Neurocomputing 512 (2022) 363–380

Contents lists available at ScienceDirect

Neurocomputing

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

A survey on machine learning models for financial time series forecasting

Yajiao Tang [a,e], Zhenyu Song [b], Yulin Zhu [a,e], Huaiyu Yuan [a,e], Maozhang Hou [a,e], Junkai Ji [c,*], Cheng Tang [d,*], Jianqiang Li [c]

[a] College of Economics, Central South University of Forestry and Technology, Changsha 410004, China
[b] College of Information Engineering, Taizhou University, Taizhou 225300, China
[c] College of Computer Science and Software Engineering, Shenzhen University, Shenzhen 518060, China
[d] Department of Electrical and Mechanical Engineering, Nagoya Institute of Technology, Nagoya 466-8555, Japan
[e] Research Center of High Quality Development of Industrial Economy, Changsha 410004, China

ARTICLE INFO

Article history:
Received 30 April 2022
Revised 5 August 2022
Accepted 4 September 2022
Available online 20 September 2022

Keywords:
Financial Time Series
Forecasting
Machine learning
Hybrid method

ABSTRACT

Financial time series (FTS) are nonlinear, dynamic and chaotic. The search for models to facilitate FTS forecasting has been highly pursued for decades. Despite major related challenges, there has been much interest in this topic, and many efforts to forecast financial market pricing and the average movement of various financial assets have been implemented. Researchers have applied different models based on computer science and economics to gain efficient information and earn money through financial market investment decisions. Machine learning (ML) methods are popular and successful algorithms applied in the FTS domain. This paper provides a timely review of ML's adoption in FTS forecasting. The progress of FTS forecasting models using ML methods is systematically summarized by searching articles published from 2011 to 2021. Focusing on the analysis of ML methods applied to the theoretical basis and empirical application of FTS data forecasting, this paper provides a relevant reference for FTS forecasting and interdisciplinary fusion research against the background of computational intelligence and big data. The literature survey reveals that the most commonly used models for prediction involve long short-term memory (LSTM) and hybrid methods. The main contribution of this paper is not only building a systematic program to compare the merits and demerits of specific FTS forecasting models but also detecting the importance and differences of each model to help researchers and practitioners make good choices. In addition, the limitations to be addressed and future research directions of ML models' adoption in FTS forecasting are identified.

1. INTRODUCTION

The financial market is affected by a series of macroeconomic and microeconomic factors, such as the general economic situation, political events, the expectations of institutional investors, the choices of individual speculators, and the operation policies of commercial firms [1,2]. However, the exact effects of these factors on the dynamic financial system remain unknown, which makes financial data more difficult to be predicted under extrinsic uncertainties and intrinsic complexity [3,4]. FTS forecasting is an important research direction in the financial field, which predicts the possible risks, market trends, and entry and exit times for investing in the financial market [3]. With the modernization of financial transactions and information systems, a large amount of information is available for traders to make FTS predictions [5–7]. Even a slight improvement in the accuracy of financial market prediction may greatly benefit investors and speculators; thus, the ability to more accurately predict financial data movements and hedge against potential market risks is beneficial to both the financial institutions and individuals involved in financial transactions [8,9].

Various forecasting techniques have been proposed to address the issues corresponding to difficult feature extraction and low prediction accuracy in FTS forecasting, and several critical works have been implemented to confirm their predictability. Historical data are used to evaluate the profitability of different technologies. The literature concerning FTS forecasting is very rich in both theo-

* Corresponding author.
E-mail addresses: t20060857@csuft.edu.cn (Y. Tang), songzhenyu@tzu.edu.cn (Z. Song), zh-y-lin@126.com (Y. Zhu), yuanhuaiyu@163.com (H. Yuan), ruiruihou@163.com (M. Hou), jijunkai@szu.edu.cn (J. Ji), tang.cheng@nitech.ac.jp (C. Tang), lijq@szu.edu.cn (J. Li).

https://doi.org/10.1016/j.neucom.2022.09.003
0925-2312/

retical methods and practical applications. Among widely applied classic FTS prediction techniques, several methods have attracted particular attention, including fundamental analysis methods, which explore economic factors and determine the market trends [10], and technical analysis methods, which use a variety of technical indicators to exploit resistance and support standards. Technical analysis models can be divided into three categories: statistical models, ML models and hybrid models. Among them, ML techniques have been reported to realize successful implementations, particularly in derivatives pricing, risk management, and market prediction [11]. Recently, applications of ML have concentrated on applying different methods to learn useful features hierarchically from a large amount of data [12]. The main goal of these ML approaches is to model complex real-world data by extracting robust features that capture the relevant information [13]. Notably, deep learning (DL) has flourished in the past few years because it is considered an excellent prediction tool in various fields. Compared with traditional methods, DL can better explain the nonlinear relationships in financial data [14]. Therefore, an increasing number of prediction models based on DL technologies have been introduced in relevant conferences and journals.

To illustrate the latest progress in the FTS forecasting domain, we aim to not only summarize the latest academic and economic perspectives on novel ML models but also explore the intrinsic and effective features of popularly applied models to offer satisfactory choices for researchers and financial investors [15]. Although many scientific papers have proposed intelligent and expert systems, technologies and algorithms for FTS forecasting problems, only a few review articles can be found in relevant journals and conference proceedings in the past decade. We pay special attention to ML implementation in FTS prediction research, briefly introduce the existing research on FTS prediction based on ML, and provide some historical and systematic perspectives by investigating the relevant literature, focusing on specific scientific research topics. Nevertheless, the research scope of the available literature is broad. It is challenging and even impossible to thoroughly address all the published papers. To review the retrieved articles, we systematically selected the most relevant literature. Objective parameters were proposed as the indicators to select more relevant articles. Thus, the following articles were included in our survey: the most recently published articles, the most-cited articles, groups of articles with the greatest bibliographic coupling, articles with the closest relationship in a cocitation network, and articles tracing the knowledge trend in a given scientific discipline.

This paper collects information from citations and publication years and selects the most important financial market prediction articles released from 2011 to 2021. It searches for the latest technologies in this scientific field, systematizes information, and points out possible challenges for future research. This review does not contain survey papers focusing on specific financial application fields except forecasting research. The relatively published articles reveal that most investigations concentrate on stock price and index, as shown in Table 2. Therefore, we focus mainly on the application of these two kinds of FTS. In addition, this paper systematically summarizes the preliminary research on financial forecasting using ML technology, points out some challenges, and discusses some open problems and future practical research in this field. The main contributions of this paper can be summarized as follows: 1) ML approaches are verified as resulting in the most effective models for FTS forecasting; 2) widely applied state-of-the-art ML FTS forecasting models are described in a detailed manner listing their advantages and disadvantages, and proper models can be decided depending on the features of the dataset; 3) various evaluation metrics used in the ML prediction models are discussed and compared; 4) hybrid ML methods are more popularly considered in FTS forecasting than single methods; and 5) research gaps in this field are identified, and corresponding future research directions are provided.

The following section briefly presents a survey of the main ML techniques covered in the articles selected for this study from the Web of Science. The remainder of this paper is organized as follows: Section 2 presents a general review of FTS forecasting. Section 3 provides an overview of the related literature we cover. The major findings in common evaluation metrics of ML methods are mentioned in Section 4. Section 5 sorts widely applied ML models, such as ANN, RF, DT, SVM, LSTM and CNN. In addition, this section summarizes the top 10 most-cited papers on the best practices of applying ML models to FTS prediction. Section 6 presents conclusions and key points of open issues that need further study.

2. FTS forecasting research review

The FTS prediction model has been widely implemented in financial trend forecasting for decades. As FTS are rich in uncertainty and noise [16], the correct assessment of future movements in the financial environment determines the profits and losses that investing organizations and individuals can gain. To analyze and predict financial market behaviour, fundamental analysis and technical analysis are commonly used. The taxonomy of the widely applied technical analysis methods for FTS forecasting is illustrated in Fig. 1.

In the big data era, digesting a massive amount of news and data from various channels to predict a market is a challenge of fundamental analysis. Based on historical financial statements and current conditions, fundamental analysis used to be the mainstream of FTS research for a period of time. It studies the economic factors that may affect the market trend, which is most suitable for the relatively long-term prediction spectrum. The influencing factors can be any information in a company and its departments, including microeconomic indicators such as the cash flow ratio and return on investment or macroeconomic indicators such as GDP and consumer price index [17,18]. In addition, basic data under research also include unstructured text data, such as global news articles, information from web boards, and company information disclosures. Although fundamental analysis has its strengths, it lacks accuracy because it requires financial efficiency [19].

On the other hand, when studying the price and trend changes in specific financial assets, technical analysis does not explicitly consider the internal and external characteristics [17,20]. The prices of financial assets are believed to already include all the fundamental factors that affect them [21], which can be abstracted without analyzing all the macro and micro economic factors. The technical index is a mathematical formula applied to price or trading volume data, which is used to model some aspects of data movement in the financial market [22]. These techniques construct a linear or nonlinear pattern to approximate the underlying function using some data as the training sample. The technical analysis usually adopts various charts, the relative strength index, the moving average, the balanced trading volume, the momentum and change rate, and the direction movement index to perform the trend analysis [23,24]. Based on these techniques, technicians usually model the historical behaviour of specific financial assets as a time series rather than analyzing subjective economic factors. They believe that history will repeat itself as historical data are distributed sequentially in time. Thus, future movements can be predicted. Therefore, the price, trading volume, trading breadth and trading activity mode can be determined because this information is enough to determine the future trend.

In technical analysis, the statistical method is a traditional modeling technology widely used in the financial market [25,26].

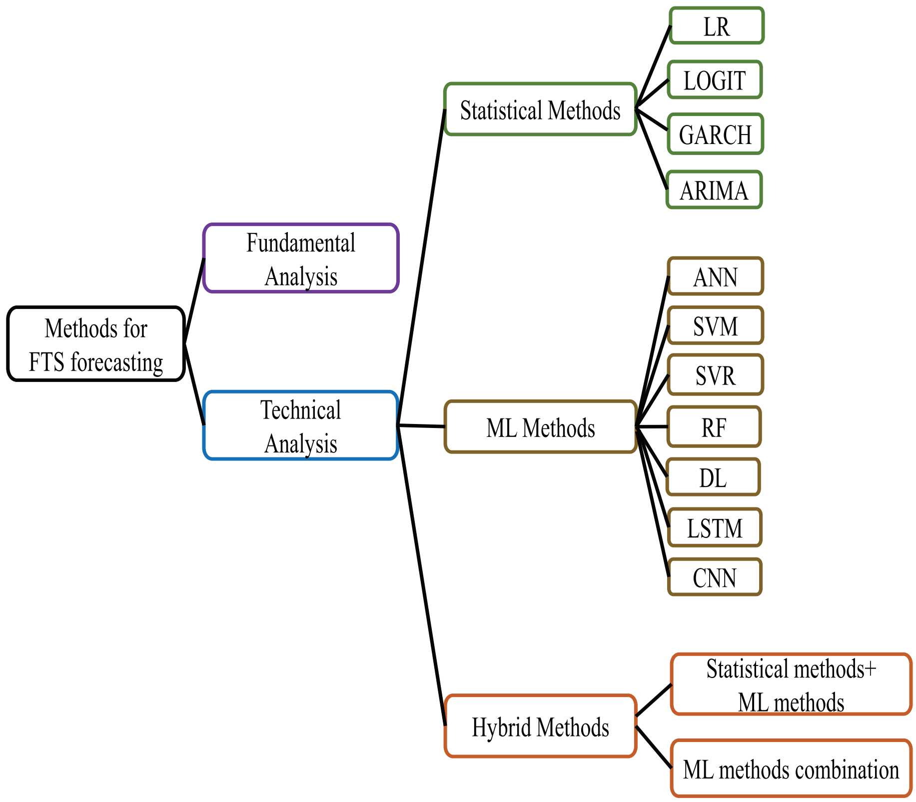

Fig. 1. Taxonomy of FTS forecasting methods.

