

## Research Article

# Stock Trend Prediction Algorithm Based on Deep Recurrent Neural Network

Ruochen Lu <sup>1</sup> and Muchao Lu<sup>2</sup>

<sup>1</sup>The University of Queensland, Brisbane 4075, Australia

<sup>2</sup>Taiyuan University of Technology, Taiyuan 030024, China

Correspondence should be addressed to Ruochen Lu; [ruochen.lu@uqconnect.edu.au](mailto:ruochen.lu@uqconnect.edu.au)

Received 2 August 2021; Revised 23 August 2021; Accepted 24 August 2021; Published 14 September 2021

Academic Editor: Yuanpeng Zhang

Copyright © 2021 Ruochen Lu and Muchao Lu. 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.

With the return of deep learning methods to the public eye, more and more scholars and industry researchers have tried to start exploring the possibility of neural networks to solve the problem, and some progress has been made. However, although neural networks have powerful function fitting ability, they are often criticized for their lack of explanatory power. Due to the large number of parameters and complex structure of neural network models, academics are unable to explain the predictive logic of most neural networks, test the significance of model parameters, and summarize the laws that humans can understand and use. Inspired by the technical analysis theory in the field of stock investment, this paper selects neural network models with different characteristics and extracts effective feature combinations from short-term stock price fluctuation data. In addition, on the basis of ensuring that the prediction effect of the model is not lower than that of the mainstream models, this paper uses the attention mechanism to further explore the predictive *K*-line patterns, which summarizes usable judgment experience for human researchers on the one hand and explains the prediction logic of the hybrid neural network on the other. Experiments show that the classification effect is better using this model, and the investor sentiment is obtained more accurately, and the accuracy rate can reach 85%, which lays the foundation for the establishment of the whole stock trend prediction model. In terms of explaining the prediction logic of the model, it is experimentally demonstrated that the *K*-line patterns mined using the attention mechanism have more significant predictive power than the general *K*-line patterns, and this result explains the prediction basis of the hybrid neural network.

## 1. Introduction

For investors, stocks meet different investment desires and investment needs, expand the range of investment options, expand investment channels, to some extent meet the possibility of investors to obtain the corresponding income, and to some extent enhance the flexibility and liquidity of capital [1]. If we look at the enterprise side, the stock can play an important role in the management and development of the joint-stock enterprise, which is conducive to the establishment and improvement of the self-restraint and self-development of the enterprise management mechanism [2]. For the country, stocks are also a great tool to counteract inflation. Stocks have three main characteristics: (1) nonre-

turnability: once sold, stocks cannot be returned to the company and cannot be requested to be refunded but can only be sold to a third party through the secondary market; (2) uncertainty of earnings: the profit and loss of stocks depend on the company's operation and the stock exchange market, both of which are uncertain and changeable, so investors need to take greater risks; and (3) speculative: the stock market fluctuates frequently and the market price is unstable. Market prices are unstable and speculative, so stocks are very risky [3]. There are many reasons that affect the volatility of stock prices, and these frequently changing factors cause stock market volatility [4]. The objective risks of the stock market can bring gains to investors and at the same time can cause economic losses and may also

negatively affect the operating conditions of shareholding companies and even bring side effects to the national economic construction.

These problems are inevitable, so stock trend prediction has become an issue of great concern for all parties [5]. The research of stock trend prediction has also become an applied research direction of financial big data, and many scholars have adopted deep neural network methods to predict the stock trend, which has become one of the popular research problems in the current academic field. In recent years, with the rapid development of computer technology, the deep neural network [6–8] has become the key research object nowadays, and the application field has been expanded and extended, including the financial information field [9]. At present, the financial market occupies a pivotal position in the whole economic system of the country, and stocks are an important component of the financial market, so buying stocks has become a popular financial management method nowadays, and a stock analysis software with various performance can meet the real needs of investors [10]. This paper designs and implements a stock trend prediction system by combining historical stock trading data, news data, and investor sentiment to recommend stocks and stock trend prediction services for investors to reduce or avoid investment risks, thus bringing investors relatively stable economic returns. Due to the continuous innovation of machine learning techniques [11–13], more and more researchers are turning to the use of machine learning techniques to analyze stock data and create well-performing methodological models to predict stock movements in the future. Stock market prediction models are built by learning from historical price data to predict future prices [14]. Common machine learning algorithms such as logistic regression, genetic algorithms, and support vector machines have been used with good results in forecasting [15]. With the rise of neural network technology, building deep neural networks to portray stock prices and predict stock movements has received a lot of attention, and some scholars have conducted in-depth research in this area [16]. In order to improve the accuracy of stock trend prediction, many improved algorithms and optimization strategies have appeared one after another and have been successfully applied in practice.

Compared with existing stock trend prediction models, the improved model used in this paper combines stock price trend, news information, and investor sentiment for prediction, not only using trading data in the stock market but also taking into account the influence of financial and political news and stock forum speech on the stock market. In this work, a framework for predicting stock price trends using financial news and sentiment dictionaries combines a two-stream gated recurrent unit for stock price trend prediction and a Stock2Vec embedding model trained on stock news and sentiment dictionaries. Firstly, we propose a plain Bayesian model based on the sentiment classification of stock forum speech and experimentally confirm that the classification effect is better when using the plain Bayesian classifier, which can obtain the investor sentiment more accurately and lay the foundation for the establishment of

the whole stock trend prediction model. Secondly, on top of the original LSTM, this model is constructed by a mixture of Bi LSTM [17] and CLSTM, with Bi LSTM extracting stock trading data and investor sentiment index-related features and CLSTM integrating and processing the contextual features of the news, and finally outputting the prediction results through the fully connected layer. In the experimental model, the stock trend is experimented with using the classification method [18–20], and the classification is obtained as the probability of the stock going up and the probability of the stock going down. The experiments use the CSI 300 stock data as the data set. The prediction effect is evaluated by accuracy and return, and the experimental results show that compared with a single LSTM prediction model, the proposed method has a certain improvement in the accuracy of stock trend prediction and can predict stock trend accurately and effectively to a certain extent. At the same time, a deep neural network stock trend prediction [21] system is designed and implemented, and the trained prediction model is uploaded to the stock prediction module. By analyzing the requirements of the stock trend prediction system and designing each functional module, the whole stock trend prediction system is completed after development and testing. Investors can make stock selection and investment with the help of this system.

