

RESEARCH ARTICLE

Improving stock trading decisions based on pattern recognition using machine learning technology

Yaohu Lin¹, Shancun Liu^{1,2}, Haijun Yang^{1,3*}, Harris Wu⁴, Bingbing Jiang⁵

1 School of Economics and Management, Beihang University, Beijing, China, **2** Key Laboratory of Complex System Analysis, Management and Decision (Beihang University), Ministry of Education, Beijing, China, **3** Beijing Advanced Innovation Center for Big Data and Brain Computing, Beihang University, Beijing, China, **4** Strome College of Business, Old Dominion University, Norfolk, Virginia, United States of America, **5** Software Engineering Center, Chinese Academy of Sciences, Beijing, China

* navy@buaa.edu.cn

OPEN ACCESS

Citation: Lin Y, Liu S, Yang H, Wu H, Jiang B (2021) Improving stock trading decisions based on pattern recognition using machine learning technology. PLoS ONE 16(8): e0255558. <https://doi.org/10.1371/journal.pone.0255558>

Editor: Stefan Cristian Gherghina, The Bucharest University of Economic Studies, ROMANIA

Received: March 19, 2020

Accepted: July 20, 2021

Published: August 6, 2021

Copyright: © 2021 Lin et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within paper and its [Supporting Information](#) files. Alternatively, the data may be accessed by downloading the files "ThreeDays_sample.zip" and "TwoDays_Sample.zip" at http://semen.buaa.edu.cn/Faculty/Finance/YANG_Haijun/Profile.htm.

Funding: Haijun Yang: Grant No. 71771006
Shancun Liu: Grant No. 71771008 The authors declare that they have no competing financial interests. This research was partially supported by National Natural Science Foundation of China (Grant No. 71771006 and 71771008). There was

Abstract

PRML, a novel candlestick pattern recognition model using machine learning methods, is proposed to improve stock trading decisions. Four popular machine learning methods and 11 different features types are applied to all possible combinations of daily patterns to start the pattern recognition schedule. Different time windows from one to ten days are used to detect the prediction effect at different periods. An investment strategy is constructed according to the identified candlestick patterns and suitable time window. We deploy PRML for the forecast of all Chinese market stocks from Jan 1, 2000 until Oct 30, 2020. Among them, the data from Jan 1, 2000 to Dec 31, 2014 is used as the training data set, and the data set from Jan 1, 2015 to Oct 30, 2020 is used to verify the forecasting effect. Empirical results show that the two-day candlestick patterns after filtering have the best prediction effect when forecasting one day ahead; these patterns obtain an average annual return, an annual Sharpe ratio, and an information ratio as high as 36.73%, 0.81, and 2.37, respectively. After screening, three-day candlestick patterns also present a beneficial effect when forecasting one day ahead in that these patterns show stable characteristics. Two other popular machine learning methods, multilayer perceptron network and long short-term memory neural networks, are applied to the pattern recognition framework to evaluate the dependency of the prediction model. A transaction cost of 0.2% is considered on the two-day patterns predicting one day ahead, thus confirming the profitability. Empirical results show that applying different machine learning methods to two-day and three-day patterns for one-day-ahead forecasts can be profitable.

1. Introduction

Analyzing and forecasting the stock market is notoriously tricky due to the high degree of noise [1] and semi-strong form of market efficiency [2], which is generally accepted. A reasonably accurate prediction may raise the potential of yielding benefits and hedging against market risks. However, financial economists often question the existence of opportunities for profitable predictions [3].

no additional external funding received for this study.

Competing interests: The authors have declared that no competing interests exist.

Technical analysis, also called candlestick charting, is one of the most common traditional analysis methods to predict the financial market. By utilizing open-high-low-close prices in chronological order, candlestick charting can reflect not only the changing balance between supply and demand [4] but also the sentiment of the investors in the market [5]. Bulkowski described the known 103 patterns in natural language [6], and then comprehensive formal specifications of the known candlestick patterns were proposed [7]. In recent years, technical analysis has been proven to be effective in stock market analysis. For example, Caginalp and Laurent tested the predictive capability of candlestick patterns and found that applying candlestick trading strategies on daily stock returns in S&P 500 stocks can result in profits [8]. Goo et al. found that many one-day candlestick and reversal patterns can help investors earn significant returns by following candlestick trading strategies [9]. Moreover, the profitability of candlestick trading strategies was further confirmed [10, 11]. More complex candlestick patterns have been used in the latest research. The predictive power of 5 two-day reversal patterns was examined [12], and 4 pairs of two-day patterns were studied [13]. Lu et al. studied the profitability of 8 three-day reversal patterns under trend conditions and different holding strategies [11].

Although these studies have shown that using the candlestick pattern strategy can be profitable, dissenting voices have emerged in academia. Fock et al. found no evidence of the predictive ability of candlestick patterns alone or in combination with other common technical indicators in the DAX stock index contract and the Bund interest rate future [14]. Duvingage et al. tested the intraday predictive power of Japanese candlesticks at the 5-minute interval on the 30 constituents of the DJIA index and concluded that candlestick trading strategies do not improve investment performance [15]. These conflicting conclusions about candlestick patterns prompt us to investigate further.

Artificial intelligence has recently been applied to address the chaotic and randomness time series data [16, 17]. The intense computational use of intelligent predictive models has commonly been studied under machine learning [18]. Compared to the more traditional models, machine learning models provide more flexibility [19], do not require distributional assumptions, and can easily combine individual classifiers to reduce variance [20]. Many machine techniques have already been applied to forecast the stock market [20–36]. For example, logistic regression (LR) and neural networks (NNs) [27, 29, 30, 36], deep neural networks (DNNs) [22], decision trees (DTs) [22, 25], support vector machines (SVMs) [24, 26, 28] or support vector regression (SVR) [21], k-nearest neighbors (KNN) [23, 33], random forests (RFs) [22], long short-term memory networks (LSTMs) [1, 31, 34] and restricted Boltzmann machines (RBMs) [32] have been used to predict stock market movements. In comparing multiple machine learning methods, Fischer et al. (2018) deploy LSTM networks for predicting out-of-sample directional movements for the constituent stocks of the S&P 500 from 1992 until 2015. They find LSTM networks to outperform RF, DNN and LR. Patel et al. (2015) compare the performance of four models (namely, ANN, SVM, RF, and naïve-Bayes) with respect to the following indexes and companies on the Indian stock market: CNX Nifty, S&P BSE Sensex, Infosys Ltd. and Reliance Industries [35]. Brownstone (1996) used a neural network to predict daily closing prices for five days and twenty-five days of the FTSE 100 Share Index in the UK and used multiple linear regression to compare the prediction results [30]. Krauss et al. (2017) implemented and analyzed the one-day effectiveness of deep neural networks (DNNs), gradient-boosted-trees (GBTs), and random forests (RFs) on all stocks of the S&P 500 from 1992 to 2015, and then the trading signals were generated based on the forecast probability. These techniques sort all stocks over the cross-section k probability in descending order. Five different investment strategies are generated by going long the top k stocks and short the bottom K stocks, with $k \in \{10, 50, 100, 150, 200\}$. These techniques achieve the best performance when $k = 10$, and RF outperforms GBT and DNN [22]. Profitable patterns may be discovered based

on more recent returns and daily data. Qiu et al. (2020) established an LSTM with a wavelet forecasting framework to predict the opening prices of stocks [34].

The existing research shows that machine learning methods can effectively predict the direction of the financial market. However, the use of machine learning technology for candlestick pattern recognition is still less prevalent. Moreover, traditional research on candlestick patterns focuses mainly on a limited number of patterns [9–15]. The latest artificial intelligence technology prompts us to consider applying pattern recognition to decision-making in the stock market. Since different machine learning methods perform differently in different scenarios [1, 22, 30, 35], a prediction framework that can adapt to different machine learning methods will greatly improve the usability of the model.

The remainder of this paper is organized as follows: Section 2 outlines the design of this paper. A pattern recognition prediction framework with four popular machine learning methods is designed. Then two other popular machine learning methods are used to replace the four machine learning methods to evaluate the dependency of the prediction framework. Section 3 presents the empirical results. Section 4 concludes the paper.

2. Methodology

This paper attempts to develop PRML, a Pattern Recognition method based on Machine Learning methods, to improve stock trading decisions, as shown in Fig 1. First, 13 forms of one-day patterns are constructed and classified, and then the corresponding technical indicators and location information are calculated. Then, the pattern-building phase starts; this phase generates all possible combinations of candlestick patterns. For example, the two-day patterns are composed of two consecutive one-day patterns; therefore, there are 13×13 possible combinations. All the pattern information, including the technical indicators and location information, is put into the machine learning models, which test the prediction accuracy of each pattern during different periods. The pattern recognition stage retains only those patterns whose accuracy exceeds the threshold value. Finally, the adaptive recommendation schedule gives corresponding stock prediction actions based on the evaluated results.

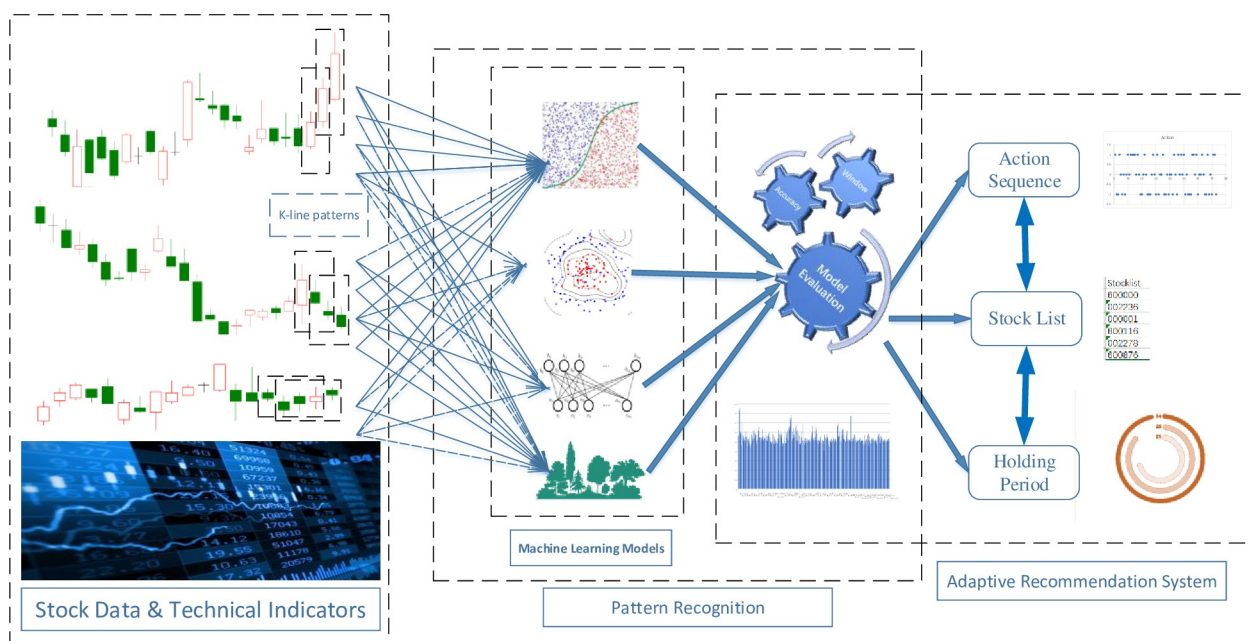


Fig 1. Overview of the PRML model.