Generally, statistical models can track patterns through historical information and data for FTS prediction. Classic statistical models have been developed, especially for linear and stationary processes [27]. Based on the assumption that linear processes generate time series, the essential processes are modelled to predict future movements. Statistical methods include mainly linear regression (LR), logistic regression (LOGIT), generalized autoregressive conditional heteroscedasticity (GARCH), and autoregressive comprehensive moving average (ARIMA) [28]. Although the modeling idea of statistical methods is relatively simple, it can achieve the prediction and analysis of small sample data. However, complex and noisy real-time series data cannot be reflected by the analytical equation containing parameters because the dynamic equation of time series data is either too complex or even unknown [29,30]. In addition, these specific characteristics of FTS data are an indispensable part of FTS data prediction. With the rapid spread and development of mobile internet, big data and other technologies, especially artificial intelligence, a large amount of financial data continue to be generated, and the correlation mode between data is becoming increasingly complex. FTS data are becoming increasingly difficult to predict. Although ARMA, GARCH and other statistical models can consider the sequence-dependent characteristics of FTS data, the determination of this factor either needs to be determined through complex econometric tests; otherwise, it is difficult to realize automatic identification based on the research problems. Even the application of these models may produce significant errors [31]. Therefore, the traditional model has limitations in forecasting FTS data with complex characteristics. There is a controversy that approaching problems encountered when we search for the most proper analysis model would limit our imagination and result in losses in accuracy and interpretability of the settlement [32]. Nonparametric methods that are not constrained by statistical assumptions are needed. Thus, nonlinear, soft computing models may resolve this dilemma.

Therefore, ML algorithms have quickly attracted many experts, who have conducted in-depth research in many aspects, and many academic results have been obtained in the past two decades. The initial concept of ML appeared in the 1950s, and since the 1990s, it has been studied as an independent field [33]. ML explores how to improve the performance of the model by means of calculation and learning experience. As a highly adaptive branch of artificial intelligence, ML relies on advanced, data-driven forecasting methods, implements simulations such as human learning, builds algorithms that perform learning and prediction [34,35], and aims to automatically learn and recognize patterns in large amounts of data [36]. Its mechanism is that the system architecture changes widely, and the interaction process changes with changes in the architecture [37]. Without knowing the input data in advance, many ML technologies actively capture the nonlinear relationship between relevant factors and extract patterns by learning historical data in the process of training to predict new data [38,38–40]. The ML model can deal with complex information without model specification, even if the functional form of data is uncertain [41]. Empirical research using ML usually includes two main stages. The first stage is to solve the selection of prediction-related variables and models, separate the data according to a specific percentage, use it for model training and verification, and later optimize the model. In the second stage, the optimized model is applied to test the data to measure the prediction performance.

Artificial neural networks (ANNs), support vector machines (SVMs), support vector regression (SVR), random forest (RF), deep learning (DL) models, such as long short-term memory (LSTM) and convolutional neural networks (CNNs), are the basic ML techniques widely applied in the relevant literature. Among them, ANNs and SVMs have demonstrated their capability and comparative advantages in financial modeling and forecasting [42–45]. ANNs were once the most commonly used methods in financial prediction. As a well-known computational model for TSF estimation [46], one of the distinctive advantages of ANNs is their robustness to noisy and erroneous data [47]. SVM is also verified to be very suitable for FTS prediction with advantages in addressing the problems of nonlinearity and local minima with small samples [48–51,48,52]. In general, different from technical analysis models, ML models are more nonparametric, nonlinear and data-driven, which can effectively deal with the nonlinear and nonstationary characteristics of FTS data, make it easier to mine the hidden patterns among them, and realize the cognition of the unknown relationship. Therefore, applying the ML algorithm to FTS forecasting has more advantages. However, ML algorithms have some defects in addressing with the sequence-dependent characteristics of FTS data. To incorporate these characteristics of time series data into the prediction model, the input characteristics need to be manually set, which will undoubtedly cause a certain subjectivity and limitations. In addition, ML has great limitations in processing complex high-dimensional data, and it is easy to encounter many problems, such as dimension disaster and invalid feature representation [53]. Therefore, ML algorithms such as the ANN and SVM have some problems in predicting FTS data with big data. Although many surveys concerning ML implementations in FTS forecasting exist, different ML techniques have not yet been surveyed comprehensively.

In recent years, among ML techniques, DL has achieved great success in fields such as speech processing and image recognition due to its strong feature extraction and complex data processing abilities. Therefore, DL has attracted the most overall interest. DL is a representation learning method with multilevel representation, which is obtained by superimposing multiple simple but nonlinear modules. From the input layer, each module converts the representation of each layer into a more abstract representation of the next layer. As long as there are enough such transformations, even very complex functions can be learned [54]. While improving the prediction accuracy of data within the sample, it is easier to alleviate the overfitting problem [55], which is very helpful in mining the complex internal characteristics of FTS data. Applying DL to FTS forecasting helps to extract the important information from time series data, to improve cognitive ability in the financial market, and to realize interdisciplinary integration research. Thus, DL has become a research hot spot in the FTS forecasting field and is undergoing rapid development.

When one intelligent paradigm is selected for a problem, the other mechanisms or different architectures of the same mechanism are excluded. To solve this problem, a hybrid mechanism combining single solutions of different methods is proposed to overcome the shortcomings of independent methods [56]. The combination of single solutions will reduce the uncertainty and randomness of parameter adjustment in the training stage [57]. Two or more intelligent methods can solve problems through their respective abilities to improve the overall forecasting performance of the whole group. Although intelligent computer systems have competitive advantages, traditional statistical methods cannot be completely abandoned for effective solutions. In related research, statistical approaches together with ML techniques have been frequently adopted in predicting markets [58,59]. In 2003, Zhang [60] first established a hybrid model that combined an ARIMA model with an SVM, which effectively improved forecasting accuracy when applied to real datasets. Later, inspired by F. Takens's theorem, Ferreira et al. provided a hybrid model including an ANN and a genetic algorithm (GA), in which the evolutionary search was adopted to determine the spatial features of the time series and a satisfactory result was concluded [61]. Hassan et al. combined an ANN, a GA and a hidden Markov model (HMM) to predict financial market behaviour in [62]. In [63], Zeng et al. proposed a forecasting approach combining empirical mode decomposition (EMD) and an ANN and applied it to the Baltic Dry Index. Chen et al. combined a self-organizing map approach with SVR and validated it on seven stock market indices [64]. Studies reveal that hybrid prediction approaches can improve prediction performance by overcoming the shortcomings of single models, reducing model uncertainty and enhancing generalization ability simultaneously. Mcdonald et al. studied the effectiveness of many ML algorithms and their integration through one-step prediction of a series of financial data and found that the hybrid model composed of a linear statistical model and a nonlinear ML algorithm is effective in predicting the future value of sequence data, especially in the future direction of the sequence [65]. Nevertheless, these models also increase the time complexity and sacrifice computational efficiency [66].

3. Overview

Forecasting a particular FTS is a popularly studied topic. The FTS prediction model and its related applications are an essential and profitable field that has been widely discussed in recent decades. The literature review investigates the research approaches and evaluation models of this topic to explore the latest developments and highlight future research directions [67]. This section provides an overview of the related literature that we will review in this study.

In this survey, in addition to forecasting research, there are no papers emphasizing specific areas of financial application. The papers discussed in this review are retrieved from Web of Science digital libraries for the specific period, and the articles selected as representative samples are validated. We adopted the publication period from 2011 to 2021 because the review articles were brief during this time interval. To maintain the diversity of our survey, our survey includes many review publications that summarize financial studies. In addition, it includes published books on stock market forecasting, trading system development, foreign exchange and market forecasting application examples. In this survey, all the works are searched and collected from the Web of Science, with financial market, ML, FTS, forecasting, prediction and computational intelligence as the keywords. The criterion for including a preliminary study in this survey is that the reference paper should recommend using any ML technology to solve the problem of FTS forecasting. The articles are reviewed and classified according to the test market adopted as the dataset, prediction variables, learning or training methods, and evaluation metrics used in the comparison. Notably, the results of this review do not involve all applications of ML methods in financial market prediction. This survey does not include papers that are not published in distinguished journals and conference proceedings, lack an adequate description of the adopted methods, or are weak in evaluation metrics. The articles cited in this paper are organized according to their primary objectives or their main contributions to the ML methods applied in the FTS forecasting literature. Most of the covered papers (129 of 146) were published in the past six years (2016–2021), as presented in Fig. 2. According to the number of papers, the top source journals' rank is shown in Table 1.

Table 3 shows the main FTS forecasting methods adopted in related articles from 2011 to 2021. Statistical approaches have been applied in 53 articles, but they are mostly adopted as the baseline or a component of a specific hybrid method. Among all the primary FTS forecasting methods in the research period, LSTM is the most widely applied method, although it has become a forecasting tool and research hotspot since 2017. In 2020, 37 articles focused on FTS forecasting by LSTM. The second commonly applied method is the hybrid method, which verifies the deduction that the hybrid method can combine the merits of individual methods and improve the prediction performance. SVM and ANN sequentially

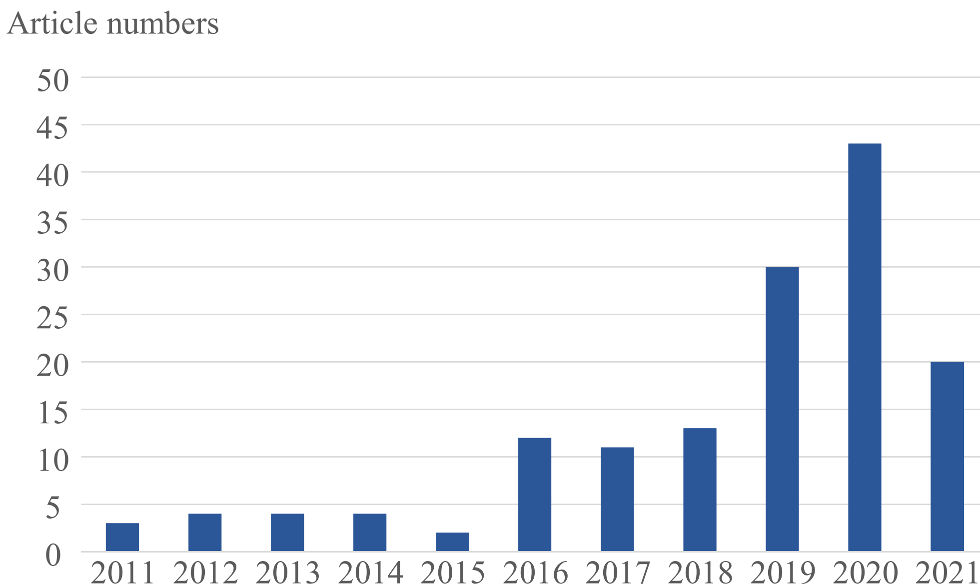

Fig. 2. Yearwise distribution of the papers.

Table 1
The 20 journals with the highest number of articles in the database searched.

Journal	Number of articles
Expert Systems with Applications	22
IEEE Access	14
Applied Soft Computing	12
Neural Computing and Applications	5
Applied Intelligence	4
Neurocomputing	4
Soft Computing	4
Applied Sciences (Basel)	3
Computational Economics	3
Economic Computation and Economic Cybernetics Studies and Research	3
International Journal of Computational Intelligence Systems	3
Concurrency Computation Practice and Experience	2
Entropy	2
European Journal of Operational Research	2
International Journal of Machine Learning and Cybernetics	2
Knowledge-Based Systems	2
Neural Processing Letters	2
Physica A: Statistical Mechanics and its Applications	2
Quantitative Finance	2
Sustainability	2

follow them as the No. 3 and No. 4 most widely applied FTS methods. Fig. 3 shows that LSTM ranks first, followed by the hybrid approach, statistical approach, SVM, ANN, SVR, CNN, RF, DT, LOGIT, evolutionary computation (EC), RNN and GA. Among them, LOGIT, EC and RNN have the same adoption publishment. In general, ML models are the mainstream methods in FTS prediction.

Table 3 summarizes the primary datasets adopted for FTS prediction from 2011 to 2021. It is evident that stock indices are the main means for evaluating the financial market, as the total number of articles was 324, and the most popular stock indices under analysis and forecasting are the S&P with 67 articles. This dataset is followed by the Shanghai Composite Index, HSI, NASDAQ Index, DAX, NIKKEI, TAIEX, NASDAQ Composite Index, and SZSE Index, with 48, 37, 35, 34, 26, 24, 19, 12 and 9 publications, respectively. The number of articles aiming to forecast stocks' movement in different stock exchange markets was 217. Among the popularly applied stock datasets, S&P500 has attracted the most attention, with 94 articles. It is also the top FTS dataset that was analyzed in the recent decade. NASDAQ, SSE, SZSE and KOPSI are the remaining hot stock datasets following S&P500. The numbers of articles forecasting the movement of foreign exchange rates and commodity prices are 35 and 20, respectively. We can conclude that the research focuses are still on stock index movements, which are traditional financial tools. Among all the datasets under forecasting and comparison, the financial data from the developed countries have absorbed the most attention. However, with the development of emerging economies, scholars have paid increasing attention to emerging markets in recent years, as we find that the Shanghai Composite Index accounts for the top 3 datasets under analysis, accounting for 48 articles. Typically, the S&P500 and S&P500 indices are traded in the U.S. market, which is the most developed financial market, and have attracted the most interest, while the SZSE, SSE and Nikkei 225 indices are from developing and underdeveloped financial markets. Notably, these findings may reveal the fact that financial data in developed financial markets may be less noisy, and prediction models can perform better in a complete and mature financial system environment. Therefore, developed markets have attracted more research attention.