## 2. Related Work

The researchers constructed a denoising hybrid stock price prediction model based on a decision tree, which first extracts relevant features from stock data, then uses the principal component analysis algorithm to dimensionally reduce the features with the decision tree algorithm, and the dimensionally reduced data [22] is fed into the fuzzy model to predict stock prices [23]. The researchers created a Bayesian neural network model that does not require preprocessing operations and cycle analysis of the data but simply feeds market prices and technical indicators into the prediction model, which is used to predict the future closing price of the stock [24]. The researchers constructed a support vector machine model with a genetic algorithm to optimize the data by dimensionality reduction and used a quantitative stock selection method to empirically analyze its stock selection performance and prediction accuracy in the short and medium to long term, respectively [25]. The researchers first used wavelet decomposition of stock price series to screen out the low- and high-frequency information in the nonstationary time series and then constructed an ARIMA model from the high-frequency information and fitted the SVM model to the low-frequency information to obtain better results [26].

The researchers performed feature construction based on relevant technical indicators and used data mining techniques for feature modelling [27]. The main method relies on the idea of maximizing returns and proposes a support vector machine for genetic parameter search with AUC values under the ROC curve, which solves the problem of poor availability of forecasting with traditional methods. The researchers used Tensor Flow, a Google AI learning

system, and thus built a multiperceptron MLP neural network model for predicting each closing price, comparing Tensor Flow with a traditional BP neural network [28] based on the stock price prediction problem [29]. The researchers used a different number of features and multivariate data preprocessing to study the effect of long- and short-term memory networks (LSTM networks) on the prediction results, and the LSTM models increased the accuracy of stock return prediction compared to random prediction methods [30]. Researchers use LSTM to predict stock prices and propose variable step integration methods and an improved MSE loss function with improved prediction performance, but the drawback is that no generalized optimal step range is derived.

Many factors can affect stock markets and cause market volatility, such as global economic conditions, domestic macroeconomic factors, and highly correlated foreign stock markets. However, most time series models use stock indices as the only factor for prediction and do not consider more variables, while the opposite can lead to better prediction results [31]. Therefore, some scholars have developed sentiment analysis models for predicting the correlation between investor sentiment and financial markets, and researchers have studied the relationship between public sentiment states obtained from Twitter and the Dow Jones Industrial Average (DJIA) [32]. Two sentiment analysis models were used to analyze the text content of daily Twitter feeds to obtain and analyze changes in the public's sentiment. Researchers' sentiments were classified as positive and negative, while GPOMS meticulously classified sentiments into six categories, including calm, alert, sure, vital, kind, and happy, with six different dimensions to measure sentiments [33]. Through the Granger causality test, a close relationship between public sentiment and the Dow Jones Average Index (DJIA) was found. Next, using a self-organizing fuzzy neural network model, the public sentiment time series was used as the independent variable, and the model was found to be effective in predicting the change of DJIA closing price, which can largely improve the accuracy of DJIA prediction [34].

The researchers propose a framework for predicting stock price trends using financial news and sentiment dictionaries, combining a two-stream gated recurrent unit (TGRU) for stock price trend prediction and a Stock2Vec embedding model trained on stock news and sentiment dictionaries [35]. The researchers predict stock trends through the impact of financial news on stock market sentiment, using more expressive features to represent the text and augmenting existing text mining methods by including market feedback as part of the feature selection process. Powerful feature selection can greatly improve classification accuracy when used in conjunction with complex feature types, and the approach allows the selection of semantically relevant features, thus reducing the problem of overfitting when applying machine learning methods. It can also be transferred to any other application area that provides textual information and corresponding effect data. In summary, many applications have been made by many scholars using neural network models for stock prediction analysis. But most of them have been analyzed for foreign stock markets,

but there is still very little research on the domestic stock market, which may be related to the fact that the domestic stock market was established late and many aspects are still not well developed. However, the Chinese stock market has grown to have more than 3,000 listed companies with a circulating market capitalization of tens of trillions of yuan, which occupies a significant share in the whole national financial system, so it is significant to study the domestic stock market. In this paper, we construct a long- and short-term memory neural network model to predict the domestic stock trend.

## 2.1. Deep Recursive-Based Stock Trend Prediction Model

*2.1.1. Factors Influencing Stock Price and Fundamental Analysis Method.* Like other countries in the world, the Chinese stock market is complex and dynamic. From the specific business environment of listed companies, the transparency of financial statements, and the psychological sentiment of investors to the national policies and regulations, unexpected news events, and the world economic situation, all of them have a certain impact on it. Because the stock market is affected by many factors, it is difficult to predict and determine the future trend of stocks, and there is a large risk and unknown. At the same time, a stock is only marketable security, which does not have real value itself. Stock price refers to the trading price of a stock in the securities market, and dividend income is obtained by buying and selling stocks. Stock prices are influenced by various economic and political factors and other external circumstances. Technical analysis is a basic forecasting method that determines the future movement of the stock market by analyzing and judging graphical charts and related technical indicators.