<https://doi.org/10.1371/journal.pone.0255558.g001>

Based on the identified patterns, the daily portfolio is dynamically built. The main pattern recognition process is shown in Fig 2.

2.1 Candlestick patterns

A candlestick, also called a K-line, consists of four basic elements: the opening price, high price, low price and closing price, as shown in Fig 3. For simplicity, these elements are often denoted as open, high, low, and close. According to the different values of open, high, low and close, one-day patterns can have 13 different forms, as shown in Fig 3.

2.1.1 Definitions. The definitions and functions used to describe the rules for classifying daily candlestick patterns are listed as follows:

2.1.1.1 Definition 1 (Candlestick). A candlestick $k = (o_t, h_t, l_t, c_t)$ is a tuple that consists of four basic prices of a stock at time t . A candlestick k is a basic element in identifying the candlestick pattern recognition. The o_t, h_t, l_t and c_t represent the opening price, high price, low price and closing price at time t , respectively. Additionally, $k_i = (o_{it}, h_{it}, l_{it}, c_{it})$ denotes the i^{th} candlestick at time t .

2.1.1.2 Definition 2 (Candlestick time series). A candlestick time series $T_n = \{k_1, k_2, \dots, k_n\}$ is a sequence of candlesticks of a stock; the sequence consists of n candlesticks from day 1 to day n . Additionally, $T_{in} = \{k_{i1}, k_{i2}, \dots, k_{in}\}$ denotes the i^{th} stock sequence.

2.1.1.3 Definition 3 (Candlestick relative position). The candlestick relative position is defined as:

$$loc = \begin{cases} BC_h & \text{where } h_t > h_{t-1}, l_t < l_{t-1}, c_t > c_{t-1} \\ BC_l & \text{where } h_t > h_{t-1}, l_t < l_{t-1}, c_t < c_{t-1} \\ BH_h & \text{where } h_t > h_{t-1}, l_t > l_{t-1}, c_t > c_{t-1} \\ BH_l & \text{where } h_t > h_{t-1}, l_t > l_{t-1}, c_t < c_{t-1} \\ BL_h & \text{where } h_t < h_{t-1}, l_t < l_{t-1}, c_t > c_{t-1} \\ BL_l & \text{where } h_t < h_{t-1}, l_t < l_{t-1}, c_t < c_{t-1} \\ BM_h & \text{where } h_t < h_{t-1}, l_t > l_{t-1}, c_t > c_{t-1} \\ BM_l & \text{where } h_t < h_{t-1}, l_t > l_{t-1}, c_t < c_{t-1} \end{cases}$$

loc_i denotes the relative position of the i^{th} candlestick.

Algorithm: Pattern recognition algorithm

Input: Pattern data, which include different indicators

Output: Best performance machine learning model, accuracy, pattern

```

0  Pattern Recognition (stock_data):
1  Defined accuracy_threshold;
2  foreach  $p$  in patterns:
3      Generate  $p\_data$  of  $p$ ;
4      LinearClassification ( $p\_data$ );
5      GridSearchCV of KNN ( $p\_data$ );
6      GridSearchCV of RF ( $p\_data$ );
7      BernoulliRBM using LogisticRegression ( $p\_data$ );
8      if (MAX_Accuracy (LinearClassification, KNN, RF, RBM) > threshold):
9          Save best performance model, accuracy,  $p$ ;
```

Output: List of best performance model, accuracy, pattern

Fig 2. Main pattern recognition algorithm.

<https://doi.org/10.1371/journal.pone.0255558.g002>

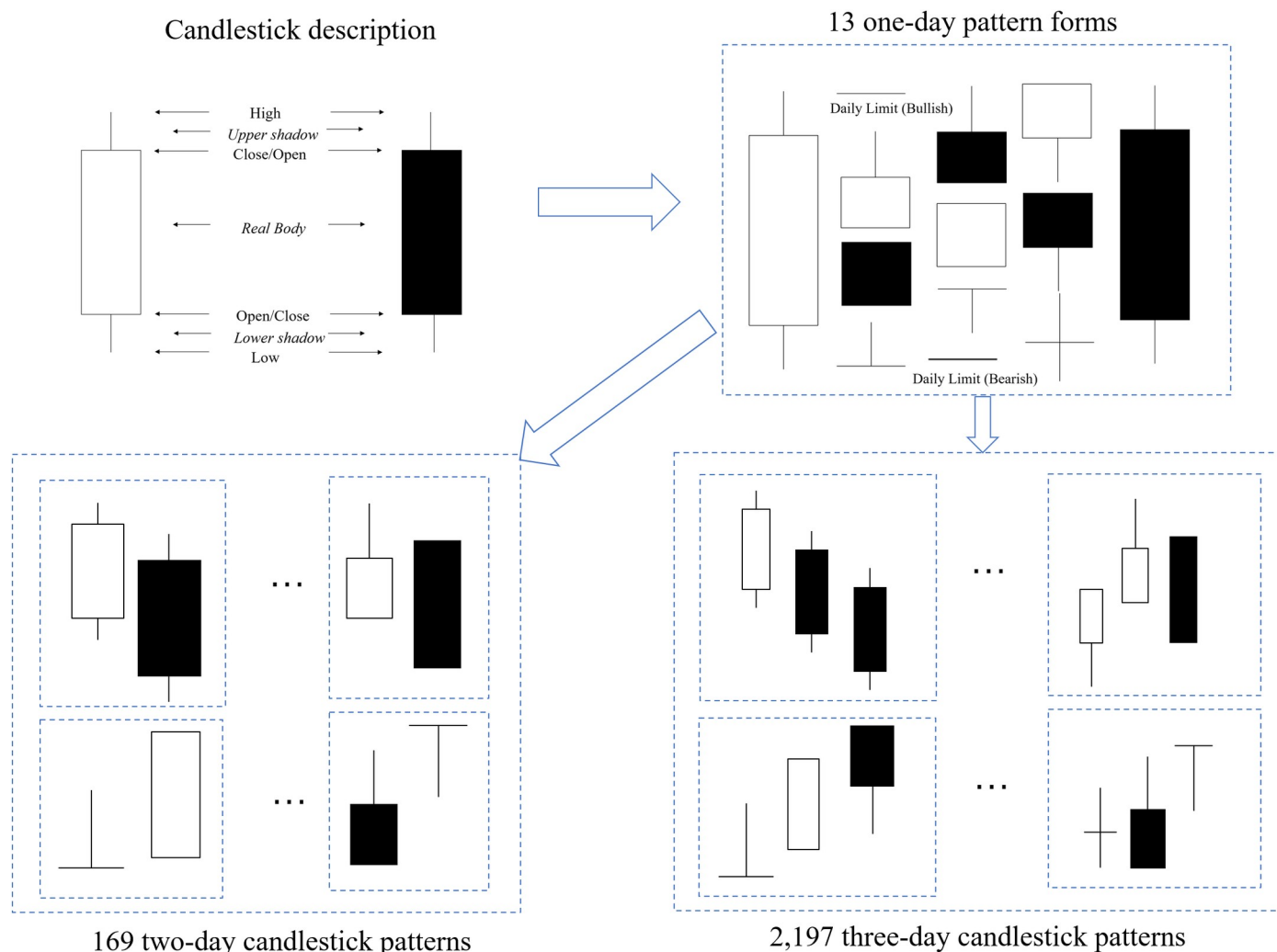


Fig 3. Description of a candlestick and candlestick patterns.

<https://doi.org/10.1371/journal.pone.0255558.g003>

2.1.1.4 Definition 4 (Candlestick relative position series). $LOC_n = \{loc_1, loc_2, \dots, loc_n\}$ is a sequence of relative position information of a stock; the sequence consists of n relative position information from day 1 to day n .

2.1.1.5 Definition 5 (Candlestick pattern). A candlestick pattern or K-line pattern $p_j = \{T_j, Loc_j\}$ is a subsequence of consecutive candlesticks; this subsequence consists of two parts: a sequence of candlesticks and a corresponding location sequence. For example, a two-day pattern can be defined as $p_2 = \{T_2, Loc_2\}$ and $T_2 = \{k_1, k_2\}$, $Loc_2 = \{loc_1, loc_2\}$.

2.2 Technical indicators

Technical indicators are mathematical calculations based on price and volume. By analyzing historical data, technical analysts use indicators to predict the future trend of the financial market [37, 38], and these indicators can potentially affect stock price prediction [39, 40]. Following Zhou et al. (2019) and Bao et al. (2017), we use feature sets containing several indicators that are commonly used in the technical analysis [3, 31]. The most prevalent indicators are defined below:

- A moving average (MA) is a calculation to analyze data points by creating a series of averages of different subsets of the full data set. The calculation formula is:

$$MA(t) = \frac{1}{m} \sum_{i=0}^{m-1} c_{t-i} \quad (1)$$

where m refers to the time interval and c is the closing price.

- An exponential moving average (EMA) is a first-order infinite impulse response filter that applies weighting factors, which decrease exponentially. The calculation formula is:

$$EMA(t) = \frac{2}{n+1} c_t + \frac{n-1}{n+1} EMA(t-1) \quad (2)$$

Where n refers to the time interval.

- The volume rate of change (ROC) shows the changes in volume. The calculation formula is:

$$ROC(t) = \frac{v_t}{\frac{1}{x} \sum_{i=0}^x v_{t-i}} \quad (3)$$

Where x refers to the time interval and v_t refers to the volume at time t .