4. Description of model performance evaluation metrics

FTS forecasting studies can be categorized into two primary groups on the basis of their expected outputs: price prediction and price movement (or trend and volatility) prediction. In most cases of FTS forecasting, correct future value prediction is not treated as important as accurate direction movement classification, although FTS forecasting is generally a regression problem. Therefore, researchers consider trend prediction, which forecasts how the price change is more crucial than exact price prediction. In that sense, trend prediction turns into a classification problem. In most studies, only upwards or downwards movements are taken into account (2-class problems), although 3-class problems also exist (up, down, or neutral movements). To measure the performance of different FTS forecasting models, some metrics have been frequently implemented [68–70].

Table 2
FTS prediction methods adopted in relative studies from 2011 to 2021.

	Number of articles											
Year	2021	2020	2019	2018	2017	2016	2015	2014	2013	2012	2011	Total
Statistical approaches	4	11	9	6	1	6	4	1	6	1	4	53
SVM	3	9	9	6	0	3	4	1	5	0	2	42
SVR	1	3	7	2	2	2	2	2	4	3	0	28
ANN	3	7	8	7	1	5	0	1	2	0	2	36
RF	3	5	1	3	1	1	1	0	0	0	0	15
DT	1	5	0	0	2	0	2	0	1	0	0	11
LOGIT	1	0	1	1	0	0	0	0	0	0	0	3
EC	1	0	1	1	0	0	0	0	0	0	0	3
GA	0	0	0	0	0	0	0	0	2	0	0	2
LSTM	27	37	9	5	2	0	0	0	0	0	0	80
CNN	12	12	1	1	0	0	0	0	0	0	0	26
RNN	0	2	0	0	0	0	0	0	0	0	1	3
Hybrid methods	13	24	9	3	3	7	1	4	2	6	2	74

Table 3
FTS datasets adopted in related studies from 2011 to 2021.

	Year											
	Number of articles											
Applied data set	2021	2020	2019	2018	2017	2016	2015	2014	2013	2012	2011	Total
S&P500	13	18	12	10	6	11	6	6	4	7	1	94
KOPSI	1	3	0	2	1	3	2	2	0	1	1	16
NASDAQ	11	8	6	6	3	5	6	2	6	5	2	60
SZSE	2	4	1	1	3	2	1	0	2	2	1	19
SSE	1	8	1	2	4	3	2	1	2	2	2	28
S&P500 Index	9	13	8	8	5	7	5	6	2	3	1	67
KOPSI200	1	1	0	1	0	0	0	1	0	1	0	5
NASDAQ Index	5	4	3	4	2	3	4	0	5	3	1	34
NASDAQ Composite Index	4	0	1	0	1	1	2	0	1	2	0	12
SZSE Index	1	1	0	1	1	1	0	0	2	2	0	9
Shanghai Composite Index	5	3	5	3	2	5	8	8	4	4	1	48
HSI	2	6	2	6	2	6	2	5	3	1	0	35
DAX	3	4	2	0	5	3	5	2	0	1	0	26
NIKKEI	5	5	1	1	2	2	3	2	0	2	1	24
DJIA	8	5	1	1	2	3	6	2	3	3	2	37
BOVESPA	0	1	0	1	2	1	0	0	0	0	0	5
ISE	1	0	0	0	0	1	0	0	0	0	1	3
TAIEX	1	0	2	1	2	3	2	3	1	3	1	19
FOREX	8	4	6	1	6	3	2	2	1	1	1	35
commodity price	1	10	1	2	3	2	1	0	0	0	0	20

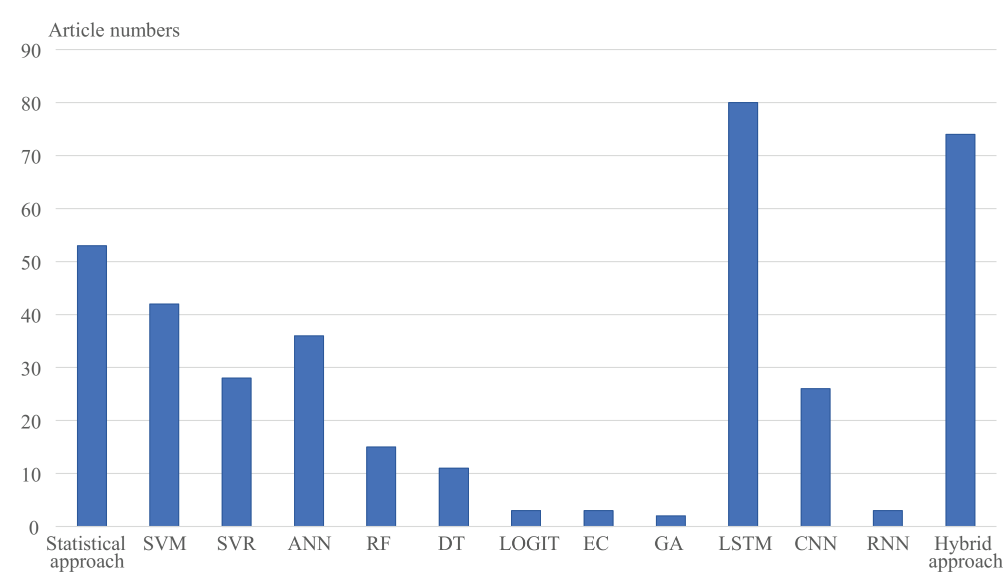

Fig. 3. Article numbers concerning the main prediction methods deployed in FTS.

4.1. Regression metrics

For stock price and stock index prediction, which are modelled as regression problems, regression metrics such as the mean squared error (*MSE*), mean absolute error (*MAE*), mean absolute percentage error (*MAPE*), root mean absolute error (*RMAE*), normalized MSE (*NMSE*), root mean squared error (*RMSE*), normalized RMSE (*NRMSE*), mean absolute percentage error (*MAPE*), root mean squared relative error (*RMSRE*), maximum positive prediction error (*MPPE*), maximum negative prediction error (*MNPE*), and correlation exponent of prediction (*R*) are commonly applied to measure model performance. *MSE* represents the distance between the actual data point and the fitted curve, while *RMSE* represents the standard deviation of the error to obtain the wrong outlier. *RMSE* can be calculated by taking the square root of *MSE*. *NRMSE* stands for normalized *RMSE*, which aims to more accurately evaluate the prediction results by applying normalization techniques such as the mean and standard deviation on *RMSE*. *NRMSE* is an unbiased quantity because it can calculate the predicted values of overfitting and underfitting. *MPPE* and *MNPE* are used to calculate the maximum positive and negative error boundary, respectively, to determine the accuracy. Among these metrics, four are widely used to measure the model prediction capability, namely, *MSE*, *MAE*, *MAPE*, and *R*. They are described in the following.

- The *MSE* of the predictor for normalized data can be given as:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(O_i - T_i)^2. \tag{1}$$

- The equation of the *MAE* is:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|T_i - O_i|. \tag{2}$$

- The *MAPE* of the prediction is:

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{T_i - O_i}{T_i}\right|. \tag{3}$$

- The *R* of the training data and the prediction data can be obtained by:

$$R = \frac{|\sum_{i=1}^{n}(T_i - \overline{T}_i)\left(O_i - \overline{O}_i\right)|}{\sqrt{\sum_{i=1}^{n}(T_i - \overline{T}_i)^2\sum_{i=1}^{n}\left(O_i - \overline{O}_i\right)^2}}, \tag{4}$$

where T_i denotes a teacher signal, O_i represents the output of the employed prediction model, and n stands for the number of data samples.

It can be concluded that the lower the prediction error is, the better the prediction model performs. For *R*, the larger the *R* value is, the better performance the prediction model achieves. However, since the difference between the *MSE*s may not be significantly different from zero, a lower *MSE* does not completely symbolize superior forecasting ability. Therefore, it is important to check whether any reduction in *MSE*s is significantly different from zero instead of comparing the *MSE*s of various forecasting models [71]. The same holds for *MAPE*, *MAE* and *R*. In addition, to test the significant differences among the results, a nonparametric test named the Wilcoxon rank-sum test [72] is widely used in this study scope. Some reviews have summarized that it is preferable to use a nonparametric statistical test instead of a parametric test to realize statistical accuracy, especially when the size of the sample dataset is small [73].

4.2. Classification merits

In most cases, the FTS forecasting problem is mapped as a classification task to determine the class of price variation [74]. Financial trading strategies are evaluated by observing their historical trading track records, so the returns are mandatory, which in turn represent the clear and simple performance of an investment indicator of interest for financial traders. Commonly applied classification efficiency metrics include accuracy (which stands for the correct prediction numbers of the movement), sensitivity, specificity, precision, recall, F1 score, Macro-average F-score, correlation coefficient (which denotes a discrete case of the Pearson correlation coefficient), average AUC score (which denotes the area under receiver operating characteristic curves) [75], Theil's *U* coefficient, hit ratio, average relative variance, and so on. Confusion matrices and boxplots are also applied for classification performance evaluation [76–78].

In general, most performance metrics attempt to evaluate how well the learned model can predict the correct class to which the new input samples belong. However, different metrics demonstrate different orderings in assessing model performance [79]. In the literature under review, classification accuracy is the most frequently adopted indicator for performance evaluation [80]. Both sensitivity and specificity are supposed to be straightforward indices. Although the AUC is not emphasized in terms of equal misclassification costs, it treats the most preferred score as an empirical probability treatment. That is, the randomly selected positive samples rank higher than the negative samples. Therefore, accuracy, sensitivity, specificity and AUC usually constitute the performance evaluation system.

$$Accuracy\ rate = \frac{TP + TN}{TP + TN + FP + FN}(\%), \tag{5}$$

where TP, TN, FP and FN indicate true positive, true negative, false positive, and false negative, respectively. True positive (TP) means that the sample is detected as positive, and the teacher label is also positive. True negative (TN) denotes that the input and the teacher target label are simultaneously negative. A false positive (FP) indicates that the input is positive, whereas the teacher target label is negative. A false negative (FN) shows that the input is negative, but the teacher target label demonstrates the opposite.

$$Sensitivity = \frac{TP}{TP + FN}(\%), \tag{6}$$

$$Specificity = \frac{TN}{TN + FP}(\%), \tag{7}$$

$$AUC(\%) = \frac{1}{2}\left(\frac{TP}{TP + FN} + \frac{TN}{TN + FP}\right) \times 100(\%), \tag{8}$$

where the sensitivity and specificity are defined as the true positive rate and true negative rate, respectively. Sensitivity measures the percent of real positive samples that are distinguished correctly. This metric shows how successfully a classifier performs by correctly identifying regular records. Therefore, whether a classifier is correct and efficient can be defined by whether it can achieve a higher sensitivity. Specificity represents how successfully a classifier can identify the proportion of true negatives, denoting abnormal records. Hence, a higher specificity means a reduction in the possibility of misclassification. A score of 100% AUC indicates that the classifier can correctly discriminate the two categories, whereas a score of 50% indicates that the two classes cannot be distinguished significantly and correctly.

Nevertheless, higher accuracy achieved in predictions does not imply higher profits in most cases. Even though the strategy is supposed to be efficient and able to make a profit on paper, if the returns accumulated by successive trades cannot exceed the associated trading costs, which conclude commissions, spreads and slippage, the strategy will eventually fail. Thus, a series of accuracy variants are adopted to evaluate the efficiency of the prediction model. Actually, the accumulated total return over a specific period is an adequate metric to execute this type of study. Cumulative return considers the following accuracies of the prediction, which can trigger successive trades. Historical trading track records are important criteria for observing financial trading strategies. The returns represent the performance of the investment indicator preferred by financial traders simply and clearly.

In addition to accuracy, the results of transaction simulation are also assessed by the following indicators. $Up_{Accuracy}$ (%): Percentage of possible accurate Up predictions during the next trading session. $Down_{Accuracy}$ (%): Percentage of possible accurate Down predictions during the next trading session. $Long_{Accuracy}$ (%): Long accuracy connects with the $Up_{Accuracy}$ since a buy order (long trade) comes from an Up prediction. It can measure the positive profit percentage of those trades without considering trade costs. $Short_{Accuracy}$ (%): Short accuracy connects with the $Down_{Accuracy}$ since a sell order (short trade) comes from a Down prediction. It can measure the positive profit percentage of those trades without considering trade costs. $CumulativeReturn$ (%): Percentage of return calculated at the end of the trading cycle. $CumulativeReturnForThei^{th}Period$: Cumulative $Return_i = (1 + CumulativeReturn_i)(Return_i) - 1$. Cumulative return takes into account the underlying accuracy of the predictions that triggered the series of trades. In addition, it represents a weighted average of the success rate, as the positive predicted trades are associated with the individual profit obtained. $MaximumDrawdown$: The historical maximum peak-to-bottom decline of the value counted by a specific equity or trading strategy uses cumulative returns to trace the movements. It indicates the maximum accumulated loss experienced during trading. $AverageReturnPerTrade$: The average return obtained in all the trades counting at the end of the trading cycle.