It is a method of making judgments about stock price changes and stock trends based on market behavior itself, combined with theories and methods related to psychology and statistics, and based on historical stock trading data such as existing prices, rates of change, and volume in the stock market, as well as combining investors' subjective judgments and analysis to find patterns. The actual trading history of a particular stock or "average" is usually recorded in graphical form, and then possible future trends are inferred from that history. Technical analysis is based on the theory that market behavior contains all information and is based on three theories: that all information is fair and open, that stock prices move along a trend, and that historical stock prices repeat themselves. It mainly includes the indicator,  $K$ -line, pattern, and wave methods. Because the stock system is highly nonlinear and stochastic, it requires the investor to analyze graphical movements and data tables to obtain forecasts, while understanding the role of relevant parameters and corrections. Therefore, this method is not applicable in today's increasingly large and complex stock market. In addition to low efficiency, difficulty, and overreliance on manual experience, it also has a series of problems such as poor integrity of stock content information, redundancy of feature data, low utilization of stock data, poor results, and poor generalization, which brings certain difficulties in predicting stock prices and makes it difficult to meet the needs

of market development. Fundamental analysis is a method to calculate the intrinsic value of a stock by looking at the basic economic factors that may affect the stock price, i.e., fundamentals.

Relevant factors considered include turnover, revenues and expenses, financial statements, the company's growth prospects, competitive factors facing the company, and the expected return on equity or assets for the industry. The purpose of this analysis is to determine a value for the stock that takes all of these potential factors into account. Investors rely on a variety of financial tools to determine a company's past, present, and future profitability in order to make investment decisions. Some of the key financial statements that investors rely on include balance sheets, cash flow statements, and income statements. Accurate financial statements are used to analyze the corresponding value of a company. In order to determine the accuracy of such statements, investors rely on independent auditors' reports to ensure the accuracy of the financial statements. The method is considered a long-term investment approach because it does not take into account short-term pricing and trading fluctuations. The method also involves some element of forward-looking expectations, and it may take some time to realize the intrinsic value, so forward-looking information needs to be evaluated before considering the use of this method. Meanwhile, the method requires a comprehensive analysis of both macro- and microinformation. Micro is difficult to guarantee access to real and valid financial information because of the asymmetry of corporate information, as shown in Figure 1. The specific parameters can be from the raw data collected by sensors, which are continuous data sequences generated by user movements over a period of time. Macroscopically, it is also difficult to predict the national plan support policies and key development industries. Therefore, the analysis method still needs to be improved in terms of accuracy and timeliness, and it relies too much on the ability of analysts, and the analysis method is relatively difficult to apply and has certain limitations, especially for ordinary stock market investors, with poor feasibility and universality. To sum up, the fundamental analysis method mainly analyzes the long-term trend of stocks, which is difficult to obtain, organize, categorize, and analyze information, requires high analytical ability of investors, and does not allow reliable and accurate short-term forecasting.

**2.1.2. Stock Trend Forecasting Model.** The operational status of a listed company can be reflected by news information about the company's stock. Based on the common feature that news information is highly correlated with stock movements, people often rely on news information to forecast stock movements. Due to the time-sensitive nature of news, the information in news articles has a short time effect on the stock market, and in general, recent news has a large impact on stock movements. In order to encode stock-related news information, data preprocessing work is first performed on the text to reduce the repetition rate and correct the comment data with wrong formatting to improve the integrity and quality of the comment data. Unreasonable cases such as noisy data, such as null values and special sym-

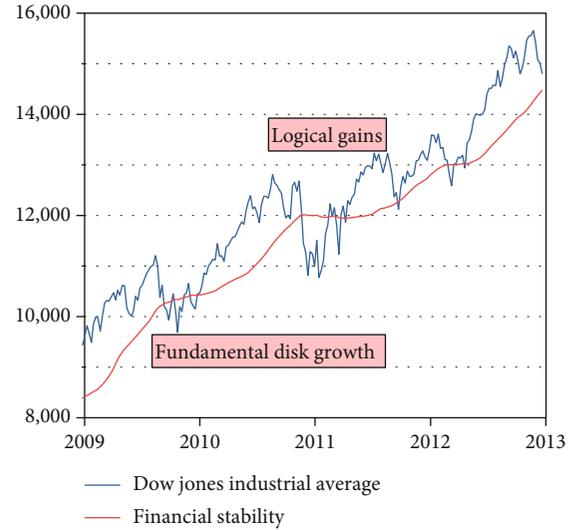


FIGURE 1: Correlation between stock price and financial factors.

bols, need to be deleted and processed. The authenticity and reliability of the research results are ensured by data preprocessing operations.

- (1) *Handling Missing Data.* The initial data obtained may suffer from data loss, which may be due to incorrectly unentered data, data that is inconsistent with data from other records and therefore deleted, data that was ignored at the time of entry, and unrecorded data changes. Missing data can be handled in a number of ways: by ignoring records, by manually filling in missing values or using global constants, or by using attribute means or most likely values through inference based on Bayesian formulas or decision trees. Missing values were observed when certain stocks or indices were recorded as null on certain dates, and to appropriately reduce the workload of this study, only stocks with up to ten null values were selected for processing, which included skipping entire records containing missing values, or filling missing values with the mean, or using inference (e.g., based on most similar instances). All null values appearing in this experiment are filled using the average of their closest left and right nonnull values, making the stock time series data complete.
- (2) *Data Noise Reduction.* Data contain a certain amount of noise due to the complexity of market dynamics. Noisy data are attributed to random errors or variations in measured variables, as well as errors in data collection tools, data entry errors, etc. The noise present in stock datasets can usually be classified into three main categories: duplicate records, inconsistent stock names, and incomplete files. Some files contain duplicate records. Therefore, all files are thoroughly checked to ensure that only unique records are available on the same date. When

a public company changes its name or a company merges, there may be inconsistencies in stock names. In this case, some of the data can use the old name while the rest of the data can use the new name. The standard is set in solving this inconsistency problem, and all the old names appearing in the records are modified in bulk with the updated records. Noise reduction is done by pywt library wavelet transform.

- (3) *Data Normalization*. Normalization is the “scaled down” transformation of data. If the value of one component of an attribute is too large, the other attributes may lose their moderation effect. For example, when stock price and trading volume are taken together as characteristic values, the difference between the two in terms of data volume is great because the stock trading volume is huge and can reach the level of billions, while the stock price is only a few tens or hundreds, but it cannot be shown that the impact on price is proportionally greater because the value of the trading volume is larger. The convergence of the model is often affected because the data scale is not consistent, the gradient is not “uniform,” the model is not stable when training, and the gradient descent algorithm does not easily converge to the optimal solution. Therefore, in this paper, the data are normalized. When normalization is performed, the size of all values is scaled to a fairly low value. The two most common methods of data normalization are min-max normalization and  $z$ -score normalization.

