https://doi.org/10.1016/j.physa.2006.12.011.

Jiang, W. (2021). Applications of deep learning in stock market prediction: Recent progress. *Expert Systems with Applications, 184*, Article 115537. https://doi.org/10.1016/j.eswa.2021.115537.

Kompella, K. C. D., Mannam, V. G. R., & Rayapudi, S. R. (2016). DWT based bearing fault detection in induction motor using noise cancellation. *Journal of Electrical Systems and Information Technology, 3*(3), 411–427. https://doi.org/10.1016/j.jesit.2016.07.002.

Kou, G., Xu, Y., Peng, Y., Shen, F., Chen, Y., Chang, K., & Kou, S. (2021). Bankruptcy prediction for SMEs using transactional data and two-stage multiobjective feature selection. *Decision Support Systems, 140*, Article 113429. https://doi.org/10.1016/j.dss.2020.113429.

Li, Z., & Tam, V. (2017, 27 Nov.-1 Dec. 2017). Combining the real-time wavelet denoising and long-short-term-memory neural network for predicting stock indexes. 2017 IEEE Symposium Series on Computational Intelligence (SSCI).

Long, J., Chen, Z., He, W., Wu, T., & Ren, J. (2020). An integrated framework of deep learning and knowledge graph for prediction of stock price trend: An application in Chinese stock exchange market. *Applied Soft Computing, 91*, Article 106205. https://doi.org/10.1016/j.asoc.2020.106205.

Lu, C.-J. (2010). Integrating independent component analysis-based denoising scheme with neural network for stock price prediction. *Expert Systems with Applications, 37* (10), 7056–7064. https://doi.org/10.1016/j.eswa.2010.03.012.

Martínez, B., & Gilabert, M. A. (2009). Vegetation dynamics from NDVI time series analysis using the wavelet transform. *Remote Sensing of Environment, 113*(9), 1823–1842. https://doi.org/10.1016/j.rse.2009.04.016.

Masini, R. P., Medeiros, M. C., & Mendes, E. F. (2020). Machine Learning Advances for Time Series Forecasting. arXiv:2012.12802. Retrieved December 01, 2020, from https://ui.adsabs.harvard.edu/abs/2020arXiv201212802M.

Masset, P. (2015). *Analysis of financial time series using wavelet methods. In*. New York: Springer-Verlag.

Mittal, A., & Goel, A. (2012). Stock prediction using twitter sentiment analysis. *Standford University, CS229 (2011 http://cs229. stanford. edu/proj2011/GoelMittal-StockMarketPredictionUsingTwitterSentimentAnalysis. pdf), 15*.

Naik, N., & Mohan, B. R. (2019, 2019//). Optimal Feature Selection of Technical Indicator and Stock Prediction Using Machine Learning Technique. Emerging Technologies in Computer Engineering: Microservices in Big Data Analytics, Singapore.

Patel, J., Shah, S., Thakkar, P., & Kotecha, K. (2015a). Predicting stock and stock price index movement using Trend Deterministic Data Preparation and machine learning techniques. *Expert Systems with Applications, 42*(1), 259–268. https://doi.org/10.1016/j.eswa.2014.07.040.

Patel, J., Shah, S., Thakkar, P., & Kotecha, K. (2015b). Predicting stock market index using fusion of machine learning techniques. *Expert Systems with Applications, 42*(4), 2162–2172. https://doi.org/10.1016/j.eswa.2014.10.031.

Phan, D. H. B., Sharma, S. S., & Narayan, P. K. (2015). Stock return forecasting: Some new evidence. *International Review of Financial Analysis, 40*, 38–51. https://doi.org/10.1016/j.irfa.2015.05.002.

Rhif, M., Ben Abbes, A., Farah, I. R., Martínez, B., & Sang, Y. (2019a). Wavelet Transform Application for/in Non-Stationary Time-Series Analysis: A Review. *Applied Sciences, 9*(7), 1345. https://www.mdpi.com/2076-3417/9/7/1345.

Rhif, M., Ben Abbes, A., Farah, I. R., Martínez, B., & Sang, Y. (2019b). Wavelet Transform Application for/in Non-Stationary Time-Series Analysis: A Review. *Applied Sciences, 9*(7). https://doi.org/10.3390/app9071345

Shah, D., Isah, H., & Zulkernine, F. (2019). Stock Market Analysis: A Review and Taxonomy of Prediction Techniques. *International Journal of Financial Studies, 7*(2), 26. https://www.mdpi.com/2227-7072/7/2/26.

Song, D., Baek, A. M. C., & Kim, N. (2021). Forecasting Stock Market Indices Using Padding-Based Fourier Transform Denoising and Time Series Deep Learning Models. *IEEE Access, 9*, 83786–83796. https://doi.org/10.1109/ACCESS.2021.3086537

Verma, A. K., Pal, S., & Kumar, S. (2020). Prediction of Skin Disease Using Ensemble Data Mining Techniques and Feature Selection Method—a Comparative Study. *Applied Biochemistry and Biotechnology, 190*(2), 341–359. https://doi.org/10.1007/s12010-019-03093-z

Vivekanandan, T., & Sriman Narayana Iyengar, N. C. (2017). Optimal feature selection using a modified differential evolution algorithm and its effectiveness for prediction of heart disease. *Computers in Biology and Medicine*, *90*, 125-136. https://doi.org/https://doi.org/10.1016/j.compbiomed.2017.09.011.

Wu, D., Wang, X., & Wu, S. (2021). A Hybrid Method Based on Extreme Learning Machine and Wavelet Transform Denoising for Stock Prediction. *Entropy, 23*(4), 440. https://www.mdpi.com/1099-4300/23/4/440.

Yan, H., & Ouyang, H. (2018). Financial Time Series Prediction Based on Deep Learning. *Wireless Personal Communications, 102*(2), 683–700. https://doi.org/10.1007/s11277-017-5086-2

Zhang, J., Cui, S., Xu, Y., Li, Q., & Li, T. (2018). A novel data-driven stock price trend prediction system. *Expert Systems with Applications, 97*, 60–69. https://doi.org/10.1016/j.eswa.2017.12.026.