To the best of our knowledge, most papers emphasize constructing efficient prediction models which are evaluated by the metrics Accuracy and/or *MSE*, which shows that FTS forecasting is treated as either a classification problem or a regression problem.

5. Analysis of primary ML studies

5.1. Review of different ML forecasting models

This section summarizes the structure, mechanism, application, advantages and disadvantages of different ML forecasting models.

5.1.1. Artificial Neural Network

Generally, an artificial neural network (ANN) is a flexible model to deal with nonlinear time series. Through error backpropagation technology, we can find the dependency and correlation between the attributes of the dataset for prediction. As a bioinspired algorithm, the ANN model is based on the brain's central nervous system, which is usually composed of three layers, namely, the input layer, hidden layers, and output layer. Fig. 4 shows a typical artificial neuron. The input layer specifies the nodes with several input features. FTS data with random attribute weights are the input data of training in the prediction mechanism. Then, the hidden layer processes the input data at various levels. Through the input operation, each neuron is either inhibited or excited, and the artificial neuron will show the corresponding state, namely, positive or negative. A group of hidden layers jointly analyze and convert the input value and predict the output value through the activation function, which can control the firing of artificial neurons. ANN models can realize high prediction accuracy by adding various transformation technologies into the hidden layer. The increase in the number of hidden layers will increase the operation complexity and improve the reliability and correctness of the results, and vice versa. In the hidden layer, each attribute is mapped to all predefined constraint models for operating and knowledge extraction. Each layer performs input data conversion, and the final output layer displays the final expected data value obtained from hidden layer processing. During the training process, the ANN learns from all the valid results generated from the effective models and utilizes the knowledge for future prediction. Then, the predicted outputs are compared with the actual values to confirm their accuracy and reliability. The model is considered valid if the results are within an acceptable relevance interval. Otherwise, the process will be implemented again with updated weights through the backpropagation model. This process will continue until the model can generate acceptable outputs.

Different from statistical models, ANNs can deal with multivariate processing flow. Backpropagation neural networks (BPNNs), multilayer perceptrons (MLPs), extreme learning machines (ELMs) where the parameters of hidden nodes need not be tuned, stochastic time effective function NNs (STEFNNs), and radial basis function networks (RBFNs) that use radial basis functions as activation functions [81,82]are categorized into the ANN family, as their structures are similar. An ANN-based method was first developed in 1988 to detect unknown regularities in the daily return of IBM stock with its prices following a random walk [83]. Similar research on ANNs in stock market prediction was carried out later [58,84]. Compared with other approaches in the field of TSF, ANNs can convincingly achieve satisfactory performance [85].

The MLP is a popular ANN model applied to predict FTS movement. The output value y of a three-layer perceptron (Fig. 1) can be formulated as:

$$y = F\left(\sum_{i=1}^{N} W_i h_i - T\right), \tag{9}$$

where N is the number of neurons in the hidden layer, W_i is the synaptic weight from neurons in the hidden layer i to the output part, i in the hidden layer to the output part, h_i is the output of neuron i, T is the threshold of the output layer and F is the sigmoid activation function of the output layer. The output value of neuron j in the hidden layer is as follows:

$$h_j = f\left(\sum_{i=1}^{M} W_{ij} x_i - T_j\right), \tag{10}$$

where W_{ij}s are the weights from the input neuron i to neuron j in the hidden layer, x_i are the input data, T_j is the threshold of neuron j, and f is the sigmoid activation function of the hidden layer. The backpropagation error algorithm is adopted as the training algorithm. It is based on the gradient descent principle and implements iteration for updating the weights and thresholds for each training vector p of the training sample.

$$\Delta W_{ij}(t) = -\alpha \frac{\partial E^p(t)}{\partial W_{ij}(t)}, \Delta T_j(t) = -\alpha \frac{\partial E^p(t)}{\partial T_j(t)}, \tag{11}$$

where α is the learning rate, $\frac{\partial E^p(t)}{\partial W_{ij}(t)}$ and $\frac{\partial E^p(t)}{\partial T_j(t)}$ are the gradients of the error function on each iteration t for the training vector p with $p \in 1, \ldots, P$, and P is the size of the training dataset.

Compared with parametric statistical models, ANNs have the following advantages that make them suitable for dealing with chaotic FTS data.

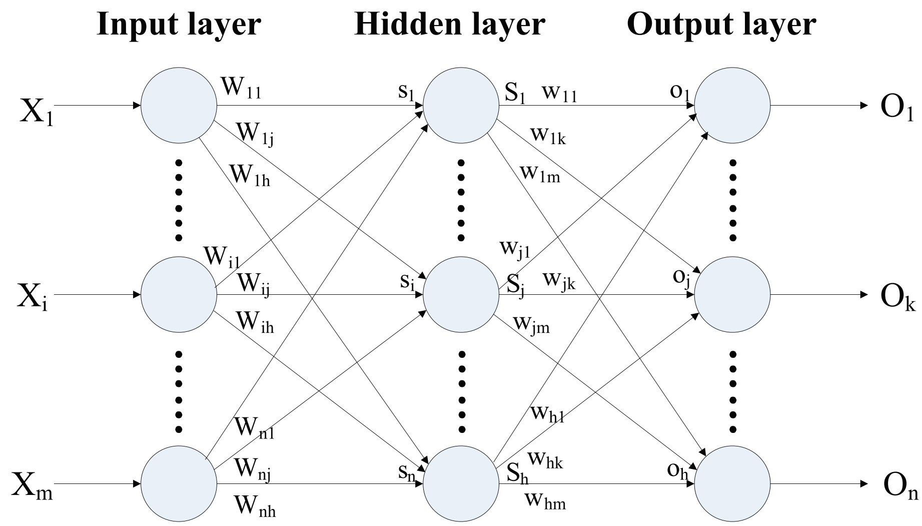

Fig. 4. The morphological architecture of ANNs.

- As a nonlinear model, an ANN can approximate any nonlinear continuous function without formal description, and due to the large number of interconnected processing elements, ANNs can be trained to learn new patterns [86,87]. Therefore, ANNs can provide more possibilities in selecting input-output relationships, and the prediction ability of a nonlinear model is stronger than that of a linear model [88].
- ANNs have no standard structural equation, making is easier to adapt to changes in the financial market, so they are more suitable for FTS data prediction [89].
- As nonparametric weak models based on a data-driven modeling idea, ANNs more easily avoid the problem of model error and are more powerful in learning the dynamics of FTS data [89].
- ANNs also have other merits, such as less complexity, less overfitting, high robustness, reliability in predictions by autocorrection, and being able to provide excellent generalized properties.

However, ANNs also have intrinsic shortcomings, such as the following:

- ANNs often perform inconsistently and unpredictably on noisy FTS data with high dimensionality [9].
- ANNs have higher computational complexity, which is more complicated than traditional prediction models.
- ANNs are difficult to interpret due to their black box nature.
- ANNs inevitably suffer from overfitting and are trapped in local minima in practical applications [49].

5.1.2. Support Vector Machines & Support Vector Regression

Support vector machines (SVMs) are applicable in FTS data modeling with its learning mechanism following the principle of structural risk minimization, realizing a kernel function considering empirical error and a regularization term and mapping the input training samples to high-dimensional feature space so that the machine can classify highly complex data [90]. The SVM constructs a hyperplane as the decision surface to identify the training samples, which contributes to maximizing the distance between the training samples and the decision boundary. As a convex function is required in the optimization process of the SVM, a unique optimal solution (global minima) can be achieved. It can adapt to various kernel functions, such as linear, polynomial and RBF. The foundation of SVM is providing linear classification for input data to select an optimal line with higher reliability through quadratic programming methods. The model has an obvious strong dependence on the data subset, which defines the classification boundaries on the training set and contributes to the decision. In short, the algorithm uses nonlinear mapping to transform the master data into higher dimensions and uses linear optimality to separate hyperplanes [91]. Due to its strong nonlinear approximation ability [69], the SVM is widely used to predict the time series of nonstationary variables, the irrationality of classical methods, or the complexity of time series [92]. The SVM has been applied to classification and regression problems, and its purpose is to predict standard variables based on predictive variables [93,94].

The SVM has many advantages in theory as follows:

- SVMs can eliminate irrelevant and scattered data to improve the prediction accuracy [95].
- The relatively easy training process is the major strength of SVMs.
- The solution of SVMs are obtained through quadratic programming with linear constraints, so it is the only global optimal solution. Unlike ANNs, local optima do not appear in the SVM operation [50].
- SVMs scale the input data moderately well to high dimensions. They can appropriate optimal parameters to explicitly control the model complexity and possible errors.
- Based on the unique principle of structural risk minimization, SVMs more easily avoid the overfitting problem and improve the prediction ability by minimizing the upper bound of the generalization error.

Nevertheless, the shortcomings of SVMs are as follows:

- The training time of SVMs is approximately two to three times the number of training set samples.
- The importance of diverse variables cannot be illustrated, and an overfitting issue will occur if the hyperparameters are not set appropriately.
- Considerable time and effort are required to construct the hyperplane of SVMs, particularly when analysing large-scale datasets.

Support vector regression (SVR) is the regression version of the SVM and is thus very similar to SVM technology, but it is used for regression rather than classification. Similar to the SVM, SVR transforms complex nonlinear input data into a high-dimensional feature space and generates a hyperplane by applying various kernel models to accurately and reliably predict the output value. This transformation makes the prediction process using a linear disjunctive model in high-dimensional feature space feasible. SVR is insensitive to the ϵ value changes. It can ensure the predicted result variance, which is always less than the maximum acceptable variance calculated by the loss function. The control of overfitting and underfitting problems in threshold detection can alleviate the deviation results caused by outliers. In the computation process, attribute correlations and dependencies can be distinguished automatically. In addition, SVR can maintain stability with assurance from the max ε deviation, resisting changes in the results with rare noisy data input. As SVR is suitable for handling small-scale training data, among ML methods, it is the best model to implement nonlinear regression with small datasets. According to [96], the nonlinear SVR function $f(x)$ is defined as follows:

$$f(x) = y = W^T \cdot \Phi(\upsilon_i) + b, \tag{12}$$

$\Phi(\upsilon_i)$ is not only the transformed high-dimensional feature space but also the scalar value and the weighted vector designed from the training data. To handle chaotic and nonlinear time series data, SVR utilizes gradient descent cost functions to make predictions [97].

SVR is used to predict stock prices, indices and exchange rates because it can reveal the relationship between several technical indicators and the price of financial assets [98]. SVR has additional advantages over the other prediction models in processing nonlinear data.

- The SVR process is easy to implement even in high-dimensional spaces.
- SVR requires fewer computational resources than other prediction modes.

However, SVR still has intrinsic shortcomings.

- SVR needs more memory space than other models to reserve all the generated support vectors.
- Deciding a suitable kernel function is not easy according to the evaluation metrics.
- It is difficult to calculate the attribute coefficiency of the dataset, interpret the whole process and generalize the regression process.
- SVR cannot allocate proper dynamic weights to predictive variables, which increases the prediction error rate.

When making financial asset value movement predictions, SVMs and ANNs are popularly adopted to classify stocks, and it is found that SVM usually outperforms the ANN. The relevant literature [99,100,48,68,101–107] concluded that SVMs perform better than ANNs when applied to financial market forecasting. Notably, kernels grant SVMs the capability to define nonlinear decision boundaries, while ANNs can also define nonlinear decision boundaries. In the abovementioned papers, most SVMs preferred the general RBF kernel function, but it is not easy to determine the best selection of the kernel function.

5.1.3. Random Forest

Random forest (RF) was first introduced by Leo Breiman [108]. It is based on a decision tree model, which is defined with classification and regression trees. Decision trees are adopted to forecast classified variables to categorize the samples into classes are called classification trees, while those used for predicting continuous variables are known as regression trees. Two crucial characteristics in constructing RF models are the adoption of the bagging method [109] and random choice in every node. The bagging method is based on bootstrapping and composition concepts. In each node of the decision tree, Breiman (2001)[110] adopted $\left[log_2^M\right] + 1$ features, where M stands for the total number of input features. The essence of the RF algorithm is to be compatible with changes and eliminate instability. More details on RFs can be found in [111,110].

The advantages of RF include the following:

- As a combination of group algorithms, RF can generate a strong learner by combining several weak learners, which prevents the input data from overfitting.
- Even if the classification baseline is part of the unstable learning algorithm, good results can still be obtained.