In order to improve the training efficiency of the samples and the generalizability of the training results, this paper uses min-max normalization to normalize the input feature sequences by which the data inputs are mapped to a predefined range  $[0, 1]$  or  $[-1, 1]$ . The min-max method normalizes the value of attribute  $A$  of the dataset according to the minimum and maximum values of the dataset. It converts the value  $a$  of attribute  $A$  to  $a$  in the range  $[\text{low}, \text{high}]$  by the following calculation:

$$a = \frac{\text{high} - \text{low}}{\max A - \min A} + \text{low}. \quad (1)$$

In particular, when low is set to 0 and high is set to 1, it is easy to see that  $a = 0$  when  $a = \min$  and  $a = 1$  when  $a = \max$ . This means that the minimum value in  $A$  maps to 0 and the maximum value in  $A$  maps to 1. Thus, the entire range of  $A$  values from the minimum to the maximum value maps between 0 and 1.

The data is then subjected to word separation processing. Since Chinese word separation techniques are relatively mature and are not the focus of this thesis, the existing jieba word separation technique, which is suitable for large-scale text separation scenarios, is used here. The TF-IDF method is a common word weighting measure that indicates the importance of a particular word in the whole document or

corpus, and the more frequently a word appears in a document to which it belongs, but hardly appears in all documents, the more it represents the key content. The importance of each word is determined by counting the total number of financial commentary documents, the number of document words, the number of times the word appears in a particular document, and the number of times the word appears in all documents. The formula for calculating the importance of a word in a given stock forum comment document is

$$T_f = \sum_{i=1}^n w_i + \text{ID}_F. \quad (2)$$

TF stands for word frequency and gives the frequency of words in each document in the corpus. It is calculated by the ratio of the number of occurrences of a word in a document to the total number of words in that document. It increases as the number of occurrences of the word in the document increases. Each document has its own TF. The formula is as follows:

$$T_f = \frac{n_{i,j}}{\sum_{i=1}^n w_i}. \quad (3)$$

$\text{ID}_F$  stands for inverse data frequency and is used to calculate the weights of rare words in all documents in the corpus. When  $T_f$  measures word frequency, the weight of words such as “of” or “and” must be reduced since they occur frequently in all documents, i.e., the document inversion frequency component. If a word appears frequently in each document, the less likely it is to be used as a keyword for a given document. The inverse document frequency of a word in a set of documents means how common or rare the word is in the entire set of documents. Thus, if the word is very common and occurs in many documents, the number will be close to 0. Otherwise, it will be close to 1. Designed to retain distinctive words as markers, words that occur rarely in the corpus have a high  $\text{ID}_F$  score. This metric can be calculated by dividing the total number of documents by the number of documents containing the word and then calculating the logarithm. It is given by the following formula:

$$\text{ID}_{F,n} = \log \left( \sum_{i=1}^n w_i \right). \quad (4)$$

Combining these two,  $\text{ID}_{F,n}$  extracts high-frequency words from the text as candidate keywords and the text inverse frequency  $\text{ID}_F$  weights the  $T_f$ - $\text{ID}_F$  scores of the words in the documents in the corpus  $w$ . Multiplying these two numbers will give the  $T_f$ - $\text{ID}_F$  scores of the words in the documents. The higher the score, the more relevant the word is in that particular document, i.e., the one with the higher weight is taken as the keyword:

$$w_{i,j} = T_{f,i,j} + \log \left( \sum_{i=0}^n w_i \right), \quad (5)$$

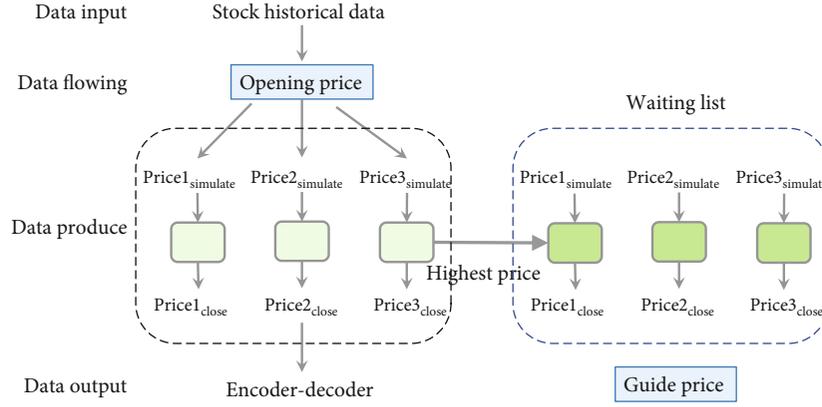


FIGURE 2: Deep recursive process.

where  $TF_{i,j}$  denotes the number of occurrences of  $i$  in  $j$  texts,  $df_i$  denotes the number of corpora containing  $i$  words, and  $N$  denotes the number of all corpora. Regarding the embedding layer, this model maps each word in a set  $K$  of size  $n$  to a corresponding word vector  $w$  by using the Word2Vec method. For the set of vectors  $w$ , the information is extracted using a fully connected neural network:

$$\text{Word}_2\text{Vec} = a\bar{w} + R_n, \quad (6)$$

where  $a$  is the weight size of each word vector,  $a$  belongs to  $R_n$ ,  $b$  is the weight size of the bias vector  $b$ ,  $b$  belongs to  $R_n$ , and  $\text{New}$  is the news information extraction result vector.