- The commodity channel index (CCI) is designed to detect the beginning and the ending market trends and measures a security's variation from the statistical mean. The calculation formula is:

$$CCI(t) = \frac{\frac{h_t + l_t + c_t}{3} - MA(n)}{0.015 * \frac{1}{n} \sum_{i=n}^n MA(i) - c_i} \quad (4)$$

- Momentum (MOM) measures the acceleration and deceleration of prices. The calculation formula is:

$$MOM(t) = c_t - c_{t-n} \quad (5)$$

- The Chaikin A/D line (AD) measures the accumulation-distribution line and is calculated as follows:

$$Clv(t) = \frac{2 * c_t - h_t - c_t}{h - l_t} \quad (6)$$

$$AD(t) = AD(t-1) + v_t * Clv(t) \quad (7)$$

- On balance volume (OBV) is a cumulative total of the up and down volume. The calculation logic of OBV is as follows:

if $c_t > c_{t-1}$

$$OBV(t) = OBV(t-1) + v_t \quad (8)$$

if $c_t < c_{t-1}$

$$OBV(t) = OBV(t-1) - v_t \quad (9)$$

else

$$OBV(t) = OBV(t-1) \quad (10)$$

where v refers to the volume.

- The true range (TRANGE) is a base calculation that is used to determine the normal trading range of a stock or commodity. The calculation formula is:

$$TR(t) = \max(H_t, C_{t-1}) - \min(L_t, C_{t-1}) \quad (11)$$

- The average true range (ATR) is a moving average of the true range. The calculation formula is:

$$ATR(t) = \frac{1}{n} \sum_{i=1}^n TR(t-i+1) \quad (12)$$

2.3 Pattern generation & generation of training and testing sets

We begin to extract more complex N-day daily patterns based on the one-day candlestick patterns. First, the one-day candlestick pattern classification is generated on the basis price of open, high, low, and close. A total of 13 one-day patterns are obtained after considering all the circumstances. Second, the values of nine technical indicators, shape and relative location information are calculated. The research in this paper focuses on short-term effects; 5 days or 10 days as the parameters of indicators are chosen. The parameters of 5 days for *MA* and 10 days for *EMA*, *CCI* and *ATR* are used for testing. All the characteristics of the one-day patterns are obtained through these two processes. Finally, the one-day candlestick patterns are combined into more complex patterns for each stock according to different time windows. A total of 169 two-day candlestick patterns are combined based on the 13 one-day candlestick patterns. Time must be continuous in this pattern combination process.

After all the patterns and features are ready, we begin to prepare the training sets and testing sets. The entire data set are divided into two parts: the machine learning data set from Jan 1, 2000 to Dec 31, 2014 and the forecasting set from Jan 1, 2015 to Oct 30, 2020. Of the machine learning data set, 80% is used as training subsets, and 20% is used as testing subsets. First of all, the data of the training subset is used to fit the parameters in the machine learning model. Next, the machine learning model that fits the parameters is predicted in the testing subset, and the predicted result is compared with the real value to get the accuracy rate. The testing subset is the validation set. The corresponding result is the next N-day's direction of the close price; N is from 1 to 10.

2.4 Prediction models

The inference engine is introduced in this phase. We use four machine learning models, logistic regression (LR), k-nearest neighbors (KNN), random forest (RF), and restricted Boltzmann machine (RBM), to predict the direction of the close price. The parameters used in the four machine learning methods are shown as [Table 1](#).

Table 1. Parameters of the four machine learning models used in prediction schedule.

MLs	Parameters
LR	Regularized = <i>L2</i> , solver_parameter = <i>warn</i> , <i>C</i> = 1.0, iteration = 100, criteria = 0.0001
KNN	n_neighbors = range(1,10), weights = ['uniform','distance'], algorithm = ['auto','ball_tree','kd_tree','brute'], leaf_size = range(1,2), optimizer = <i>GridSearchCV</i> , cv = 10
RBM	learning_rate = 0.06, iteration = 10, <i>C</i> = 6000, components = 100
RF	n_estimators = range(10,100,5), criterion = [gini, entropy], min_samples_leaf = [2, 4, 6,50], max_depth = range(1,10), optimizer = <i>GridSearchCV</i> , cv = 10

<https://doi.org/10.1371/journal.pone.0255558.t001>

2.4.1 Logistic regression (LR). Logistic regression is the most basic machine learning algorithm. The logistic regression model returns an equation that determines the relationship between the independent variables and the dependent variable. The model calculates linear functions and then converts the result into a probability. Finally, the model converts the probability into a label.

In the empirical stage, we use *L2* as a regularized parameter and specify *warn* as the solver parameter that determines our optimization method for the logistic regression loss function. In terms of the termination parameters of the algorithm, we set the maximum number of iterations that are taken for the solvers to converge to 100 and set the tolerance for the stopping criteria parameter to 0.0001.

2.4.2 K-nearest neighbors (KNN). K-nearest neighbors (KNN) is another machine learning algorithm. The k-NN algorithm looks for 'k' nearest records within the training data set and uses most of the classes of the identified neighbors for classifying. Subha used k-NN to classify the stock index movement [23], and Zhang et al. used ensemble empirical mode decomposition (EEMD) and a multidimensional k-nearest neighbor model (MKNN) to forecast the closing price and high price of the stocks [33].

In the experimental stage, we obtain the best performance from the grid search algorithm, which sets different neighbors, leaves, and weights. Different parameter combinations produce different clustering effects.

2.4.3 Restricted Boltzmann machine (RBM). A restricted Boltzmann machine (RBM) is a generative stochastic artificial neural network that can learn a probability distribution over its input sets. Recently, due to their powerful representation, RBMs have been used as generative models of many types of data, including text, images and speech. Liang et al. used an RBM to predict short-term stock market trends [32].

In the empirical stage, we connected a logistic regression to the output of the RBM for classification. Different parameter combinations may produce different classification effects; therefore, we set 10 iterations and 100 components to improve training results.

2.4.4 Random forest (RF). The following brief description of random forests follows Breiman (2001). Random forests are a combination of tree predictors such that each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest. Random forests (RFs) are nonparametric and nonlinear classification and regression algorithms [41]. Random forests not only use a subset of the training set but also select only a subset of the feature set when the tree is established in the decision tree. In the training stage, RF repeats *n* times to select a random sample with replacement of the training set and selects *k* features randomly to build a decision tree. Then repeat the above steps T_{num} times to build T_{num} decision trees. After each tree decision, the final result is confirmed by voting. Booth (2014) used RFs to construct an automated trading mechanism [42].

In the empirical stage, RF repeats 5000 times to select a random sample with replacement of the training set and selects \sqrt{K} features, where *K* is the number of the input features. There are

three tuning parameters: the number of trees T_{num} , their maximum depth T_{depth} , and the minimum number of samples required to be at a leaf node T_{sample} . We set these parameters as $T_{num} \in [10, 100]$, $T_{depth} \in [1, 10]$ and $T_{sample} \in \{2, 4, 6, 50\}$ into a relatively large selectable range and use the grid search algorithm to find the optimal performance.

2.5 Two other testing machine learning models

2.5.1 Multilayer perceptron network (MLP). A multilayer perceptron (MLP) is a class of feedforward artificial neural network (ANN). A MLP consists of an input layer, one or more hidden layers and an output layer. The following brief description of MLP follows Moghadam et al. (2016) [36]. The input layer matches the feature space, and the output layer matches the output space, which may be a classification or regression layer. In the network, each neuron in the previous layer is fully connected with all neurons in the subsequent layer and represents a certain weight. Each non-output layer of the net has a bias unit, serving as an activation threshold for the neurons in the subsequent layer. In this study, the most common three-layer MLP model is constructed based on experience. In order to improve the generalization of the model, the number of inputs is set to 64, which is greater than the number of features. Finally, a three-layer MLP network, including an input layer with 64 nodes, a hidden layer with 64 neurons and an output layer with 1 neuron is developed. The *ReLU* activation, 0.1 learning rate, 20 epochs, 128 batch sizes and an *RMSProp* optimizer for the objective function of *binary_crossentropy* are used in the learning phase.

2.5.2 Long short-term memory neural networks (LSTMs). Long short-term memory neural networks (LSTMs) are one of the most common forms of recurrent neural networks (RNNs), which are a type of deep neural network architecture. This description of LSTMs follows the description of Fischer et al. (2018), Bao et al. (2017) and Qiu et al. (2020) [1, 31, 34]. The LSTM consists of a set of memory cells that replace the hidden layer neurons of the RNN. The memory cell consists of three components: the input gate, the output gate, and the forget gate. The gates control the interactions between neighboring memory cells and the memory cell itself. The input gate controls the input state, while the output gate controls the output state, which is the input of other memory cells. The forget gate can choose to remember or forget its previous state. In this study, a common three-layer LSTM model is constructed based on experience. The number of inputs is set to 64 to improve the generalization. Finally, a three-layer LSTM network that includes 64 neurons input and an *Adam* optimizer for the objective function of *binary_crossentropy*, a middle layer with 64 neurons and an output layer with 1 neuron is developed; 10 *epochs* are constructed to increase the stability. A total of 56,129 parameters are generated by the LSTM in the prediction process.

2.6 Model evaluation

The machine learning models are used to forecast stock fluctuations. To evaluate the performance of the prediction models, the common evaluation criterion of *Accuracy* is used in this study. *Accuracy* is used to evaluate the overall classification ability of the model. The formula is as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (13)$$

where *TP* (*true positive*) indicates that both the model prediction and the real sample are true; *TN* (*true negative*) indicates that both the model prediction and the real sample are false; *FP* (*false positive*) indicates that the model prediction is true but that the real sample is false; and *FN* (*false negative*) indicates that the model prediction is false while the real sample is true. In

the empirical stage, patterns are selected into the strategy pool when the model's accuracy rate is higher than 55%.

In the dependence testing stage, the *F-measure* is also introduced to evaluate the performance of MLP and LSTM. Following Patel et al. (2015) and Zhou et al (2019), the *F-measure* evaluation method and the additional metrics are defined as follows [3, 35]:

$$Precision = \frac{TP}{TP + FP} \quad (14)$$

$$Recall = \frac{TP}{TP + FN} \quad (15)$$

$$F - measure = \frac{2 * Precision * Recall}{Precision + Recall} \quad (16)$$

Recall is also called sensitivity or the true positive rate. The *F-measure* is an increasing function when *Precision* and *Recall* are equally important.

Four statistical measures are chosen to evaluate the trading performances, including Sharpe Ratio, Maximum Drawdown, Average Annual Return and Information Ratio. These four measures are defined as follows:

$$SharpeRatio = \frac{\mu(R_t) - r_f}{\sigma(R_t)} \quad (17)$$

Where R_t is the cumulative return until date t , $\mu(R_t)$ and $\sigma(R_t)$ are the corresponding mean and standard deviation of the return R_t , r_f is the risk-free return.

$$MaximumDrawdown = \max_{\tau \in (0, T)} (\max_{t \in (0, \tau)} (R_t - R_\tau)) \quad (18)$$

Where T denotes a period.

$$AverageAnnualReturn = \left(\prod_{i=1}^N (1 + r_i) \right)^{\frac{1}{N}} - 1 \quad (19)$$

Where r_i denotes the return of year i .

$$InformationRatio = \frac{R_t - R_b}{\sigma_t} \quad (20)$$

Where R_b is the benchmark return.