The disadvantages of RF are as follows:

- Small changes in training data will lead to major changes in the model established by RF.
- With the characteristics of random features, in every node of each decision tree, a few input features are selected randomly, but for node division, the best feature with the most efficient information is selected for the tree's growth instead of exploring all the features.
- RFs would consume a large amount of computational resources for the CART decision tree algorithm to be adopted such that each tree grows to its maximum size without pruning.

5.1.4. Deep learning

DL is usually considered a subset of ML and has been successfully applied to classification tasks [11]. It was first introduced in 2005 and has been considered since 2012 [112]. In fact, DL represents a new attitude towards the idea of NNs, which means a new method to study ANNs [113,114]. Table 1 shows that LSTM has been the most commonly used FTS forecasting method in the last five years. With the further exploration of DL models in stock forecasting, their ratio as a baseline has increased in the past few years [115]. DL is a representation learning method with multilevel representation, which is obtained by superimposing multiple simple but nonlinear modules. Each module converts the representation of each layer (starting from the input layer) into a more abstract representation of the next layer. With enough such transformations, very complex functions can be learned [54]. While improving the prediction accuracy of data within the sample, it is easier to alleviate the overfitting problem [55], which will be very helpful in mining the complex internal characteristics of FTS data.

LSTM and its variations present superior performance in FTS prediction, utilizing data varied over time with feedback embeded behaviours. Publication statistics reveal that most researchers preferred to adopt LSTM to conduct FTS forecasting because a large amount of financial data contains time-dependent components. LSTM is a special DL model derived from a more general classifier series, namely, recurrent NNs (RNNs). In fact, more than half of publications on TSF concern RNNs. Regardless of the problem, whether price or trend prediction, researchers prefer to consider RNNs and LSTM as practical options because of the ordinal nature of the data. Therefore, RNN and LSTM models were selected in some studies for comparison with other developed models as the benchmark.

(1) Long Short-Term Memory (LSTM): LSTM is widely used in many sequential tasks, such as financial market forecasting. The topological structure of LSTM is composed of a set of recurrent subnetworks, which are called memory blocks. Each block has one or more autoregressive storage units and three or more units, namely, the input gate, forget gate, and output gate. These gates manage

the information flow in the cells, which are analogues of the cells' continuous writing, reading, and regulation operations. Each cell can remember the optimal values over random time spans with these natural features. In LSTM, long-term memory stands for learned weights and short-term memory stands for the internal status of the cells. The long-term affiliations that can be properly learned and referenced are the main feature of LSTM [116]. Learning time is not important for LSTM's operation mechanism because learning is already completed before the system starts to run, so there is enough time available. The more important issues are the accuracy of learning and computation error reduction. Bao et al. (2017)[117] pointed out that, compared with traditional ANNs, SVMs and other algorithms, LSTM can extract robust features to describe more complex real data. At the same time, it captures the sequence-related information of FTS data, so LSTM fits for FTS prediction, and the verification results also confirm that the LSTM is superior to other prediction models, which can obtain more accurate prediction results with better profitability. Fischer & Krauss (2018) [118]believed that LSTM could extract important information from FTS data with a large amount of noise. The verification results show that, compared with RF and logistic regression models, LSTM has a better prediction effect.

The following equations show the early development form of the LSTM unit [119](x_t: input vector to the LSTM unit, f_t: forget gate's activation vector, i_t: input gate's activation vector, o_t: output gate's activation vector, o_t: output vector of the LSTM unit, c_t: cell state vector, σ_g: sigmoid function, σ_c, σ_h: hyperbolic tangent function, $*$: elementwise (Hadamard) product, W, U: weight matrices to be learned, b: bias vector parameters to be learned).

$$f_t = \sigma_g(W_f x_t + U_f h_{t-1} + b_f), \tag{13}$$

$$i_t = \sigma_g(W_i x_t + U_i h_{t-1} + b_i), \tag{14}$$

$$o_t = \sigma_g(W_0 x_t + U_0 h_{t-1} + b_0), \tag{15}$$

$$c_t = f_t * c_{t-1} + i_t * \sigma_c(W_c x_t + U_c h_{t-1} + b_c), \tag{16}$$

$$h_t = o_t * \sigma_h(c_t). \tag{17}$$

LSTM has the following merits:

- LSTM performs better than other sequence models (such as RNNs), especially when there is a large amount of data.
- LSTM can effectively deal with the sequence-related problems of FTS data. LSTM has solved the problem of the long-term dependence on sequences with memory modules and has achieved great success in sequential data applications such as sequential text translation [120].
- LSTM performs better than other sequence models (such as RNN), especially when there is a large amount of data.
- LSTM can store more memory than a conventional RNN by avoiding the gradient descent problem. Compared with RNNs' maintenance of a single hidden state only, LSTM has more parameters and can better control the status in which memories are saved and discarded at a specific time step.

LSTM has the following shortcomings:

- The main practical limitation of gradient descent in the LSTM's operation mechanism is that the model cannot learn long-term dependencies.
- LSTM models present the challenges of inconvenient configuration and original data format conversion requirements before learning.

(2) Convolutional Neural Network (CNN): CNNs have been increasingly applied in recent years because they perform better in classification problems, and they are more apt to perform either static data representation or nontime varying problems. They are used mainly for two-dimensional image recognition and classification [121–123]. Their topological architecture has several layers, namely, convolutional, max-pooling, dropout, and fully connected MLP layers. Every neuron group, also called the filter, executes a convolution process motivated by the kernel window function on the input images from different regions, and the same weights are assigned to different neurons, which reduces the parameter numbers in contrast with the dense feedforward NN and thus benefits computing and storage. CNN extracts local features by restricting the local perception of hidden units. The convolutional layer performs a convolution (filtering) operation. Eq. (18) shows a basic convolution operation, where t represents time, s represents feature mapping, w represents the kernel, x represents the input, and a represents the variable. In addition, the convolution operation is realized in two dimensions. Eq. (19) shows the convolution operation of a two-dimensional image, where I represents the input image, K represents the kernel, (m, n) represents the image size, and i and j represent the variables. The continuous convolution layer and maximum pooling layer constitute the depth network. Eq. (20) describes the NN structure, where W represents weights, x represents input, b represents bias, and z represents the output of neurons. At the end of the network, the softmax function helps calculate the output. Eqs. (21) and (22) illustrate the softmax function, where y denotes the output [124].

$$s(t) = (x * w)(t) = \sum_{a=-\infty}^{\infty} x(a) w(t-a), \tag{18}$$

$$S(i,j) = (I * K)(i,j) = \sum_m \sum_n I(m,n) K(i-m, j-n), \tag{19}$$

$$z_i = \sum_j W_{i,j} x_j + b_i, \tag{20}$$

$$y = \mathrm{softmax}(z), \tag{21}$$

$$\mathrm{softmax}(z_i) = \frac{\exp(z_i)}{\sum_j \exp(z_j)}. \tag{22}$$

Scholars have explored the applicability of CNNs to FTS prediction. Tsantekidis et al. applied CNN to stock price prediction [125]. The empirical results show that the prediction effect of this method is better than ANN, SVM and other models. A CNN was used by Sezer & Ozbayoglu to identify buying, selling and holding opportunities, [126]and the empirical results show that this strategy outperforms buying Holding strategies and other algorithmic trading strategies.

Generally, CNNs have the following advantages:

- CNNs includes a convolution operation, pooling layer operation and other steps, which greatly reduce the number of parameters that need to be trained, thus reducing the complexity of CNNs and making the learning and training process simpler.
- CNNs can effectively process the spatial correlation characteristics of data.

However, there are still some limitations to CNNs' application to FTS forecasting.

- The hidden state must be updated in each training step, so CNNs cannot determine whether the memory is saved or discarded.

- It is difficult to capture the sequence-correlation characteristics of FTS data during CNN implementation.

In summary, scholars have confirmed the feasibility and effectiveness of applying DL algorithms such as RNNs, LSTM and CNNs to FTS prediction. At the same time, in the empirical prediction effect, the prediction effect of DL is better than that of ANNs, SVMs and other ML algorithms. Not only does DL have a stronger ability to extract hierarchical features but also some DL algorithms, such as LSTM, can describe the sequence-related features of FTS well. Therefore, compared with other FTS prediction models, such as econometric models, in-depth learning can more effectively mine the important information contained in FTS data and has more advantages in the field of FTS forecasting.

5.2. Reproducibility of the selected top 10 most-cited studies

Aiming to address the evolution of the state-of-the-art ML model applied to FTS forecasting, this part comments on the top 10 most cited articles according to the Web of Science database,

Table 4
The ten most cited articles in the compiled database (The number of citations refers to the citations in the Web of Science database from 2011 to 2021).

Sequence	Title	Journal	Authors	Year	Total citations	Citations per year
1	Deep learning with long short-term memory networks for financial market predictions	European Journal of Operational Research	Thomas et al.	2018	368	92
2	A hybrid stock selection model using genetic algorithms and support vector regression	Applied Soft Computing	Huang et al.	2012	113	11.3
3	Deep learning-based feature engineering for stock price movement prediction	Knowledge-Based Systems	Wen et al.	2019	83	27.67
4	ModAugNet: A new forecasting framework for stock market index value with an overfitting prevention LSTM module and a prediction LSTM module	Expert Systems with Applications	Baek et al.	2018	74	18.5
5	A novel hybrid model using teaching–learning-based optimization and a support vector machine for commodity futures index forecasting	International Journal of Machine Learning and Cybernetics	Prasad et al.	2018	63	15.75
6	Fractional Neuro-Sequential ARFIMA-LSTM for Financial Market Forecasting	IEEE Access	Hussain et al.	2020	56	28
7	Stock market one-day ahead movement prediction using disparate data sources	Expert Systems with Applications	Bin et al.	2017	53	10.6
8	Bridging the divide in financial market forecasting: machine learners vs. financial economists	Expert Systems with Applications	Hsu et al.	2016	52	8.67
9	Technical analysis and sentiment embeddings for market trend prediction	Expert Systems with Applications	Andrea et al.	2019	50	16.67
10	Complex Neurofuzzy ARIMA Forecasting A New Approach Using Complex Fuzzy Sets	IEEE Transactions on Fuzzy Systems	Li et al.	2013	49	5.44

Table 5
Reproducibility of the top 10 most cited articles.

Sequence	Data sets	Time duration	Assets	Predictive variables	Predictions	Main methods	Performance measures
1	S&P500	December,1992 - October,2015	Index	Return indices	Price	LSTM	Accuracy, Standard deviation
2	Taiwan Stock Exchange	1996–2010	Stock	Stock return	Price	hybrid GA-CSVR	Accuracy, Standard deviation
3	CSI 300	December 9, 2013 - December 7, 2016	Stock	Price	Volatility	MFNN	Accuracy
4	KOSPI200 S&P500	January 4, 2000 - July 27, 2017	Stock	Price	Volatility	LSTM	MSE,MAPE MAE
5	MCX COMDEX	January 1,2010 - May 7, 2014	Commodity futures index	Index	Volatility	SVM-CTLBO	RMSE,NMSE MAE,DS
6	Fauji Fertilizer Company (FFC)	January 1, 2009 - May 30, 2018	Stock	Price	Price	ARFIMA-LSTM	MAE,RMSE, MAPE
7	AAPL(Apple NASDAQ)	May 1,2012 - June 1,2015	Stock	Price	Volatility	DT,NN SVM	Accuracy,AUC, Precision, Sensitivity, Specificity
8	TickWrite Data Inc.	2008-C2014	Stock indices	Price	Direction	SVM ANN	ROI Accuracy
9	NASDAQ100	July 3,2017 - June 14,2018	Stock	Price	Direction	RF,SVM NN	Accuracy Recall
10	NASDAQ composite index TAIEX DJI	January 3, 2007 - December 20, 2010 1999 to 2004 1999 to 2004	Stock index	Index	Index value	CNFS-ARIMA	MSE NRMSE NRMSE

which is listed in Table 5. When the investigated papers are used as a baseline in future research, the discussion on implementation and reproducibility will be very useful. This part reproduces the most influential literature retrieved during the time interval 2011–2021, which can represent the latest research focus and future research orientation. Tables 5 and 6 summarize the characteristics of the top 10 most cited papers. Six of them were published in the past three years, which announced the latest theoretical and practical application of ML models in FTS forecasting. Four of them were published in Expert Systems with Applications, which verifies Table 2's statistics and shows that the journal is the most popular outlet for sharing ML's implementation in FTS forecasting. Regarding the prediction object, six of them treat stocks as the prediction scope, four concentrates on indices, and one article considers both the stock and the index, which further indicates that indices and stocks are important financial instruments that attract prediction preferences. This finding is in accordance with the implications of Table 4. The datasets that caught the most attention are from established markets, with six papers adopting this focus, three papers considering emerging markets, and one paper investigating both markets. The papers that concentrate on emerging markets were published in the past three years, which indicates the latest development of the research topic. Regarding the main prediction method, seven of these studies adopt hybrid methods involving either statistical methods or ML methods, especially LSTM. Except for LSTM, the other individual method has played an important role in paving how financial markets will be analyzed for several decades. Considering the research time span, four studies employ a data interval longer than ten years, and one adopts a time interval of less than one year. Among these representative papers, seven explore and exploit the superiority of specific forecasting models, while the remainder favour feature selection and integration. When the performances of the employed prediction methods are measured in these papers, six of them adopt classification evaluation metrics, especially accuracy, while regression evaluation metrics such as *MSE* and *MAE* are applied in the remaining methods. It is obvious that although FTS forecasting is a regression problem, it is framed as a classification problem in most research. Forecasting the volatility and trend of the financial market was supposed to be more valued until now. The main contents and contributions of these ten most cited publications are provided here.