In order to correspond the stock names to the news data, each stock name is transformed into the corresponding stock vector  $S$  by embedding operation. By the Word2Vec method, it is transformed into the corresponding word vector  $c_v$ . The generated word vector is fed into CLSTM, and the stock name Name is used as Topic to process the stock news keyword Keys; in particular, the output matrix of CLSTM is used here as output to obtain the corresponding implied layer matrix information  $c_h$ :

$$c_h = \text{clastm}(\text{name}, \text{keys}). \quad (7)$$

**2.1.3. Deep Recursive Process.** The input layer of the model includes three major parts: stock historical data information, news information, and investor sentiment. The main part of the model firstly uses Bi LSTM to process the features in terms of data information separately, secondly uses news data as the contextual information input of CLSTM, and introduces the attention mechanism to give different attention to the news text sequences to ensure that the model captures information from the news that is more relevant to the stock price movement. Secondly, a plain Bayesian-based sentiment classifier is used to analyze the data from the forum, and the sentiment data obtained from it is combined with other parts of the data to serve as training data for the long- and short-term memory time series learning model. Finally, a multilayer fully connected neural network is used to process all the data. Satisfactory results can be obtained

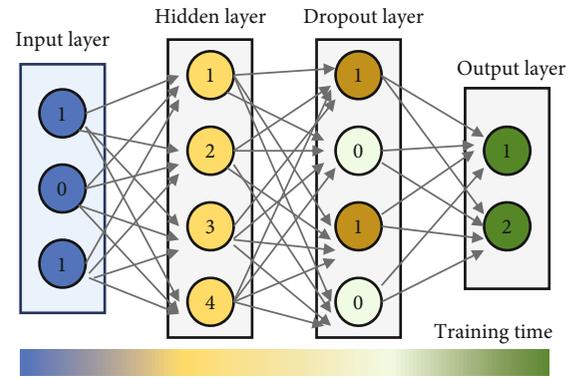


FIGURE 3: Schematic of adding dropout neural network.

by training the neural network, as shown in Figure 2. It also takes into account the effect of some randomness to avoid the occurrence of falling into a local minimum that leads to a global minimum not being reached.

The type of optimizer used optimizes the efficiency with which the algorithm converges to a minimum. The model is optimized with the Adam optimizer, which combines the advantages of both ADAGRAD and RMSPROP optimizers. The reason behind ADAGRAD is that infrequently used parameters must have a large learning rate, while frequently used parameters must have a small learning rate. The stochastic gradient descent update of ADAGRAD becomes

$$\beta_{t+1,i} = \beta_{t,i} - \eta, \quad (8)$$

$$\eta_{t,i} = \nabla J_{i,j} K. \quad (9)$$

The learning rate is derived by calculating the historical gradient of each parameter. Thus,

$$\beta_{t+1} - \beta_t = \frac{\theta}{\sqrt{Q + \chi}} h_t, \quad (10)$$

where  $H$  is the sum-of-squares matrix of the historical gradients. The problem with this optimization is that as

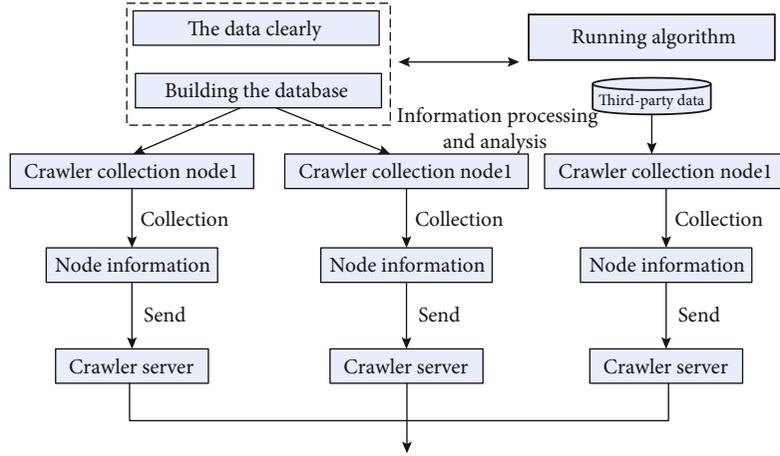


FIGURE 4: Web crawler flow.

the number of iterations increases, the learning rate begins to rapidly decrease and disappear. RMSprop slows down the decline in the learning rate by using a certain number of historical gradients. Updated to

$$\beta_{t+1} - \beta_t = \frac{\theta}{\sqrt{Q\sqrt{[t^2 - r]}}} h_t, \quad (11)$$

$$Q\sqrt{[t^2 - r]} = Q\sqrt{t} + 0.1t^2. \quad (12)$$

After using two hidden layer neural networks in this model, as shown in Figure 3, two hidden layer standard neural networks are used, and the dropout method is adopted to randomly discard some neurons, and the neurons marked with a fork in the right figure are randomly inactivated neurons. In practice, the dropout value is set too low, and the effect can be ignored; if the dropout value is set too high, it may lead to underfitting results. Therefore, the dropout is set to 20% in this model.

## 2.2. Experiments and Result Analysis

**2.2.1. Experimental Data.** Stock forums provide a platform for investors to discuss online. Stock forum sites allow users to request and exchange information therein. In addition, the stock forum site also allows users to view forum posts and post messages therein. When posting a message, users can create a new topic or post in an existing topic. Stock forums consist of user-generated content (UGC), and to get investor sentiment from a forum site, its content should be downloaded first. Use web crawler technology to crawl content and obtain data. A web crawler, an important part of current search engines, is a computer program or automated script that automatically crawls and downloads web information according to certain rules. In a narrow sense, it is usually considered a software program that traverses the information space of the World Wide Web based on web hyperlinks and web document retrieval methods (e.g., depth-first or breadth-first) using the standard HTTP protocol. Web crawlers obtain the content of the web pages corre-

sponding to each URL by determining the queue of URLs to be crawled, parse the web page content, and store the corresponding data. In order to efficiently and accurately obtain comment data, the crawler should start from the entry URL of the forum. Web crawlers are classified into several categories according to the system structure and technical mode, such as general-purpose web, focused web crawlers, incremental web crawlers, and deep web crawlers. In this paper, we use a general-purpose web crawler to crawl the stock forum comment data. It is very expansive, crawling objects involving URLs all the way to the entire Web, collecting data for web search engines and large web service providers. Due to the large scope and number of crawls, high requirements for crawl speed and storage space, and low requirements for the order of crawled pages, usually web crawlers work in parallel. Therefore, the application value is strong. The specific flow chart of the generic web crawler operation is shown in Figure 4.