The Sharpe Ratio depicts the risk-adjusted return, the Maximum Drawdown denotes the largest cumulative loss over the period of investment and the Information Ratio measures the portfolio returns beyond the returns of a benchmark.

2.7 Investment strategy

Based on the above evaluation criteria, the investment strategy is constructed. According to the actual situation of the Chinese stock market, short selling is limited to some stocks with securities margin trading. Therefore, we consider only long selling and build the corresponding investment strategy with different time windows from 1 day to 10 days. This article assumes that we will invest at the close price at time t and will be clear at the close price at time $t+N$, where N is from 1 to 10. Suppose that we have initial capital M and the current stock can be bought without affecting price fluctuations; then, the specific construction steps of the equal-weight investment strategy are as follows:

First, all the possible two-day and three-day candlestick patterns of all stocks traded on the Chinese stock market are checked at time t . Next, the machine learning methods are used to predict the rise or fall of $t+N$ based on the above evaluation model. Only the best-performing of the machine learning methods and forecasting results will be saved after comparing the prediction results of these machine learning models. If the predicted result is long and consistent with the real result, the $t+N$ profit is recorded. If the prediction is wrong, the negative profit of $t+N$ is recorded as a loss value. Furthermore, it would do nothing if the predicted result is short. Then, the above steps are repeated to calculate $t+1$, $t+2$, etc. Finally, the recommendation stage for investment is carried out. The specific patterns where accuracy exceeds the threshold are screened out in the training and testing stage. The daily investment portfolio is adaptively built according to the filtered patterns pools. Five different strategy pools are used to form 5 comparable strategies: the *All*, *Adjust*, *TOP10*, *TOP5* and *TOP3* candlestick pattern pool. The *All* baseline candlestick pattern pool contains all the candlestick patterns whose accuracy is greater than 55% in the testing stage. After excluding from the rule pool the patterns that appear fewer than 1000 times in the machine learning set, from the *All* candlestick pattern pool, we obtain the *Adjust* patterns pool. The 10 most accurate patterns in the adjust pool were used to form the *TOP10* candlestick patterns pool; the 5 most accurate patterns in the adjust pool were used to form the *TOP5* candlestick patterns pool; and the 3 most accurate patterns in the adjust pool were used to form the *TOP3* candlestick patterns pool. The investment flowchart is shown as Fig 4.

In this equal-weight investment strategy, there is M/N capital each day. Suppose there are P_i patterns in the i -th investment strategy, then each pattern has $\frac{M}{N} * \frac{1}{P_i}$ capital for investment. Suppose there are K stocks in pattern p_i at time t , then a total amount of $\frac{M}{N} * \frac{1}{P_i} * \frac{1}{K_i}$ capital will be invested on each stock. Finally, the average daily return of the portfolio is calculated in the forecasting stage to test prediction performance.

3. Empirical results

3.1 Data and training environment

In this study, we use the daily data on the Chinese stock market from the period of Jan 1, 2000 to Oct 30, 2020; a list of 9,745,597 rows of original data is used in our study.

Stock data are collected from CCER, a local data provider in China. First, a data cleaning phase schedule is carried out to guarantee the validity of the training data. We remove the daily data for a given stock if the trading volume is zero, which is a sign of stopped trading due to, e.g., company reorganization. Rows that contain one or more missing values out of range are removed from the database.

Then, feature information for each stock is generated on each day t . According to the definition of the candlestick chart, the feature of *Shape* is labeled. The feature of *Loc* is generated by using the definition of candlestick relative position. The other 9 feature values are calculated by indicator formulas (1)-(12). Therefore, the daily data contains these 11 features for each stock. To ensure effectiveness, three rounds of training are carried out. We randomly choose 5,000 rows of daily stock data for each of the patterns from the database. To ensure the balance of classification during training, for each intraday pattern, we choose half of the training data with rising prices (closing price lower than next N-day's) and half of the training data with falling prices. Finally, the average accuracy is obtained based on the three rounds of results.

As a result, a list of 5,445,915 rows of the two-day patterns data with 26 data columns and 5,420,650 rows of the three-day patterns data with 38 data columns is generated. The *Date* and *Result return* information used for investment return calculation will not be used in machine

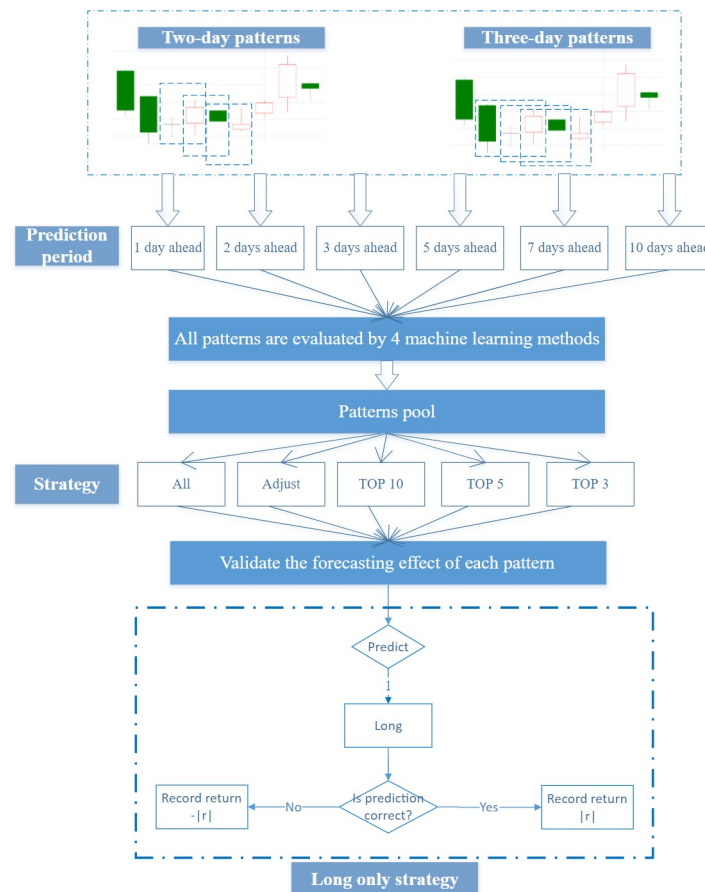


Fig 4. Flowchart of investment strategy construction for two-day patterns and three-day patterns.

<https://doi.org/10.1371/journal.pone.0255558.g004>

training. Therefore, 22 distinct features will be used as input data for two-day candlestick patterns recognition and 33 distinct features will be used as input data for three-day candlestick patterns recognition. Regardless of the two-day or three-day patterns, the *Result direction* data is used for training and evaluation of machine learning results. The features of the two-day patterns are composed of the features of the first day, the next day, and the result features. In addition to the features of the two-day patterns, the three-day patterns also include the features of the third day. The data sample which used in the machine learning models is shown as [Table 2](#).

3.2 Model comparison and evaluation

Four complete predictions are made in this study. First, all two-day candlestick patterns and three-day candlestick patterns data are put into the four machine models for prediction without distinguishing patterns. The histograms in [Fig 5](#) show the forecast accuracy in the training sets for 1, 2, 3, 5, 7, and 10 days ahead. Then, a prediction method for a basic segmentation pattern is performed. All possible combinations of different patterns for two-day and three-day patterns are put into four machine learning models to make predictions and record the best machine learning method and accuracy rate for each pattern. During training, patterns that appear fewer than 100 times are discarded. Three full-round calculating works have been done and the error bars are calculated based on the standard deviation of the three groups of calculation results. In the forecasts 2, 3, 5, 7 and 10 days ahead, the pure ML methods forecast two-

Table 2. Data samples used in the machine learning models.

First day features											
Date	Shape	Loc	MA5	EMA10	ROC(1)	CCI10	MOM10	AD	OBV	TR	ATR10
20081119	12	6	10.59	10.26	0.75	82.99	2.11	1473163	34274056	1	0.89
The next day features											
Date	Shape	Loc	MA5	EMA10	ROC(1)	CCI10	MOM10	AD	OBV	TR	ATR10
20081120	11	7	10.64	10.31	1.12	65.39	2.09	1161726	33662716	0.57	0.86
Third day features											
Date	Shape	Loc	MA5	EMA10	ROC(1)	CCI10	MOM10	AD	OBV	TR	ATR10
20081121	11	5	10.56	10.29	0.89	6.77	1.44	1115630	33116148	0.83	0.85
Result features											
Result direction				Result return							
0				-4.4							

<https://doi.org/10.1371/journal.pone.0255558.t002>

day patterns have lower accuracy than forecasting 1 day ahead, which means that there is a greater risk, leading to greater uncertainty. Similarly, in the forecast 3, 5, 7 and 10 days ahead, the pure ML methods forecast three-day patterns have lower accuracy than forecasting 1 day and 2 days ahead, which means that it may suffer continuous losses in the future. The line chart in Fig 5 shows the predicted averages for all the segment patterns.

For different candlestick patterns, four machine models have different prediction effects. Four machine learning models are used to make predictions separately for each pattern. The machine learning model with the highest prediction accuracy corresponding to each pattern will be recorded. Fig 6 shows the number of machine learning methods whose prediction accuracy exceeds the threshold. Taking the two-day candlestick patterns which have 169 combinations in theory to predict one day ahead as an example, 159 patterns have a prediction accuracy rate that exceeds our threshold, thus supporting 1,299,028 rows of data. From the bottom half of Fig 6, we can see that 14 patterns perform well using the KNN prediction method, 10 patterns make the best predictions by using the LR method, 39 patterns are supported by the RBM model, and 96 models make the best predictions by using the RF method. Taking the three-day candlestick patterns which have 2,197 combinations in theory to predict one day ahead as an example, the prediction accuracy of 451 patterns exceeded our threshold and supported 665,243 rows of data. From the upper half of Fig 6, we can see that 51 patterns perform well using the KNN prediction method, LR supports 96 models, 120 patterns make the best predictions by using the RBM method, and RF supports 184 patterns. RF outperforms LR, KNN and RBM in two-day and three-day-pattern forecasting. Except for forecasting 5 days ahead, the number of patterns that exceeds the accuracy threshold value decreases significantly as the forecast period becomes longer based on two-days pattern forecasting. The

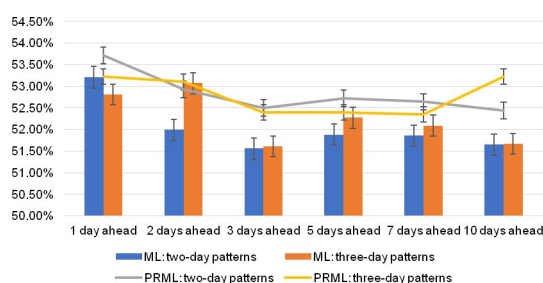


Fig 5. Accuracy overviews of PRML and pure machine learning models.