Table 6
List of Acronyms.

Acronyms	Definition
FTS	Financial Time Series
ML	Machine Learning
DL	Deep Learning
LR	Linear Regression
LOGIT	Logistic Regression
GARCH	Generalized Autoregressive Conditional Heteroscedasticity
ARIMA	Autoregressive Comprehensive Moving Average
ANN	Artificial Neural Network
SVM	Support Vector Machine
SVR	Support Vector Regression
RF	Random Forest
LSTM	Long-Short Term Memory
CNN	Convolutional Neural Network
GA	Genetic Algorithm
HMM	Hidden Markov Model
EMD	Empirical Mode Decomposition
NN	NN
DE	Differential Evolution
EC	Evolutionary Computation
BPNN	Backpropagation Neural Network
MLP	Multilayer Perceptron
ELM	Extreme Learning Machines
STEFNN	Stochastic Time Effective Function NN
RBFN	Radial Basis Function Network
TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative
MSE	Mean Squared Error
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
RMAE	Root Mean Absolute Error
NMSE	Normalized Mean Squared Error
RMSE	Root Mean Squared Error
NRMSE	Normalized RMSE
MAPE	Mean Absolute Percentage Error
RMSRE	Root Mean Squared Relative Error
MPPE	Maximum Positive Prediction Error
MNPE	Maximum Negative Prediction Error
R	the Correlation Exponent of Prediction

Fischer Thomas et al. (2018) was the most popular cited paper since its publication, and the frequency of its citation reveals that LSTM is the research hotspot. Fischer Thomas et al. (2018) [118] focused on LSTM networks used in FTS forecasting. Constituent S&P 500 stocks from 1992 to 2015 were chosen as the target, and the directional motion outside the samples was predicted by using the LSTM network. Memoryless classification methods, such as RF, deep neural net, and logistic regression classifier, were adopted as the baseline models. The comparison result showed that LSTM performed better than the other baselines, with a daily mean return of 0.46 % before and 0.26 % after transaction cost. LSTM's superior performance relative to the specific market was very obvious from 1992 to 2009. However, 2010 was an exception with excess returns because LSTM's profitability fluctuated close to zero after the transaction cost, which seemed to be arbitraged. Further regression analysis revealed that compared with the three baselines, LSTM presented a low exposure return to the common source of system risk. It can be concluded that LSTM is a better choice in terms of prediction accuracy and daily yield. The authors discovered that as a form of DL, the LSTM network constituted an advancement in this research domain. The main contribution of this paper is that the authors successfully proved that the LSTM network can effectively extract meaningful information from noisy FTS data to promote predictional accuracy and daily returns.

Chien Feng Huang (2012) [127] took the 200 largest market value constituent stocks listed on the Taiwan Stock Exchange as the investment universe and retrieved the annual financial statement data and stock returns from the TEJ (Taiwan Economic Journal Co. Ltd., http://www.tej.com.tw/) database from 1996 to 2010. The authors developed a hybrid method combining SVR with GA for effective stock selection. They concluded that there were significant opportunities to resolve the issues of ranking stocks for effectively and successfully constructing portfolios. The cumulative average annual return rates of 200 stocks in the investment field were regarded as the benchmark. Their research was composed of three steps. First, GA was used to optimize model parameters and select features to obtain the optimal subset of input variables because it is a simple binary coding method and has the effectiveness of feature selection. Then, SVR was used to generate alternatives to the actual stock return to provide a reliable stock ranking. Later, the top-ranked stocks were selected to form a portfolio. The experimental results showed that this hybrid GA-SVR methodology outperformed the benchmark in terms of investment returns. This paper focuses on the contribution of feature selection before the construction of the prediction model and illustrates the advantages of the hybrid method in the domain of FTS prediction. Since its publication, it has ranked in the top 2 most highly cited papers, reflecting the research trend of applying the hybrid method.

Wen Long et al. (2019) [128] adopted China's stock market data of CSI 300 at a 1-min frequency from December 9th, 2013, to

December 7th, 2016, for training and testing and proposed a new end-to-end model called the multifilter NN (MFNN) for extreme market prediction and signal-based trading simulation. By fusing convolution neurons and recursive neurons, the MFNN extracted the characteristics of FTS data and predicted price changes. In this study, traditional ML models, statistical models and single structure (namely, convolution, recursion and LSTM) networks were used as baseline models. The experimental results showed that the MFNN is superior to the baselines in terms of accuracy, profitability, and stability. The best prediction results calculated by the MFNN were 11.47% and 19.75% better than the best ML method and statistical method, respectively. In particular, compared with RNNs and CNNs, the integration and network structure of the filter improved the accuracy by 7.19% and 6.28%, respectively. For market simulation, the performance of the MFNN combined with the DL method was also better than the baselines, that is, 15.41% better than the best results obtained by the traditional ML method and 22.41% better than the statistical method. The performance of the proposed method was clearly superior in terms of profitability when making prediction simulations, but there were still problems of risk and stability. This paper focused mainly on the superior advantage of the DL hybrid method in predicting financial asset value movement.

Baek Yujin et al. (2018) [129] proposed a new data enhancement method called the ModAugNet framework, which was composed of an overfitting prevention LSTM module and a prediction LSTM module. The construction of this hybrid model was based on the recognition that DL had the disadvantage of overfitting owing to its limited data availability during training. Two different representative stock market datasets, the S&P500 and Korea Composite Stock Index 200 (KOSPI200), were the prediction objectives. The research treated the single DL methodologies DNN, RNN, and SingleNet as the baselines. The experimental results confirmed the superior forecasting performance of the proposed model. Compared with a comparative model such as SingleNet, ModAugNet yielded a lower test error by overcoming overfitting that the LSTM module lacked. With the evaluation metrics MSE, MAPE, MAE and forecasting error for S&P500 and KOSPI200 all decreasing to satisfactory levels, this hybrid DL model demonstrated its strong ability in financial data prediction. The significant contribution of this study is that the proposed model is applicable to all situations, even if it is challenging to model artificially increased data.

Shom Prasad Das and Sudarsan Padhy (2015)[130] constructed a hybrid method based on recently developed ML techniques and tested the commodity future contract index (MCX COMDEX) data obtained from the website *http : //www.mcxindia.com* from January 1, 2010, to May 7, 2014. This hybrid method combined a SVM with teaching–learning-based optimization (TLBO) to modify the user-specified control parameters, which were required by other optimization methods. The authors tested the practicability of the novel TLBO algorithm for choosing optimal free parameters and operated an SVM model to handle FTS data. This study used the particle swarm optimization (PSO)-SVM hybrid model and the standard SVM model as the baselines. The experimental results showed that the proposed SVM-TLBO hybrid regression model could effectively obtain the optimal parameters, and its performance was better than that of the standard SVM model. Compared with the standard SVM regression, the model performed better not only in the evaluation matrix *MAE* but also in *RMSE*. In terms of *MAE* and *RMSE*, there were similar advantages compared with the PSO-SVM hybrid model. In addition, the experimental results also showed that the proposed SVM-TLBO hybrid regression model had advantages over the standard SVM and PSO-SVM hybrid model. This study revealed the advantages of parameter modification and method integration in financial forecasting.

Ayaz Hussain Bukhari et al. (2020) [131] proposed a novel autoregressive fractional integrated moving average (ARFIMA)-LSTM hybrid recurrent network model. This model could extract potential information and capture the nonlinearity feature of the data in the residual values with exogenous dependent variables. In addition, the DL model's intrinsic dynamical features predict sudden random changes in financial markets, such as the advantage of fractional order derivative and the ability of long-term dependency, which could match the sequences of the input with the output. The daily open prices of Fauji Fertilizer Company data from January 1, 2009, to May 30, 2018, were adopted as the experimental data. Experimental results revealed that this hybrid ARFIMA-LSTM model achieved better performance in terms of prediction accuracy. Notably, the augmentation of the exogenous input of dependent variables improved the prediction accuracy compared with the independent adoption of ARIMA, ARFIMA, and GRNN. This research enhanced ARFIMA filters' linear tendencies in hybrid scenarios. Moreover, error analysis also verified the superiority of the proposed model's performance to that of traditional individual models on the metrics of *MAPE* and accuracy rate. The superiority of the proposed hybrid model proved that models with perfect parameters could significantly improve the FTS forecasting performance.

Weng Bin et al. (2017)[132] proposed a data-driven financial expert approach to predict the daily movement of a stock that consisted of three main phases, namely, data acquisition and feature generation, identification of important features, and ML model comparison and evaluation. They applied three individual ML methods to forecast the stock price fluctuation, though the focus of this research was feature selection. From May 1, 2012, to June 1, 2015, 37 months of large and feature-rich datasets of Apple Inc. were collected as experimental data. This study showed that considering the prediction factors from multiple sources could improve the prediction performance of the proposed model, and the performance of the ML model under different objectives was variable. This study combined completely different online data sources with traditional time series and technical indicators to build a more intelligent and effective daily trading expert system. The proposed expert system was tested to outperform the reported results. The research concluded that (1) so-called nontraditional experts such as Google and Wikipedia benefited the knowledge-based financial expert system from the captured data, and ML models could utilize Google News indicators to increase the prediction accuracy of the proposed expert system; (2) integrating data from different sources could diversify the knowledge background, thus improving the performance of financial expert systems; and (3) with a rich knowledge database, the adoption of simple ML models for reasoning and rule generation was effective. According to the seven scenarios of data aggregation, it could be concluded that the supplementation of online data sources such as Google News and features generated lately could significantly improve the model's prediction performance. This research focused on the importance of data accumulation and feature selection to improve prediction performance as well as model construction.

Ming-Wei Hsu et al. (2016)[133] explored the impact of methodological factors on prediction behaviour instead of prediction accuracy optimization, with the belief that the information efficiency of the financial market affected the price predictability of assets and the possibility of profitable transactions. The simulation adopted financial index data from 34 markets collected by Tick Write Data Inc., which included both developed and developing markets over six years from 2008 to 2014. The forecasting experiment merely adopted the most popular ML techniques, namely, the SVM and the ANN. Through 34 financial market dimensions, two prediction intervals, two covariate combinations and two simulation methods, 272 individual SVM and ANN prediction results were obtained, which confirmed the superiority of the best ML approaches over the best econometric approaches on the evaluation metric of forecasting accuracy. The results also verified the previous conclusion that the SVM performed better than the ANN in financial market modeling with higher accuracy, greater stability and more robustness. The results showed that the predictability of a financial market and the feasibility of trading are affected by a series of mixed factors, such as the maturity of the target market, the prediction method, the time span, and the simulation used to evaluate the model. These factors represent the uncertainty of FTS forecasting. In fact, although the authors merely adopted the already existing ML methods to make the prediction, their findings imply that the financial market targeted for prediction played a significant and substantial role in achieving a certain forecasting accuracy and trading profitability.

Picasso, A et al. (2019)[134] exploited a hybrid approach that employed RF, an SVM, and a feed-forward NN to solve the value movement classification problem. The adopted dataset was composed of the 20 top capital companies listed in the NASDAQ100 index, and the time interval was from July 3, 2017 to June 14, 2018. A two-step evaluation was performed. The first step was executed to evaluate the classifier's behaviour from the statistical perspective, and the second focused on testing the model's prediction effectiveness. Both technical analysis indicators and the views from news articles were treated as inputs. The model aimed to forecast the portfolio development trend and avoid liquidity while making trade simulations. This proposed approach's actual effectiveness was proven by the simulation achieving an annual return of 80%, which was higher than the average. In addition, this research focused on comparing three feature sets, namely, the Price, News, and Price& News sets. Compared with the use of individual indicator Price sets only, the integration of emotion embedding and technical indicators resulted in better performance. However, the latter approach was not superior when using the News sets alone. This finding revealed a practical future direction to overcome this weakness, that is, the design and adoption of a useful feature integration technique. In addition, the authors predicted that once the applied period was extended, the forecasting performance of the model would be improved. This conclusion also suggests a future research direction. Future selection and applied time intervals will be important factors that affect the prediction performance.