**2.2.2. Evaluation Based on Accuracy and Yield.** The training performed on the above data is used to predict the stock trend of that day. After trying methods such as fitting stock movements and classifying stock movements, experiments were conducted using the classification method, which is a dichotomous classification (up/down) of stock movements to make predictions. The experiments show that reducing the problem to a classification problem is more accurate than fitting stock movements. To evaluate the impact of financial news on stock prices over time, we set different time intervals (i.e., 1, 2, 5, 7, and 10 days) for the prediction experiments. In the case of a one-day interval, it means that news affects stock prices within 24 hours. Similarly, the impact of news varies at other time intervals. The results are shown in Figure 5, where the accuracy obtained by the prediction model proposed in this paper decreases with time, with the highest accuracy of the experimental results in the first 24 hours and decreases with time. This also illustrates the impact of financial news and the rapid reflection of the stock market.

The experiment uses the above stock selection strategy to backtest the test set by selecting one hundred consecutive

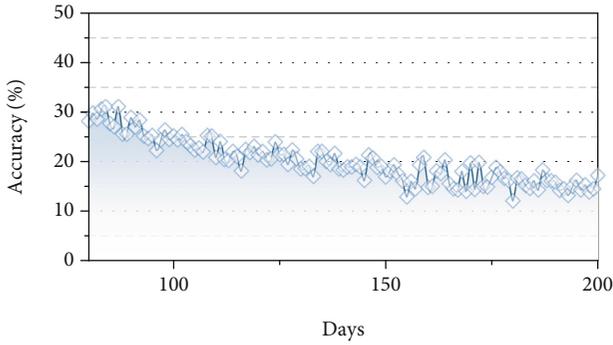


FIGURE 5: Accuracy vs. number of days line graph.

days of data and removing the cases where the data on the day of backtest is less than 100 shares and conducting the test using the return as a measure, the results of which are shown in Figure 6. Daily return represents the relationship between the return available for each investment and the number of days, where the horizontal coordinate indicates the number of days and the vertical coordinate indicates the return, and the points in the figure are the return for each investment with the dashed 0 coordinate, where the number of days with a return greater than 0 is  $G_{days}$ . The results indicate that the actual probability of a daily return greater than 0 is close to 0.7041.

Figure 7 shows the return versus time assuming that the investment strategy given by the model is adopted every day, where the horizontal coordinates indicate the number of days and the vertical coordinates indicate the return compared to day 0, where the coordinates of the highest point are (98, 0.2849), i.e., if the stock selection is done with this model, then at 98 days, a cumulative return equal to 28.49% of the principal can be obtained.

The number of input features of the network model is adjusted, and the same training data and test data are experimented with in a single LSTM model, a news-based LSTM model, and a news- and investor sentiment-based LSTM model, and the results are shown in Figure 8. In the training process of the model, the accuracy of stock trend prediction by LSTM with a different number of layers is compared by continuously modifying the LSTM layers. However, the number of layers is positively correlated with the computational redundancy and consumption, which is not conducive to the efficiency of the overall model. Therefore, with the proper number of network layers, increasing the number of network layers is costly and the improvement of prediction accuracy is not significant. Therefore, a two-layer LSTM network model is used to predict stock movements.

In this paper, stocks and their corresponding news are extracted to obtain information about possible stock trends from the news, such as national policy support that may lead to a rise in the stock. Since each news item is too long, thus, TF-IDF, a common weighting technique used for information retrieval and data mining, is used to process each news item and extract the news-related keywords as input. In order to align with numerical data, historical data from November 5, 2018, to November 8, 2019, were collected,

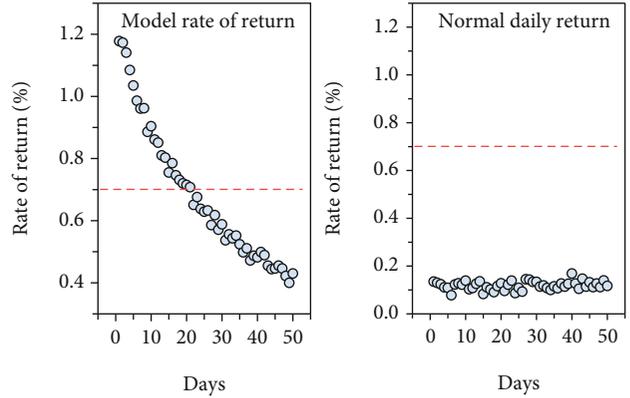


FIGURE 6: Daily yield.

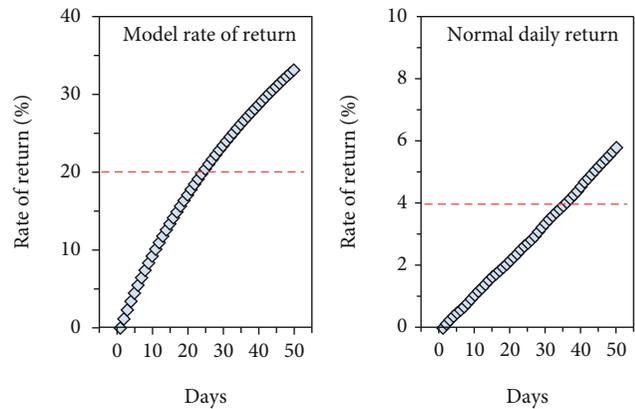


FIGURE 7: Total return.