<https://doi.org/10.1371/journal.pone.0255558.g005>

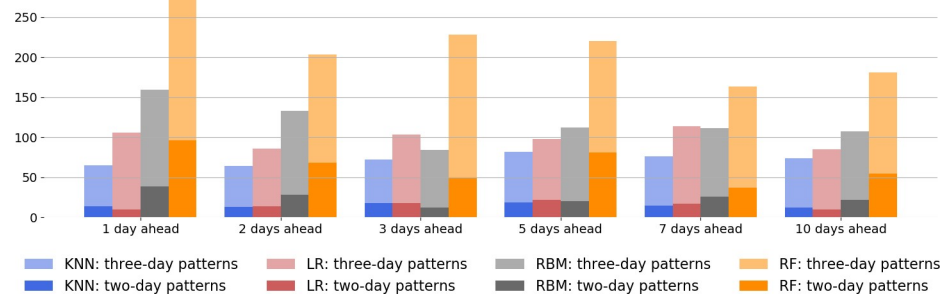


Fig 6. The number of the four machine learning methods supporting highest prediction accuracy.

<https://doi.org/10.1371/journal.pone.0255558.g006>

number of patterns to meet the conditions for forecasting 1, 2, 3, 5, 7, and 10 days ahead are 159, 123, 98, 142, 95, and 99 respectively. In terms of the three-day patterns, the number of patterns in forecasting 1 day ahead is significantly higher than in forecasting other periods. The number of three-day patterns to meet the conditions for forecasting 1, 2, 3, 5, 7, and 10 days ahead are 451, 363, 389, 370, 369, and 348 respectively. Regardless of two-day patterns or three-day patterns, the number of selecting patterns for forecasting 1 day ahead is significantly higher than forecasting other periods, which means that it is possible to obtain better performance in forecasting 1 day than in other periods.

3.3 Investment strategy result

Based on the PRML prediction framework of this paper, we conducted an investment validation in the forecasting sets. The cumulative return performance of PRML and pure ML, including forecasting 1, 2, 3, 5, 7, and 10 days ahead, is shown as Fig 7. The Shanghai Composite Index during the same period was used as a benchmark, as shown by the blue line in the figure. The performance of PRML with respect to two-day patterns and three-day patterns is better than pure ML when forecasting 1 day ahead and shows more stability when forecasting 2, 3, 5, 7 and 10 days ahead. For the two-day patterns, the pure ML models predict and give investment recommendations for 13×13 patterns every day. For the three-day patterns, the pure ML models predict $13 \times 13 \times 13$ patterns and give investment recommendations every day. However, PRML only gives investment recommendations for the patterns whose prediction accuracy exceeds the threshold. Pure ML methods predict and give investment recommendations for each stock every day, while PRML invests according to the performance of different patterns, and the average number of stocks invested per day is relatively small. This can be seen from the number of selected patterns in Fig 6 and the number of theoretical combinations. Lower forecast accuracy and frequent transactions may lead to a significant decline in earnings, as we can see from the forecast results of 3 and 7 days ahead in Fig 7. Pure ML methods have a better effect in predicting two days in the three-day patterns, which is consistent with the higher accuracy in Fig 5.

The finance performance of PRML and ML when predicting one day ahead is shown as Table 3. The average annual return of ML3, PRML2 and PRML3 show that the previous patterns are profitable. For both the two-day patterns and the three-day patterns, the financial performance of PRML is better than that of ML. With a maximum of 10.75% annual returns, the two-day patterns based on the PRML model are more profitable than the market. During the same period, two-day patterns using pure ML have a large drawdown of 75.45%, and our portfolio drawdown using PRML is smaller, thus indicating that the proposed model has less risk. In the one-day forecasting scenario, the two-day patterns are better than the three-day patterns.

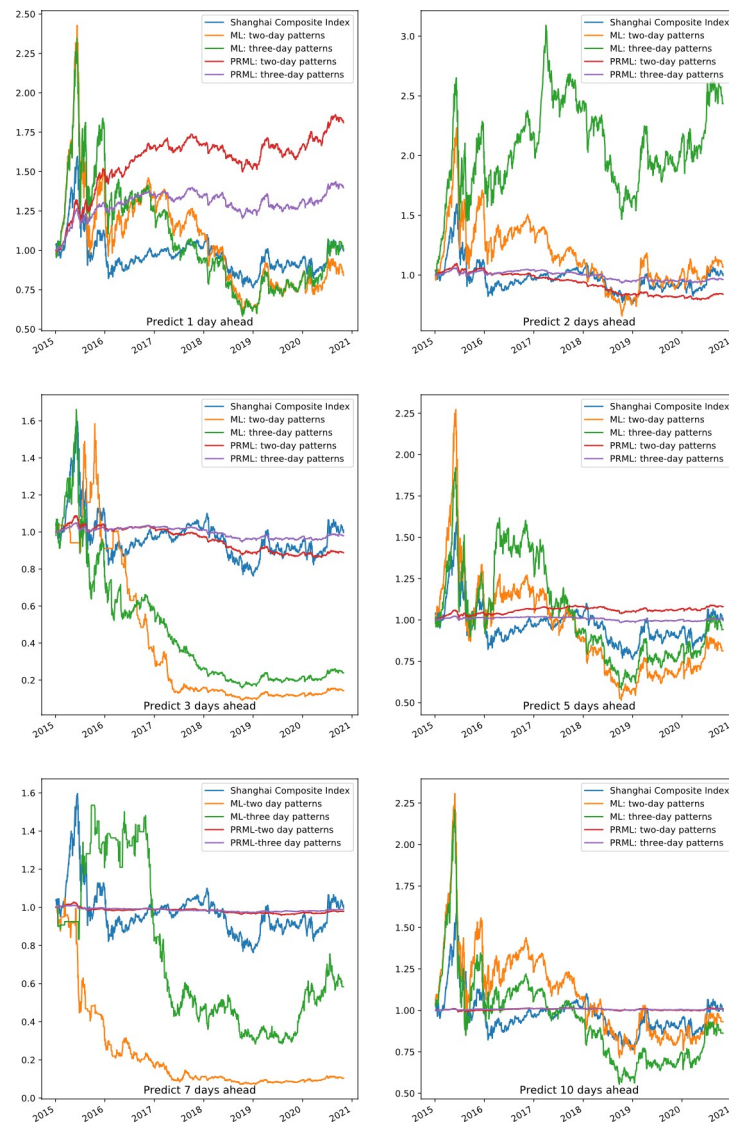


Fig 7. Portfolio forecasting performance of 1, 2, 3, 5, 7, 10 days ahead: PRML vs. ML.

<https://doi.org/10.1371/journal.pone.0255558.g007>

Table 3. Finance performance of forecasting one day ahead: PRML vs. ML.

	SH Index	ML2	PRML2	ML3	PRML3
Average Annual Return	-0.66%	-2.89%	10.75%	0.41%	5.90%
Max Drawdown	-52.28%	-75.45%	-13.81%	-75.18%	-14.05%
Annual Sharpe Ratio	-0.19	-0.25	0.17	-0.28	-0.2
Information Ratio	0	0.41	2.38	0.38	2.08
Standard Deviation	1.48	2.09	0.49	2.11	0.42

* SH Index represents the Shanghai Composite Index, ML2 represents using two-day patterns without detail and pure ML to forecast, PRML2 represents using two-patterns and PRML to forecast, ML3 represents using three-day patterns without detail and pure ML to forecast, PRML3 represents using three-day patterns and PRML to forecast. Risk-free uses the yield to maturity of China's government bonds in the past five years, with a value of 3.08%.

<https://doi.org/10.1371/journal.pone.0255558.t003>

Then, 5 more detailed investment strategies are constructed based on PRML. First, 5 different strategy pools are generated based on machine learning results and investment strategy. Next, the prediction effects of different periods from 1 day to 10 days are examined. For each day of data in the prediction set, we use the corresponding machine learning method according to the pattern rule pool to forecast. If the prediction result is long, a buy operation is performed. Then, the result with the actual situation is compared and the average daily return is calculated with equal weight.

Fig 8 shows the average portfolio return of two-day candlestick patterns, including forecasting 1, 2, 3, 5, 7, and 10 days ahead. *All* patterns indicate that we construct the portfolio according to all the patterns in the pattern rule pool. *Adjust* patterns show that we exclude the patterns in the baseline candlestick pattern pool that appear fewer than 1000 times in the machine learning set. The *TOP10* patterns indicate that we use the 10 most accurate candlestick patterns, *TOP5* means that we use only the 5 most accurate candlestick patterns, and

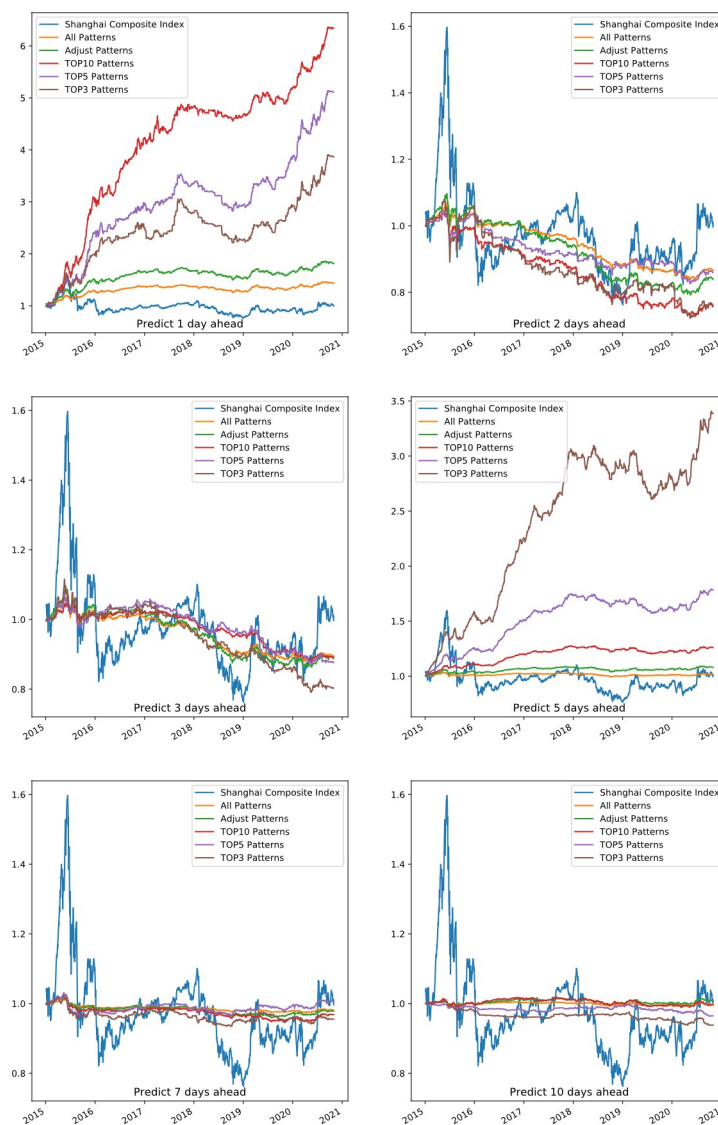


Fig 8. Portfolio return performance of two-day patterns predicting 1, 2, 3, 5, 7, 10 days ahead.