Chunshien Li and Tai-Wei Chiang [135] presented an innovative hybrid approach named the CNFS-ARIMA approach, which integrated both a complex neuro-fuzzy system (CNFS) and ARIMA because ARIMA models fit for time series linear modeling and CNFS matches nonlinear function mapping. Since the real part and imaginary part were used for two different function mappings, the model output was complex to some extent, which is the so-called dual-output property. To construct the architecture of CNFS-ARIMA, both topology and parameter learning were necessary for realizing self-organization and self-tuning. The National Association of Securities Dealers Automated Quotation (NASDAQ), the Taiwan Stock Exchange Capitalization Weighted Stock Index (TAIEX), and the Dow Jones Industrial (DJI) Average Index were used to perform the dual-output forecasting experiments. This study used 1000 data points from the daily opening and closing indices of the NASDAQ from January 3, 2007, to December 20, 2010. For TAIEX and DJI, their daily closing indices were collected for six years from 1999 to 2004. The mean *RMSE* was adopted as the metric to evaluate the performance between the proposed hybrid model and other single models. Experimental results showed that CNFS-ARIMA models had outstanding nonlinear mapping ability with obviously lower *RMSE*. This hybrid approach's superiority in forecasting time series movement has been verified by the synergistic advantages of the CNFS and ARIMA. This study verified that the hybrid method could combine the advantages of the individual forecasting methods, which is predicted to be the research trend in this domain.

6. Conclusion and future directions

Nonlinear FTS forecasting is a popular research subject in real economic life. However, the chaotic and irregular features of financial data make this task increasingly complex and error prone. Currently, many prediction methods and modeling paradigms have been developed. In practice, well-designed financial prediction models can assist governments in devising better strategies and making smarter decisions to stabilize and stimulate economic developments at the macro level and help institutions and individuals earn more profits at the micro level. In this paper, a detailed literature review of commonly applied ML approaches is presented. The abovementioned prediction models are discussed, and their capabilities and limitations are presented. In addition, the primary studies concerning this topic are summarized, the possible challenges of implementing ML techniques to predict FTS data are presented, and future research directions are discussed.

Through the literature review, we can conclude that an efficient and scientific FTS forecasting model should possess the following characteristics. First, it should have a strong and robust feature learning ability and should be able to fully capture the high noise, nonlinear, nonstationary and other complex features of FTS data to extract the important influence information that affects the changes in FTS data. Second, it should effectively reflect the nonlinear dynamic interaction of the influencing factors of FTS and then incorporate the nonlinear interaction between FTS data into the prediction model. Third, it should have a strong generalization ability and realize the accurate prediction of data outside the sample to improve the applicability of the prediction model in practice. Fourth, the long-term and short-term sequence dependence characteristics of FTS data should be considered to realize automatic identification of sequence dependence length based on input data.

ML approaches such as ANNs, SVMs and RF are more nonparametric, nonlinear and data-driven models, which can effectively deal with the nonlinear and nonstationary characteristics of FTS data, make it easier to mine the hidden patterns in FTS data, and capture the unknown relationships between FTS data. Therefore, applying the ML algorithm to FTS forecasting has more advantages. However, ML has some shortcomings in dealing with the sequence-dependent characteristics of FTS data. In fact, ML methods far from providing the best solution for FTS prediction when considering complex natural instincts, irregular market behaviour, and the sudden appearance of extreme events, which compromise the learning patterns or generalization as observed in relevant studies. ML models can capture nonlinear relationships in time series, but a single ML model has difficulty fitting all the circumstances in the financial market. Hybrid models can combine several models to obtain a stable and excellent model with a strong ability to reduce deviation and variance by effectively reducing the risk of overfitting a single model. Hence, hybrid models can not only effectively identify the nonlinear relationship in financial data but also avoid excessive noise in model fitting and enhance the generalization ability of the model. In addition, the trade-off between the benefits of forecasting performance improvement and the risk of efficiency loss in the hybrid models should be considered accordingly.

Most scholars and traders focus on the long-term and intermediate-term trend predictions of financial asset movement instead of short-term predictions such as weekly and monthly predictions because long-term trends and intermediate-term trends are related to the price of the scrip only, while short-term trends are related to the random behaviour of the prices. Short-term trends are rarely modelled because it is difficult to find variables that influence the development pattern [136]. This difficulty is reflected in the top ten most highly cited papers in this survey. Financial markets are predictable to a higher degree in the long run. Updated ML approaches, including the LSTM and RNN, have recently been explored in long-term forecasting. Regarding the dataset scale, there is a trade-off between sufficiency and efficiency. A small dataset is not sufficient to test its effectiveness and or the risk of overfitting, while a big dataset encounters the risk of traversing various market styles and presenting out-of-date results. Small datasets usually limit the generalization of the proposed model, and increasing the dimension of the dataset will be a solution to make the performance of the model more credible [137]. Nearly all the top ten most-cited papers reproduced in this survey adopted a big dataset and presented satisfactory forecasting performance.

According to the bibliography retrieval and comprehensive analysis of the literature published in the recent decade, LSTM, the most popular applied ML method, has achieved great success in FTS forecasting against the background of big data and artificial intelligence. The main reasons are as follows. First, LSTM has a strong feature learning ability, which can mine the complex features of FTS data and overcome the deficiencies that make it difficult for traditional measurement models to model FTS data with complex features. Second, LSTM can effectively incorporate the sequence-dependent features of time series data to realize the dynamic identification of the long-term and short-term dependence, thus overcoming the deficiencies that make it difficult for other prediction models to incorporate the sequence-related features of FTS data and realize the automatic identification of the sequence-dependent length. Third, LSTM has a better generalization ability, which can effectively improve the prediction ability and thus enhance the applicability of the prediction model to practical problems. Fourth, LSTM can realize the modeling of massive and unstructured financial data and improve the mining ability. Applying DL to FTS forecasting has important theoretical and practical significance for improving prediction accuracy, realizing FTS forecasting and multidisciplinary cross-fusion research, and expanding existing FTS forecasting models.

The analysis of the top 10 most cited articles suggests that these popularly applied models consume low computational costs in the training phase, which benefits their implementation of frequent trades in a given timeframe. Although the primary targets of the ML approach are optimizing accuracy for classification prediction and minimizing *MSE* for regression prediction, they might not be the most suitable metrics to measure the approach performance when applied in financial trading. In addition, establishing a relationship between accuracy and profitability and balancing them are also crucial in ML-based trading because financial trading incurs considerable costs that need to be included in the performance assessment. Moreover, multiple one-dimensional FTSs can be combined into two-dimensional financial images, and a new tag generation algorithm can be adopted to mark financial images into three states: buy, sell or hold to solve the problem of category imbalance.

Since economists consider risk management and investment return as two major objectives, we expect to exploit promising research to optimize these multiple objectives simultaneously in future works. More systematic approaches are required to find optimal trade-offs to maximize both cumulative profits and accuracies, with the particularities of different financial instruments taken into account. It is notable that improvements in computation speed, model resilience, prediction accuracy, and result reliability are the crucial requirements of the FTS forecasting system's end users. For future research, the selection and integration of certain forecasting approaches together with optimal feature selection and special handling of large local variations still have great room for further improvement. In addition, tailor-made, specific forecasting techniques could be employed for forecasting different time-scale series and thereby improve FTS forecasting.

Declaration of Competing Interest

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

Acknowledgements

This research was supported by the National Key R&D Program of China under Grant 2020YFA0908700, the Social Science Fund Project of Hunan Province, China (Grant No. 20YBA260), the Natural Science Fund Project of Changsha, China (Grant No. kq2202297), the Project of the Guangdong Basic and Applied Basic Research Fund (No. 2019A1515111139), the National Natural Science Foundation of China under Grants 62106151 and 62073225, and the Natural Science Foundation of Guangdong Province-Outstanding Youth Program under Grant 2019B151502018.

References

[1] Y.S. Abu-Mostafa, A.F. Atiya, Introduction to financial forecasting, Applied intelligence 6 (3) (1996) 205–213.
[2] G.J. Deboeck, Trading on the edge: neural, genetic, and fuzzy systems for chaotic financial markets, vol. 39, John Wiley & Sons, 1994.
[3] Y. Kara, M.A. Boyacioglu, Ö.K. Baykan, Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul Stock Exchange, Expert systems with Applications 38 (5) (2011) 5311–5319.
[4] E.F. Fama, Efficient capital markets a review of theory and empirical work, The Fama Portfolio (2021) 76–121.
[5] Y. Chen, Y. Hao, A feature weighted support vector machine and K-nearest neighbor algorithm for stock market indices prediction, Expert Systems with Applications 80 (2017) 340–355.
[6] T. Anbalagan, S.U. Maheswari, Classification and prediction of stock market index based on fuzzy metagraph, Procedia Computer Science 47 (2015) 214–221.
[7] R.K. MacKinnon, C.K. Leung, Stock price prediction in undirected graphs using a structural support vector machine, in: 2015 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT), vol. 1, IEEE, 548–555, 2015.
[8] M.T. Leung, H. Daouk, A.-S. Chen, Forecasting stock indices: a comparison of classification and level estimation models, International Journal of forecasting 16 (2) (2000) 173–190.
[9] M. Kumar, M. Thenmozhi, Forecasting stock index movement: A comparison of support vector machines and random forest, in: Indian institute of capital markets 9th capital markets conference paper, 2006.
[10] R.C. Cavalcante, R.C. Brasileiro, V.L. Souza, J.P. Nobrega, A.L. Oliveira, Computational intelligence and financial markets: A survey and future directions, Expert Systems with Applications 55 (2016) 194–211.
[11] H. Lee, R. Grosse, R. Ranganath, A.Y. Ng, Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations, in: Proceedings of the 26th annual international conference on machine learning, 2009, pp. 609–616.
[12] Y. Bengio, A. Courville, P. Vincent, Representation learning: A review and new perspectives, IEEE transactions on pattern analysis and machine intelligence 35 (8) (2013) 1798–1828.
[13] G.E. Hinton, S. Osindero, Y.-W. Teh, A fast learning algorithm for deep belief nets, Neural computation 18 (7) (2006) 1527–1554.
[14] P. Gao, R. Zhang, X. Yang, The application of stock index price prediction with neural network, Mathematical and Computational Applications 25 (3) (2020) 53.
[15] J.S. Liu, L.Y. Lu, W.-M. Lu, B.J. Lin, Data envelopment analysis 1978–2010: A citation-based literature survey, Omega 41 (1) (2013) 3–15.
[16] B. Vanstone, G. Finnie, An empirical methodology for developing stockmarket trading systems using artificial neural networks, Expert systems with Applications 36 (3) (2009) 6668–6680.
[17] M.C. Thomsett, Getting Started in Stock Analysis, John Wiley & Sons, 2015.
[18] M.C. Thomsett, Getting started in fundamental analysis, John Wiley & Sons, 2006.