along with the corresponding stock name information as input, and according to its news release time, 4:00 p.m. was used as the dividing line, with news released before 4:00 p.m. counted as financial news of the day and news released after 4:00 p.m. counted as the next trading day. Also, obtaining stock forum data relies on web crawlers. A plain Bayesian classifier trained using the method described in Section 3 is used to obtain investor sentiment for predicting stock movements. After the above experiments, we verified the conjecture that when the attention mechanism of the hybrid neural network assigns a weight to the  $K$ -line pattern of a sample that is significantly higher than the overall mean of the sample, it is highly likely that the  $K$ -line pattern carries a stronger predictive power than the general  $K$ -line pattern, and when the pattern appears, there is a high probability that the later market will appear to be consistent with the history. It is also true that  $K$ -line patterns are an intuitive analysis tool, and when they show a strong trading signal, human researchers will also focus on the signal to make predictions about the future of the market. Therefore, compared to other deep learning models, the hybrid neural network model proposed in this paper guarantees the predictive effect and performs the predictive inference work in a way that humans can intuitively feel and understand. In addition,

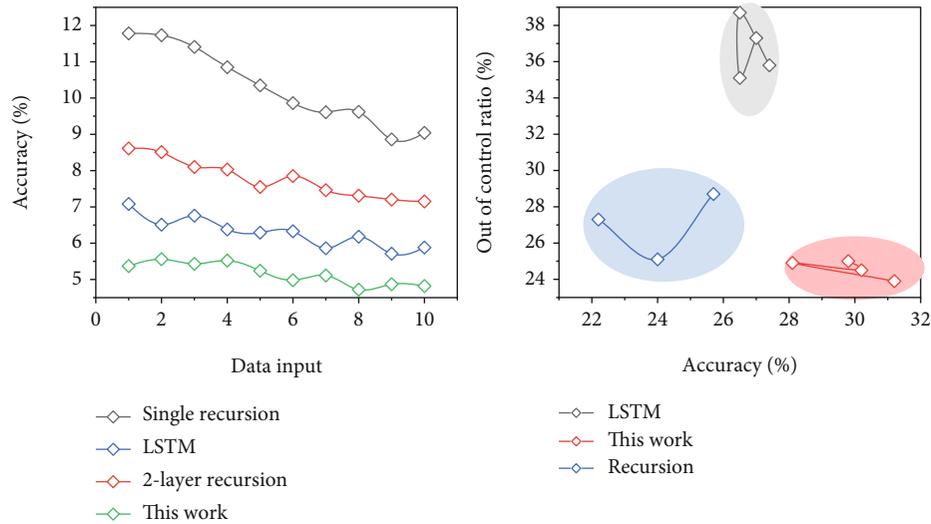


FIGURE 8: Comparison of the accuracy of different prediction models.

using the powerful computing power and storage capacity of computers, we can use the model to mine more  $K$ -line patterns that may have strong predictive power and continuously expand the knowledge base of human researchers.

### 3. Conclusion

This paper focuses on the problem of stock trend prediction based on deep learning. Due to the many factors affecting the stock movement, the number of stocks and the huge trading volume make this research challenging and difficult. In this paper, we consider multiple influencing factors at the same time, combining three aspects: historical stock trading data, news information, and investor sentiment index, and use an improved LSTM to model the prediction. We also construct a stock trend prediction system and apply the improved model to make stock recommendation and stock trend prediction for investors. It has high operability and practical value. In this paper, we classify the sentiment of stock forum remarks and calculate to obtain the investor sentiment index. Based on the sentiment dictionary, a stock market sentiment dictionary is formed by adding stock market specialized vocabulary. A sentiment classification model is constructed on the basis of the plain Bayesian algorithm to classify the sentiment of stock forum remarks. Experiments show that the classification effect is better using this model, and the investor sentiment is obtained more accurately, and the accuracy rate can reach 85%, which lays the foundation for the establishment of the whole stock trend prediction model. In this paper, we design and implement a deep neural network-based stock trend prediction system and upload the trained prediction model to the stock prediction module. Through the requirement analysis of the stock trend prediction system and the design of each functional module, the whole stock trend prediction system is completed after development and testing. It enables investors to make relatively correct investment decisions based on the stock recommendation results and stock prediction results of the

system, thus reducing the possible investment risks to a certain extent and obtaining high and stable investment returns.

### Data Availability

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

### Conflicts of Interest

All the authors do not have any possible conflicts of interest.

### References

- [1] M. Wen, P. Li, L. Zhang, and Y. Chen, "Stock market trend prediction using high-order information of time series," *IEEE Access*, vol. 7, pp. 28299–28308, 2019.
- [2] W. Chen, M. Jiang, W.-G. Zhang, and Z. Chen, "A novel graph convolutional feature based convolutional neural network for stock trend prediction," *Information Sciences*, vol. 556, pp. 67–94, 2021.
- [3] M. Z. Asghar, F. Rahman, F. M. Kundi, and S. Ahmad, "Development of stock market trend prediction system using multiple regression," *Computational and Mathematical Organization Theory*, vol. 25, no. 3, pp. 271–301, 2019.
- [4] H. Ni, S. Wang, and P. Cheng, "A hybrid approach for stock trend prediction based on tweets embedding and historical prices," *World Wide Web*, vol. 24, no. 3, pp. 849–868, 2021.
- [5] F. Zhou, H.-m. Zhou, Z. Yang, and L. Yang, "EMD2FNN: a strategy combining empirical mode decomposition and factorization machine based neural network for stock market trend prediction," *Expert Systems with Applications*, vol. 115, pp. 136–151, 2019.
- [6] M. Zhao, A. Jha, Q. Liu et al., "Faster mean-shift: GPU-accelerated clustering for cosine embedding-based cell segmentation and tracking," *Medical Image Analysis*, vol. 71, article 102048, 2021.