<https://doi.org/10.1371/journal.pone.0255558.g008>

TOP3 indicates that we use the top 3 most accurate patterns to invest. It seems that the two-day combination patterns have a certain prediction effect for 1 and 5 days. This is consistent with the higher number of selected patterns in forecasting 1 day and 5 days ahead than 2, 3, 7 and 10 days ahead in Fig 6. In the case of two-day patterns forecasting 5 days ahead, the performance difference among *TOP3*, *TOP5* and *TOP10* is obvious, indicating that its return is significantly affected by a certain pattern. Of the two-day pattern's prediction effect for six different periods, the 1-day prediction effect is the best. The very short-term forecasts may be related to China's emerging stock markets.

The finance performance of two-day candlestick patterns in predicting one day ahead is shown in Table 4. The average annual return of *TOP3*, *TOP5* and *TOP10* shows that the previous patterns are profitable. With a maximum annual return of 36.73% returns, five different portfolios based on the PRML model are more profitable than the market. During the same period, the market has a large drawdown of 52.28%, and our portfolio drawdown is smaller, thus indicating that the proposed model has less risk. The *All* and *Adjust* patterns show that using many patterns can reduce risk, but the corresponding profit will be reduced.

The portfolio average return of three-day patterns for forecasting 1, 2, 3, 5, 7, and 10 days ahead is shown as Fig 9. It shows that the three-day combination patterns have certain prediction effects only for 1 day. Among the prediction effect for six different periods, the one-day prediction effect is the best. This effect may be related to short-term market volatility.

Table 5 shows the finance performance of three-day patterns to predict one day ahead. The average annual returns of the five strategies show that the previous patterns are profitable when they predict one day ahead. With a maximum annual return of 8.29%, five different portfolios based on the PRML model are more profitable than the market. During the same period, the market has a large drawdown of 52.28%, and our portfolio drawdown is smaller, thus indicating that the proposed model has less risk. The *All* and *Adjust* patterns show that using many patterns can reduce risk, but the corresponding profit will be reduced.

In the *All* strategy, although the three-day candlestick combination patterns have a lower return than do the two-day candlestick combination patterns, the max drawdown of the three-day candlestick combination patterns has also decreased and has greater stability. Portfolios constructed by the top 10 three-day candlestick patterns have not obtained a return larger than the return obtained by the top 5 portfolios; however, more patterns can bring greater benefits in the two-day candlestick patterns because there is a fewer supporting number of three-day patterns in the training set and the machine learning model is prone to overfitting. However, adding more three-day combination patterns can effectively increase stability.

Although we have used all the stock data of Chinese listed companies for 15 years, the training data corresponding to each pattern of the two-day patterns and the three-day patterns are

Table 4. Finance performance of two-day patterns predicting one day ahead.

	SH Index	All	Adjust	TOP10	TOP5	TOP3
Average Annual Return	-0.66%	6.35%	10.75%	36.73%	33.76%	26.17%
Max Drawdown	-52.28%	-10.90%	-13.81%	-16.43%	-15.34%	-27.62%
Annual Sharpe Ratio	-0.19	-0.08	0.17	0.81	0.99	0.62
Information Ratio	0	1.94	2.38	2.37	2.02	2.26
Standard Deviation	1.48	0.34	0.49	0.78	0.84	0.96

* All represents a portfolio constructed by all patterns whose training accuracy is higher than 55%, Adjust represents a portfolio constructed by the patterns whose training accuracy is higher than 55% and whose support number is higher than 1000, TOP10 represents a portfolio constructed by the top 10 training accuracy patterns, TOP5 represents a portfolio constructed by the top 5 training accuracy patterns, and TOP3 represents a portfolio constructed by the top 3 training accuracy patterns.

<https://doi.org/10.1371/journal.pone.0255558.t004>

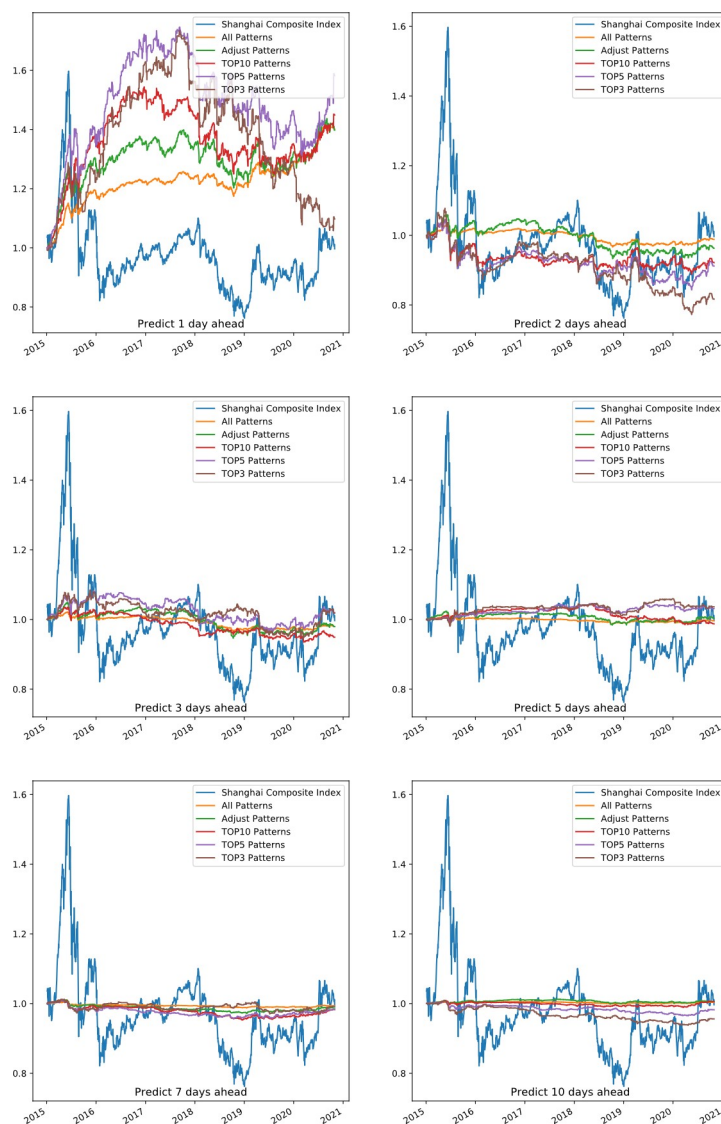


Fig 9. Portfolio return performance of three-day patterns predicting 1, 2, 3, 5, 7, 10 days ahead.

<https://doi.org/10.1371/journal.pone.0255558.g009>

different. Especially the three-day patterns which have as many as 2,197 combinations, the data of each pattern of the three-day patterns is still relatively small, which will easily cause the over-fitting of machine learning, leading to obvious differences in the prediction effects which are shown in Tables 4 and 5.

3.4 Prediction model dependence testing

In recent years, new developments in deep learning have allowed for multiple levels of abstraction. The deep learning models have performed well in speech recognition, text processing, etc. Following Krauss et al. (2017) and Yu et al. (2020), two popular deep learning models are applied to verify the dependence of the pattern recognition framework [22, 43]. The MLP and LSTM forecasting models are used in two-day patterns and three-day patterns to predict one day ahead. Following Patel et al. (2015) and Zhou et al. (2019), *Accuracy* and *F-measure* are

Table 5. Finance performance of three-day patterns predicting one day ahead.

	SH Index	All	Adjust	TOP10	TOP5	TOP3
Average Annual Return	-0.66%	5.95%	5.90%	6.62%	8.29%	1.74%
Max Drawdown	-52.28%	-6.63%	-14.05%	-19.10%	-24.90%	-39.00%
Annual Sharpe Ratio	-0.19	0.10	-0.2	-0.21	0.03	-0.22
Information Ratio	0	1.53	2.08	2.10	2.32	1.64
Standard Deviation	1.48	0.23	0.42	0.59	0.74	0.88

<https://doi.org/10.1371/journal.pone.0255558.t005>

used for screening and identifying patterns, and the qualified patterns will be automatically put into the strategy pool [3, 20].

The portfolio return performance of MLP and LSTM is shown in Fig 10. The performance of the Shanghai Composite Index during the same period was used as a benchmark, which is indicated by the blue line in the figure. The figure shows that MLP and LSTM have obviously prediction effects in terms of one-day prediction. Applying other machine learning model methods to pattern recognition can also obtain excess returns.

The best strategy's financial performance when using two-day and three-day patterns for one-day predictions is shown in Table 6. Among the strategies, with a maximum of 21.83% annual returns and 16.26% max drawdown, the TOP10 strategy using LSTM seems to obtain the best performance. The best-performing patterns during the training period are not profitable when using MLP to make one day predictions in two-day patterns; i.e., the stability is poor. Similar to the four traditional machine learning methods used in the previous PRML, MLP and LSTM are still profitable and have good prediction effects.

3.5 Further analysis

The performance of the PRML model on sliding window data is also tested. Table 7 shows the performance of two-day patterns predicting one day ahead. It can be seen from the table that our forecasting model still has a good forecasting effect on four different data intervals, and the strategies are all profitable with a smaller retracement rate than the index.

Finally, the effects of stock transaction costs, whose importance was emphasized by Park and Irwin [44] are examined. Caginalp and Laurent, and, Bessembinder and Chan note that the largest cost in the U.S. stock market is the bid-ask spread, which ranges from 0.1% to 0.39% [4, 8]. In the Chinese stock market, broker commissions range from 0.015% to 0.3%, and the stamp duty is 0.1% when selling shares. Therefore, we tested the two-day candlestick patterns' one-day-ahead predictions with a total transaction cost of 0.2%.