- [7] A. Moghar and M. Hamiche, "Stock market prediction using LSTM recurrent neural network," *Procedia Computer Science*, vol. 170, pp. 1168–1173, 2020.
- [8] M. Zhao, Q. Liu, A. Jha et al., "VoxelEmbed: 3D instance segmentation and tracking with voxel embedding based deep learning," 2021, <https://arxiv.org/abs/2106.11480/>.
- [9] J. Zhang, S. Cui, Y. Xu, Q. Li, and T. Li, "A novel data-driven stock price trend prediction system," *Expert Systems with Applications*, vol. 97, pp. 60–69, 2018.
- [10] A. Thakkar and K. Chaudhari, "CREST: cross-reference to exchange-based stock trend prediction using long short-term memory," *Procedia Computer Science*, vol. 167, pp. 616–625, 2020.
- [11] W. Sun, P. Zhang, Z. Wang, and D. Li, "Prediction of cardiovascular diseases based on machine learning," *ASP Transactions on Internet of Things*, vol. 1, no. 1, pp. 30–35, 2021.
- [12] M. Nabipour, P. Nayyeri, H. Jabani, S. Shahab, and A. Mosavi, "Predicting stock market trends using machine learning and deep learning algorithms via continuous and binary data; a comparative analysis," *IEEE Access*, vol. 8, pp. 150199–150212, 2020.
- [13] M. Vijh, D. Chandola, V. A. Tikkiwal, and A. Kumar, "Stock closing price prediction using machine learning techniques," *Procedia Computer Science*, vol. 167, pp. 599–606, 2020.
- [14] S. Paul and S. Vishnoi, "Real-time stock trend prediction via sentiment analysis of news article," *Computer Engineering and Intelligent Systems*, vol. 9, no. 7, pp. 21–28, 2018.
- [15] T. Kim and H. Y. Kim, "Forecasting stock prices with a feature fusion LSTM-CNN model using different representations of the same data," *PLoS One*, vol. 14, no. 2, article e0212320, 2019.
- [16] W. Long, L. Song, and Y. Tian, "A new graphic kernel method of stock price trend prediction based on financial news semantic and structural similarity," *Expert Systems with Applications*, vol. 118, pp. 411–424, 2019.
- [17] M. A. I. Sunny, M. M. S. Maswood, and A. G. Alharbi, "Deep learning-based stock price prediction using LSTM and bi-directional LSTM model," in *2020 2nd Novel Intelligent and Leading Emerging Sciences Conference (NILES)*, pp. 87–92, Cairo, Egypt, 2020, October.
- [18] H. S. Ibrahim, A. Z. Khan, S. Attia, and Y. Serag, "Classification of heritage residential building stock and defining sustainable retrofitting scenarios in Khedivial Cairo," *Sustainability*, vol. 13, no. 2, p. 880, 2021.
- [19] T. Colliri and L. Zhao, "Stock market trend detection and automatic decision-making through a network-based classification model," *Natural computing*, pp. 1–14, 2021.
- [20] Y. Gu, A. Chen, X. Zhang, C. Fan, K. Li, and J. Shen, "Deep learning based cell classification in imaging flow cytometer," *ASP Transactions on Pattern Recognition and Intelligent Systems*, vol. 1, no. 2, pp. 18–27, 2021.
- [21] Y. Yan and D. Yang, "A stock trend forecast algorithm based on deep neural networks," *Scientific Programming*, vol. 2021, Article ID 7510641, 7 pages, 2021.
- [22] Z. Wang, P. Zhang, W. Sun, and D. Li, "Application of data dimension reduction method in high-dimensional data based on single-cell 3D genomic contact data," *ASP Transactions on Computers*, vol. 1, no. 2, pp. 1–6, 2021.
- [23] J. P. Verma, S. Tanwar, S. Garg, I. Gandhi, and N. H. Bachani, "Evaluation of pattern based customized approach for stock market trend prediction with big data and machine learning techniques," *International Journal of Business*, vol. 6, no. 3, pp. 1–15, 2019.
- [24] Y. Chen, W. Lin, and J. Z. Wang, "A dual-attention-based stock price trend prediction model with dual features," *IEEE Access*, vol. 7, pp. 148047–148058, 2019.
- [25] M. Y. Chen, C.-H. Liao, and R.-P. Hsieh, "Modeling public mood and emotion: stock market trend prediction with anticipatory computing approach," *Computers in Human Behavior*, vol. 101, pp. 402–408, 2019.
- [26] M. Ananthi and K. Vijayakumar, "Stock market analysis using candlestick regression and market trend prediction (CKRM)," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 5, pp. 4819–4826, 2021.
- [27] S.-F. Huang, M. Guo, and M.-R. Chen, "Stock market trend prediction using a functional time series approach," *Quantitative Finance*, vol. 20, no. 1, pp. 69–79, 2020.
- [28] L. Huang, G. Xie, W. Zhao, Y. Gu, and Y. Huang, "Regional logistics demand forecasting: a BP neural network approach," *Complex & Intelligent Systems*, pp. 1–16, 2021.
- [29] Y. Fang, "Feature selection, deep neural network and trend prediction," *Journal of Shanghai Jiaotong University (Science)*, vol. 23, no. 2, pp. 297–307, 2018.
- [30] C.-s. Lim, "Fair performance evaluation method for stock trend prediction models," *The Journal of the Korea Contents Association*, vol. 20, no. 10, pp. 702–714, 2020.
- [31] L. Lei, "Wavelet neural network prediction method of stock price trend based on rough set attribute reduction," *Applied Soft Computing*, vol. 62, pp. 923–932, 2018.
- [32] C. Patel, "Impact of the financial result on stock trend prediction using SVM and M-score," *International Journal for Advance Research and Development*, vol. 3, no. 5, pp. 6–11, 2018.
- [33] A. Thakkar and K. Chaudhari, "A comprehensive survey on portfolio optimization, stock price and trend prediction using particle swarm optimization," *Archives of Computational Methods in Engineering*, vol. 28, no. 4, pp. 2133–2164, 2021.
- [34] C.-s. Lim, "A study on stock trend determination in stock trend prediction," *Journal of the Korean Society of Computer and Information*, vol. 25, no. 12, pp. 35–44, 2020.
- [35] V. Gupta and M. Ahmad, "Stock price trend prediction with long short-term memory neural networks," *International Journal of Computational Intelligence Studies*, vol. 8, no. 4, pp. 289–298, 2019.