The validity of the prediction results is shown as Fig 11. Although the returns have decreased, the results are not qualitatively changed even when a higher transaction cost is considered. The finance performance of two-day candlestick patterns in predicting one day ahead with 0.2% transaction cost is shown in Table 8. With a maximum average annual return of 24.45%, four different portfolios based on the PRML model investment strategy are all more profitable than the market. Additionally, the max drawdown of our portfolio is smaller than the max drawdown of the market during the same period.

4. Conclusion

Forecasting the direction of the daily changes of stocks is an essential yet challenging task. The newest data mining and artificial intelligence methods can be used to improve the effectiveness of financial market forecasting. This paper attempts to develop PRML, a pattern recognition model that uses machine learning methods to improve stock trading decisions. Different time

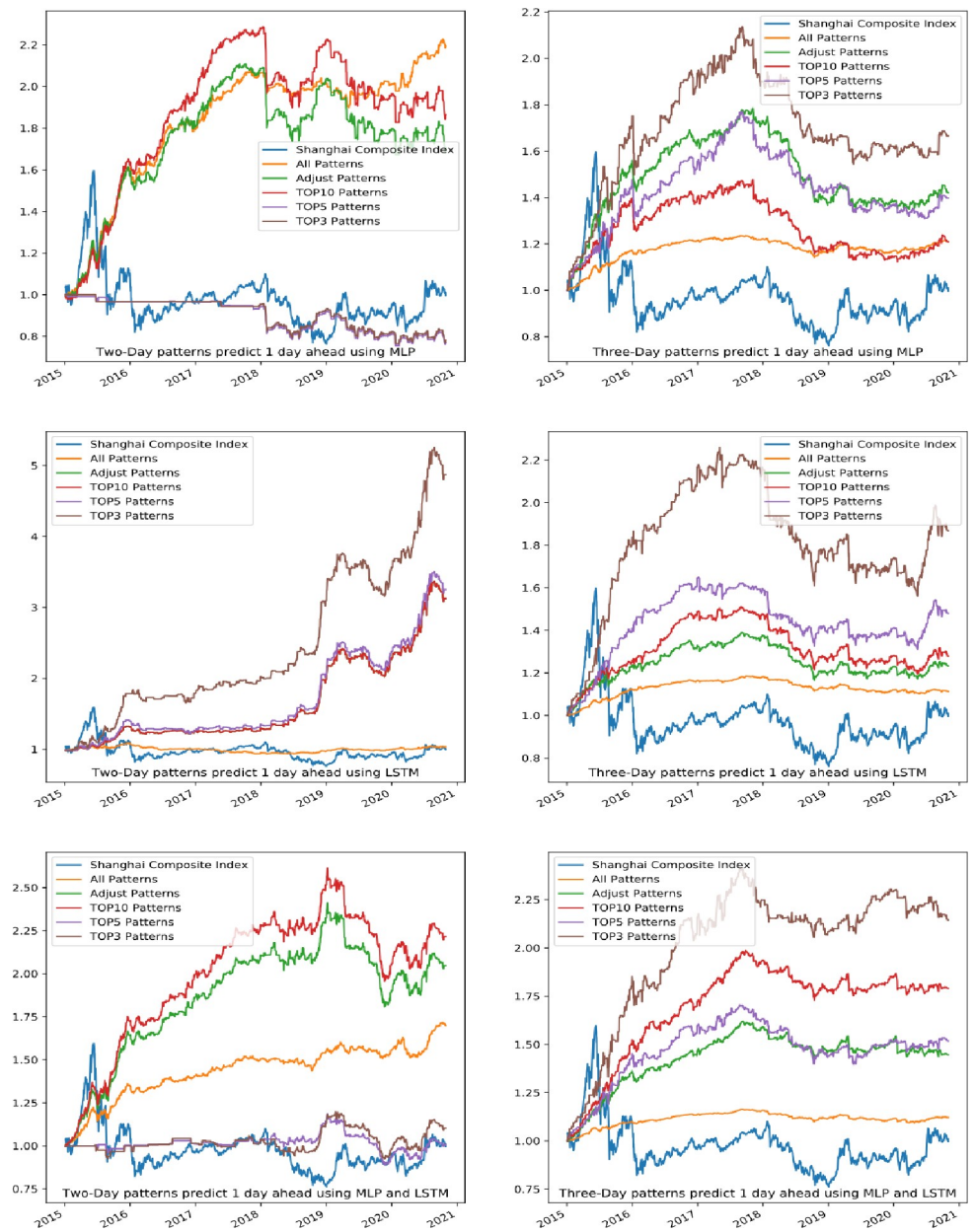


Fig 10. One day ahead forecasting performance of MLP and LSTM.

<https://doi.org/10.1371/journal.pone.0255558.g010>

Table 6. One day ahead performance of MLP and LSTM.

	SH Index	Two-day predictions one day ahead			Three-day predictions one day ahead		
		TOP10 (MLP)	TOP10 (LSTM)	TOP10 (MLP+LSTM)	TOP3 (MLP)	TOP3 (LSTM)	TOP3 (MLP+LSTM)
Average Annual Return	-0.66%	11.43%	21.83%	14.72%	9.16%	11.34%	14.01%
Max Drawdown	-52.28%	-20.03%	-16.26%	-25.10%	-27.70%	-30.91%	-14.87%
Annual Sharpe Ratio	-0.19	0.04	0.75	0.21	-0.42	-0.14	0.04
Information Ratio	0	2.16	1.00	2.26	2.51	2.49	2.49
Standard Deviation	1.48	0.64	0.75	0.57	0.78	0.76	0.75

<https://doi.org/10.1371/journal.pone.0255558.t006>

Table 7. Finance performance of sliding windows in two-day patterns predicting one day ahead.

Different training and testing sets	Benchmark	SH Index	All	Adjust	TOP10	TOP5	TOP3
Training: Jan 1,2001-Dec 31,2015	Average Annual Return	-0.46%	10.44%	3.72%	3.97%	7.42%	9.78%
Testing: Jan 1,2016-Oct 30,2020	Max Drawdown	-30.73%	-9.16%	-17.61%	-17.32%	-23.48%	-28.49%
	Standard Deviation	1.19	0.27	0.45	0.51	0.73	0.96
Training: Jan 1,2002-Dec 31,2016	Average Annual Return	0.73%	12.19%	1.19%	1.12%	4.36%	3.55%
Testing: Jan 1,2017-Oct 30,2020	Max Drawdown	-30.73%	-9.87%	-18.94%	-21.44%	-21.30%	-5.99%
	Standard Deviation	1.11	0.27	0.46	0.53	0.73	0.29
Training: Jan 1,2003-Dec 31,2017	Average Annual Return	-1.32%	25.16%	11.65%	13.06%	33.58%	12.39%
Testing: Jan 1,2018-Oct 30,2020	Max Drawdown	-30.73%	-3.10%	-14.64%	-16.27%	-22.14%	-30.26%
	Standard Deviation	1.25	0.28	0.58	0.65	1.03	1.32
Training: Jan 1,2004-Dec 31,2018	Average Annual Return	15.80%	45.04%	44.83%	51.92%	21.79%	0.59%
Testing: Jan 1,2019-Oct 30,2020	Max Drawdown	-18.66%	-1.67%	-3.84%	-6.92%	-11.11%	-3.79%
	Standard Deviation	1.25	0.28	0.46	0.58	0.65	0.35

<https://doi.org/10.1371/journal.pone.0255558.t007>

windows from one to ten days are used to detect the prediction effect at different periods. Empirical results show that the two-day candlestick patterns and three-day candlestick patterns have the best prediction effect when forecasting one day ahead. In general, the results do not qualitatively change even when a 0.2% transaction cost is considered. Two other popular machine learning models are used to test the dependence of the prediction model. The MLP and LSTM forecasting models also perform well when predicting one day ahead. The results show that an investment strategy constructed according to the PRML model can be profitable.

This study contributes to the literature in four ways: First, a candlestick pattern recognition model based on machine learning methods is proposed. The prediction results obtained by this model are more accurate than those obtained by purely using machine learning methods. Four machine learning methods—LR, KNN, RBM, and RF—are applied to all possible different combinations of daily patterns in the pattern recognition process. We incorporate the shape, location and nine commonly used technical indicators as features into each machine learning model to improve the accuracy of predictions.

Second, we divide the data set into two parts—the machine learning set and the prediction set—and use the trained parameters to perform prediction tests on new unknown data in the experimental stage. Conventional machine learning methods generally divide the data set only into a training set and a testing set. We added a prediction set to test whether the identified candlestick patterns still have a prediction effect on unknown data. We split the data set into the following two parts during the experimental testing phase of China's stock market: the machine learning data set from Jan 1, 2000 to Dec 31, 2014 and the prediction set from Jan 1, 2015 to Oct 30, 2020.

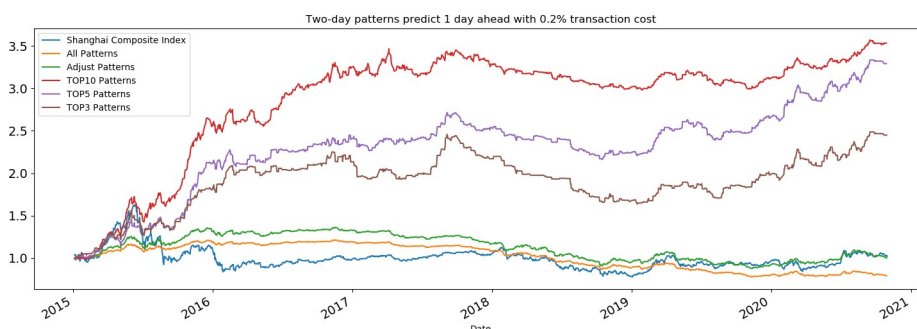


Fig 11. Portfolio return of two-day predictions one day ahead with 0.2% transaction cost.

<https://doi.org/10.1371/journal.pone.0255558.g011>

Table 8. Finance performance of two-day predictions one day ahead with 0.2% transaction cost.

	SH Index	All	Adjust	TOP10	TOP5	TOP3
Average Annual Return	-0.66%	-3.92%	0.01%	24.45%	23.02%	16.65%
Max Drawdown	-52.28%	-35.85%	-35.66%	-17.29%	-20.38%	-33.31%
Annual Sharpe Ratio	-0.19	-1.37	-0.71	0.32	0.50	0.25
Information Ratio	0	0.16	0.86	2.68	2.31	2.49
Standard Deviation	1.48	0.41	0.62	0.74	0.84	0.96

<https://doi.org/10.1371/journal.pone.0255558.t008>

Third, this paper examines the effects of predicting at different periods. Based on the two-day and three-day candlestick patterns identified by PRML, an investment strategy was constructed to dynamically build a daily investment portfolio. The average number of stocks invested in the portfolio is significantly less than that of pure ML methods. Compared with the pure ML methods, the PRML can effectively improve the accuracy of prediction, thereby further reducing the risk of uncertainty. For the number of patterns that exceed the accuracy threshold, the 1 day ahead forecasting is significantly higher than other forecasting periods. The strategy results also show that we can obtain the best performance when predicting one day ahead. These very short-term forecasting effects may be related to the characteristic of China's emerging stock markets. Compared to the -0.66% annual return of the market, all the identified two-day candlestick patterns yield an average annual return of 6.35%, and all the eligible three-day candlestick patterns have a 5.95% average annual return over the same period. Relative to the market drawdown, the max drawdown of the two-day candlestick pattern is 10.90%, and the max drawdown of the three-day candlestick pattern is only 6.63%.

Finally, five different strategy pools were used to form five comparable strategies to build daily portfolios dynamically. Five strategies (including *TOP3*, *TOP5*, and *TOP10*, which have the highest accuracy in the machine learning set), *All* patterns with an accuracy rate of more than 55%, and *Adjust* patterns that remove the candlestick patterns that occur fewer than 1000 times are used to dynamically build a daily portfolio. The results show that more two-day candlestick patterns can be profitable. After filtering, the two-day candlestick patterns have the best prediction effect when forecasting one day ahead and using the *TOP10* strategy, which, in this case, obtains an average annual return, an annual Sharpe ratio, and an information ratio as high as 36.73%, 0.81, and 2.37, respectively. With a maximum annual return of 8.29% from the *TOP5* strategy, three-day candlestick patterns after screening also present a profitable effect when forecasting one day ahead, thus showing more stable characteristics. During the same period, the market has a significant drawdown of 52.28%, and the max drawdowns of our portfolios are all less than 40%, thus indicating that the proposed model has less risk than the market.

5. Limitations and future work

Although we have used all the stock data of Chinese listed companies from 2000 to 2014 for learning, the number of each combination pattern is still relatively small. We hope to identify the candlestick patterns that have a certain frequency in the market. We did not identify more complex candlestick patterns in the empirical stage because more complex patterns require larger data sets. A more careful hyperparameter optimization of different machine learning methods, especially deep learning models, may still need to be considered in future research to get better prediction performance.

In future work, we will consider researching candlestick patterns in more mature markets. We also plan to utilize additional predictive factors, such as the newest deep learning methods and more technical indicators, to improve forecasting results.

Supporting information

S1 Data.
(ZIP)

S2 Data.
(ZIP)

Acknowledgments

We are truly grateful to an anonymous referee whose comments vastly improved this paper.

Author Contributions

Conceptualization: Yaohu Lin, Haijun Yang.

Data curation: Yaohu Lin, Harris Wu, Bingbing Jiang.

Formal analysis: Yaohu Lin, Haijun Yang.

Funding acquisition: Shancun Liu, Haijun Yang.

Methodology: Haijun Yang.

Software: Yaohu Lin.

Validation: Harris Wu, Bingbing Jiang.

Writing – original draft: Yaohu Lin.

Writing – review & editing: Shancun Liu, Haijun Yang, Harris Wu.

References

1. Fischer T, Krauss C. Deep learning with long short-term memory networks for financial market predictions. *Eur J Oper Res*. 2018; 270(2):654–69.
2. Malkiel BG, Fama EF. Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*. 1970; 25(2):383–417.
3. Zhou F, Zhang Q, Sornette D, Jiang L. Cascading logistic regression onto gradient boosted decision trees for forecasting and trading stock indices. *Applied Soft Computing*. 2019; 84:105747.
4. Bessembinder H, Chan K. Market efficiency and the returns to technical analysis. *Financ Manag*. 1998;5–17.
5. Marshall BR, Young MR, Rose LC. Candlestick technical trading strategies: Can they create value for investors? *J Bank Financ*. 2006; 30(8):2303–23.
6. Bulkowski TN. *Encyclopedia of Candlestick Charts*. 2008.
7. Hu W, Si Y-W, Fong S, Lau RYK. A formal approach to candlestick pattern classification in financial time series. *Applied Soft Computing*. 2019; 84:105700.
8. Caginalp G, Laurent H. The predictive power of price patterns. *Appl Math Financ*. 1998; 5:181–205.
9. Goo YJ, Chen DH, Chang YW. The application of Japanese candlestick trading strategies in Taiwan. *Investment Management and Financial Innovations*. 2007;(4, Iss. 4):49–79.
10. Lu T-H. The profitability of candlestick charting in the Taiwan stock market. *Pacific Basin Financ J*. 2014; 26:65–78.
11. Lu T-H, Chen Y-C, Hsu Y-C. Trend definition or holding strategy: What determines the profitability of candlestick charting? *J Bank Financ*. 2015; 61:172–83.
12. Zhu M, Atri S, Yegen E. Are candlestick trading strategies effective in certain stocks with distinct features? *Pacific Basin Financ J*. 2016; 37:116–27.
13. Chen S, Bao S, Zhou Y. The predictive power of Japanese candlestick charting in Chinese stock market. *Phys A Stat Mech its Appl*. 2016; 457:148–65.

14. Fock JH, Klein C, Zwergel B. Performance of candlestick analysis on intraday futures data. *The Journal of Derivatives*. 2005; 13(1):28–40.
15. Duvinage M, Mazza P, Petitjean M. The intra-day performance of market timing strategies and trading systems based on Japanese candlesticks. *Quant Financ*. 2013; 13(7):1059–70.
16. Yan D, Zhou Q, Wang J, Zhang N. Bayesian regularisation neural network based on artificial intelligence optimisation. *Int J Prod Res*. 2017; 55(8):2266–87.
17. Wang J-J, Wang J-Z, Zhang Z-G, Guo S-P. Stock index forecasting based on a hybrid model. *Omega*. 2012; 40(6):758–66.
18. Henrique BM, Sobreiro VA, Kimura H. Literature review: machine learning techniques applied to financial market prediction. *Expert Syst Appl*. 2019; 124:226–51.
19. Zhang Y, Wu L. Stock market prediction of S&P 500 via combination of improved BCO approach and BP neural network. *Expert Syst Appl*. 2009; 36(5):8849–54.
20. Patel J, Shah S, Thakkar P, Kotecha K. Predicting stock market index using fusion of machine learning techniques. *Expert Syst Appl*. 2015; 42(4):2162–72.
21. Gupta D, Pratama M, Ma Z, Li J, Prasad M. Financial time series forecasting using twin support vector regression. *PLoS ONE*. 2019; 14(3). <https://doi.org/10.1371/journal.pone.0211402> PMID: 30865670
22. Krauss C, Do XA, Huck N. Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the S&P 500. *Eur J Oper Res*. 2017; 259(2):689–702.
23. Subha MV, Nambi ST. Classification of Stock Index movement using k-Nearest Neighbours (k-NN) algorithm. *WSEAS Trans Inf Sci Appl*. 2012; 9(9):261–70.
24. Lee M-C. Using support vector machine with a hybrid feature selection method to the stock trend prediction. *Expert Syst Appl*. 2009; 36(8):10896–904.
25. Wu M-C, Lin S-Y, Lin C-H. An effective application of decision tree to stock trading. *Expert Syst Appl*. 2006; 31(2):270–4.
26. Pai P-F, Lin C-S. A hybrid ARIMA and support vector machines model in stock price forecasting. *Omega*. 2005; 33(6):497–505.
27. Frances PH, Marches M, Murray A. A hybrid genetic-neural architecture for stock index forecasting. *Information Science*. 2005; 17(1):3–37.
28. Kim K-j. Financial time series forecasting using support vector machines. *Neurocomputing*. 2003; 55(1–2):307–19.
29. Chen A-S, Leung MT, Daouk H. Application of neural networks to an emerging financial market: forecasting and trading the Taiwan Stock Index. *Computers & Operations Research*. 2003; 30(6):901–23.
30. Brownstone D. Using percentage accuracy to measure neural network predictions in stock market movements. *Neurocomputing*. 1996; 10(3):237–50.
31. Bao W, Yue J, Rao Y. A deep learning framework for financial time series using stacked autoencoders and long-short term memory. *PLoS ONE*. 2017; 12(7):e0180944. <https://doi.org/10.1371/journal.pone.0180944> PMID: 28708865
32. Liang Q, Rong W, Zhang J, Liu J, Xiong Z, editors. Restricted Boltzmann machine based stock market trend prediction. *International Joint Conference on Neural Networks (IJCNN)*; 2017: IEEE.
33. Zhang N, Lin A, Shang P. Multidimensional k-nearest neighbor model based on EEMD for financial time series forecasting. *Phys A Stat Mech its Appl*. 2017; 477:161–73.
34. Qiu J, Wang B, Zhou C. Forecasting stock prices with long-short term memory neural network based on attention mechanism. *PLoS ONE*. 2020; 15(1). <https://doi.org/10.1371/journal.pone.0227222> PMID: 31899770
35. Patel J, Shah S, Thakkar P, Kotecha K. Predicting stock and stock price index movement using Trend Deterministic Data Preparation and machine learning techniques. *Expert Syst Appl*. 2015; 42(1):259–68.
36. Moghaddam AH, Moghaddam MH, Esfandiyari MJJoEF, Science A. Stock market index prediction using artificial neural network. 2016:89–93.
37. Jabbarzadeh A, Shavvalpour S, Khanjarpanah H, Dourvash D. A multiple-criteria approach for forecasting stock price direction: nonlinear probability models with application in S&P 500 Index. *International Journal of Applied Engineering Research*. 2016; 11(6):3870–8.
38. Minh DL, Sadeghi-Niaraki A, Huy HD, Min K, Moon H. Deep learning approach for short-term stock trends prediction based on two-stream gated recurrent unit network. *IEEE Access*. 2018; 6:55392–404.
39. Kim K-j, Han I. Genetic algorithms approach to feature discretization in artificial neural networks for the prediction of stock price index. *Expert Syst Appl*. 2000; 19(2):125–32.
40. Tsai C-F, Hsiao Y-C. Combining multiple feature selection methods for stock prediction: Union, intersection, and multi-intersection approaches. *Decis Support Syst*. 2010; 50(1):258–69.

41. Breiman L. Random forests. *Mach Learn.* 2001; 45(1):5–32.
42. Booth A, Gerding E, McGroarty F. Automated trading with performance weighted random forests and seasonality. *Expert Syst Appl.* 2014; 41(8):3651–61.
43. Yu P, Yan X. Stock price prediction based on deep neural networks. *Neural Comput Appl.* 2020; 32(6):1609–28.
44. Park CH, Irwin SH. What do we know about the profitability of technical analysis? *Journal of Economic Surveys.* 2007; 21(4):786–826.