[19] A.S. Wafi, H. Hassan, A. Mabrouk, Fundamental analysis models in financial markets–review study, Procedia economics and finance 30 (2015) 939–947.
[20] B. Rockefeller, Technical analysis for dummies, John Wiley & Sons, 2019.
[21] J.J. Murphy, Technical analysis of the financial markets: A comprehensive guide to trading methods and applications, Penguin, 1999.
[22] L.A. Teixeira, A.L.I. De Oliveira, A method for automatic stock trading combining technical analysis and nearest neighbor classification, Expert systems with applications 37 (10) (2010) 6885–6890.
[23] M. Lam, Neural network techniques for financial performance prediction: integrating fundamental and technical analysis, Decision support systems 37 (4) (2004) 567–581.
[24] P.E. Tsinaslanidis, D. Kugiumtzis, A prediction scheme using perceptually important points and dynamic time warping, Expert Systems with Applications 41 (15) (2014) 6848–6860.
[25] T. Bollerslev, Generalized autoregressive conditional heteroskedasticity, Journal of econometrics 31 (3) (1986) 307–327.
[26] G.E. Box, G.M. Jenkins, G.C. Reinsel, G.M. Ljung, Time series analysis: forecasting and control, John Wiley & Sons, 2015.
[27] R.F. Engle, Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation, Econometrica: Journal of the econometric society (1982) 987–1007.
[28] M. Ballings, D. Van den Poel, N. Hespeels, R. Gryp, Evaluating multiple classifiers for stock price direction prediction, Expert systems with Applications 42 (20) (2015) 7046–7056.
[29] G.W. Taylor, Composable, distributed-state models for high-dimensional time series, Citeseer (2009).
[30] S.-H. Hsu, J.P.-A. Hsieh, T.-C. Chih, K.-C. Hsu, A two-stage architecture for stock price forecasting by integrating self-organizing map and support vector regression, Expert Systems with Applications 36 (4) (2009) 7947–7951.
[31] O. Akbilgic, H. Bozdogan, M.E. Balaban, A novel Hybrid RBF Neural Networks model as a forecaster, Statistics and Computing 24 (3) (2014) 365–375.
[32] R. Engle, GARCH 101: The use of ARCH/GARCH models in applied econometrics, Journal of economic perspectives 15 (4) (2001) 157–168.
[33] R.S. Michalski, J.G. Carbonell, T.M. Mitchell, Machine learning: An artificial intelligence approach, Springer Science & Business Media, 2013.
[34] J.B. Anon, Computer-aided endoscopic sinus surgery, The Laryngoscope 108 (7) (1998) 949–961.
[35] I. Portugal, P. Alencar, D. Cowan, The use of machine learning algorithms in recommender systems: A systematic review, Expert Systems with Applications 97 (2018) 205–227.
[36] Y. Lu, Artificial intelligence: a survey on evolution, models, applications and future trends, Journal of Management Analytics 6 (1) (2019) 1–29.
[37] S.R. Beeram, S. Kuchibhotla, A Survey on state-of-the-art Financial Time Series Prediction Models, in: 2021 5th International Conference on Computing Methodologies and Communication (ICCMC), IEEE, 596–604, 2021.
[38] G.S. Atsalakis, K.P. Valavanis, Surveying stock market forecasting techniques–Part II: Soft computing methods, Expert Systems with applications 36 (3) (2009) 5932–5941.
[39] J. Wu, J. Wang, H. Lu, Y. Dong, X. Lu, Short term load forecasting technique based on the seasonal exponential adjustment method and the regression model, Energy Conversion and Management 70 (2013) 1–9.
[40] M.-C. Lee, Using support vector machine with a hybrid feature selection method to the stock trend prediction, Expert Systems with Applications 36 (8) (2009) 10896–10904.
[41] S.H. Kim, S.H. Chun, Graded forecasting using an array of bipolar predictions: application of probabilistic neural networks to a stock market index, International Journal of Forecasting 14 (3) (1998) 323–337.
[42] E. Akman, A.S. Karaman, C. Kuzey, Visa trial of international trade: evidence from support vector machines and neural networks, Journal of Management Analytics 7 (2) (2020) 231–252.
[43] A.B. Kock, T. Teräsvirta, Forecasting performances of three automated modelling techniques during the economic crisis 2007–2009, International Journal of Forecasting 30 (3) (2014) 616–631.
[44] C. Wegener, C. von Spreckelsen, T. Basse, H.-J. von Mettenheim, Forecasting government bond yields with neural networks considering cointegration, Journal of Forecasting 35 (1) (2016) 86–92.
[45] C. Xie, Z. Mao, G.-J. Wang, Forecasting RMB exchange rate based on a nonlinear combination model of ARFIMA, SVM, and BPNN, Mathematical Problems in Engineering (2015).
[46] Z. Song, Y. Tang, J. Ji, Y. Todo, Evaluating a dendritic neuron model for wind speed forecasting, Knowledge-Based Systems 201 (2020) 106052.
[47] P. Paliyawan, Stock market direction prediction using data mining Classification, future 5 (2006) 6.
[48] K.-J. Kim, Financial time series forecasting using support vector machines, Neurocomputing 55 (1–2) (2003) 307–319.
[49] Z. Luo, W. Guo, Q. Liu, Z. Zhang, A hybrid model for financial time-series forecasting based on mixed methodologies, Expert Systems 38 (2) (2021) e12633.
[50] W. Huang, Y. Nakamori, S.-Y. Wang, Forecasting stock market movement direction with support vector machine, Computers & operations research 32 (10) (2005) 2513–2522.
[51] C.-L. Huang, C.-Y. Tsai, A hybrid SOFM-SVR with a filter-based feature selection for stock market forecasting, Expert Systems with applications 36 (2) (2009) 1529–1539.
[52] M. Kumar, M. Thenmozhi, Support vector machines approach to predict the S&P CNX NIFTY index returns, in, 10th Capital Markets Conference, Indian Institute of Capital Markets Paper (2007).
[53] Y. Bengio, Y. LeCun, et al., Scaling learning algorithms towards AI, Large-scale kernel machines 34 (5) (2007) 1–41.
[54] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, nature 521 (7553) (2015) 436–444.
[55] J. Heaton, N.G. Polson, J.H. Witte, Deep learning in finance, arXiv preprint arXiv:1602.06561.
[56] V. Tresp, Committee machines, Handbook for neural network signal processing (2001) 1–18.
[57] N. Kourentzes, D.K. Barrow, S.F. Crone, Neural network ensemble operators for time series forecasting, Expert Systems with Applications 41 (9) (2014) 4235–4244.
[58] T. Kolarik, G. Rudorfer, Time series forecasting using neural networks, ACM Sigapl Apl Quote Quad 25 (1) (1994) 86–94.
[59] J. Yao, C.L. Tan, H.-L. Poh, Neural networks for technical analysis: a study on KLCI, International journal of theoretical and applied finance 2 (02) (1999) 221–241.
[60] G.P. Zhang, Time series forecasting using a hybrid ARIMA and neural network model, Neurocomputing 50 (2003) 159–175.
[61] T.A. Ferreira, G.C. Vasconcelos, P.J. Adeodato, A new intelligent system methodology for time series forecasting with artificial neural networks, Neural Processing Letters 28 (2) (2008) 113–129.
[62] M.R. Hassan, B. Nath, M. Kirley, A fusion model of HMM, ANN and GA for stock market forecasting, Expert systems with Applications 33 (1) (2007) 171–180.
[63] Q. Zeng, C. Qu, A.K. Ng, X. Zhao, A new approach for Baltic Dry Index forecasting based on empirical mode decomposition and neural networks, Maritime Economics & Logistics 18 (2) (2016) 192–210.
[64] M.-Y. Chen, Visualization and dynamic evaluation model of corporate financial structure with self-organizing map and support vector regression, Applied Soft Computing 12 (8) (2012) 2274–2288.
[65] S. McDonald, S. Coleman, T.M. McGinnity, Y. Li, A. Belatreche, A comparison of forecasting approaches for capital markets, in: 2014 IEEE Conference on Computational Intelligence for Financial Engineering & Economics (CIFEr), IEEE, 32–39, 2014.
[66] R.S. Tsay, Analysis of financial time series, vol. 543, John wiley & sons, 2005.
[67] M.L. Junior, M. Godinho Filho, Variations of the kanban system: Literature review and classification, International Journal of Production Economics 125 (1) (2010) 13–21.
[68] F.E. Tay, L. Cao, Application of support vector machines in financial time series forecasting, omega 29 (4) (2001) 309–317.
[69] J.M. Box-Steffensmeier, J.R. Freeman, M.P. Hitt, J.C. Pevehouse, Time series analysis for the social sciences, Cambridge University Press, 2014.
[70] G.E. Box, D.A. Pierce, Distribution of residual autocorrelations in autoregressive-integrated moving average time series models, Journal of the American statistical Association 65 (332) (1970) 1509–1526.
[71] L. Harris, Trading and exchanges: Market microstructure for practitioners, OUP USA, 2003.
[72] P.E. McKnight, J. Najab, Mann-Whitney U Test, The Corsini encyclopedia of psychology (2010) 1.
[73] J. Demšar, Statistical comparisons of classifiers over multiple data sets, The, Journal of Machine Learning Research 7 (2006) 1–30.
[74] M. Maggini, C. Giles, B. Horne, Financial time series forecasting using k-nearest neighbors classification, Nonlinear Financial Forecasting, in: Proceedings of the First INFFC, 1997, pp. 169–181.
[75] T. Fawcett, An introduction to ROC analysis, Pattern recognition letters 27 (8) (2006) 861–874.
[76] L. Guang, W. Xiaojie, L. Ruifan, Multi-scale RCNN model for financial time-series classification, arXiv preprint arXiv:1911.09359.
[77] X. Zhong, D. Enke, Forecasting daily stock market return using dimensionality reduction, Expert Systems with Applications 67 (2017) 126–139.
[78] Z. Zhang, S. Zohren, S. Roberts, Deeplob: Deep convolutional neural networks for limit order books, IEEE Transactions on Signal Processing 67 (11) (2019) 3001–3012.
[79] D.J. Hand, Assessing the performance of classification methods, International Statistical Review 80 (3) (2012) 400–414.
[80] V. García, A.I. Marqués, J.S. Sánchez, An insight into the experimental design for credit risk and corporate bankruptcy prediction systems, Journal of Intelligent Information Systems 44 (1) (2015) 159–189.
[81] R. d. A. Araújo, N. Nedjah, A.L. Oliveira, R. d. L. Silvio, A deep increasing–decreasing-linear neural network for financial time series prediction, Neurocomputing 347 (2019) 59–81.
[82] J. Wang, J. Wang, Forecasting stock market indexes using principle component analysis and stochastic time effective neural networks, Neurocomputing 156 (2015) 68–78.
[83] H. White, Economic prediction using neural networks: The case of IBM daily stock returns, ICNN 2 (1988) 451–458.

[84] A.N. Refenes, M. Azema-Barac, A. Zapranis, Stock ranking: Neural networks vs multiple linear regression, in: IEEE international conference on neural networks, IEEE, 1419–1426, 1993.
[85] A. Garg, S. Sriram, K. Tai, Empirical analysis of model selection criteria for genetic programming in modeling of time series system, in: 2013 IEEE conference on computational intelligence for financial engineering & economics (CIFEr), IEEE, 90–94, 2013.
[86] B.E. Hansen, Threshold effects in non-dynamic panels: Estimation, testing, and inference, Journal of econometrics 93 (2) (1999) 345–368.
[87] L. Medsker, E. Turban, R.R. Trippi, Neural network fundamentals for financial analysts, The Journal of Investing 2 (1) (1993) 59–68.
[88] A. Lendasse, E. de Bodt, V. Wertz, M. Verleysen, Non-linear financial time series forecasting-Application to the Bel 20 stock market index, European Journal of Economic and Social Systems 14 (1) (2000) 81–91.
[89] E. Guresen, G. Kayakutlu, T.U. Daim, Using artificial neural network models in stock market index prediction, Expert Systems with Applications 38 (8) (2011) 10389–10397.
[90] N. Cristianini, J. Shawe-Taylor, et al., An introduction to support vector machines and other kernel-based learning methods, Cambridge University Press, 2000.
[91] J. Han, J. Pei, M. Kamber, Data mining: concepts and techniques, Elsevier, 2011.
[92] G. Tian, Research on stock price prediction based on optimal wavelet packet transformation and ARIMA–SVR mixed model, Journal of Guizhou University of Finance and Economics 6 (6) (2015) 57–69.
[93] G.-Q. Xie, The optimization of share price prediction model based on support vector machine, in: 2011 International Conference on Control, Automation and Systems Engineering (CASE), IEEE, 1–4, 2011.
[94] S.P. Das, S. Padhy, Support vector machines for prediction of futures prices in Indian stock market, International Journal of Computer Applications 41 (3).
[95] Z. Wei, A SVM approach in forecasting the moving direction of Chinese stock indices, Lehigh University, 2012.
[96] C. Cortes, V. Vapnik, Support-vector networks, Machine learning 20 (3) (1995) 273–297.
[97] L. Kilian, M.P. Taylor, Why is it so difficult to beat the random walk forecast of exchange rates?, Journal of International Economics 60 (1) (2003) 85–107
[98] J. Chen, SVM application of financial time series forecasting using empirical technical indicators, in: 2010 International Conference on Information, Networking and Automation (ICINA), vol. 1, IEEE, V1–77, 2010.
[99] L. Cao, F.E. Tay, Financial forecasting using support vector machines, Neural Computing & Applications 10 (2) (2001) 184–192.
[100] L.-J. Cao, F.E.H. Tay, Support vector machine with adaptive parameters in financial time series forecasting, IEEE Transactions on neural networks 14 (6) (2003) 1506–1518.
[101] K.-Y. Chen, C.-H. Ho, An improved support vector regression modeling for Taiwan Stock Exchange market weighted index forecasting, in: 2005 International conference on neural networks and brain, vol. 3, IEEE, nil15–1638, 2005.
[102] L. Yu, S. Wang, K.K. Lai, Mining stock market tendency using GA-based support vector machines, in: International Workshop on Internet and Network Economics, Springer, 336–345, 2005.
[103] W. Xie, L. Yu, S. Xu, S. Wang, A new method for crude oil price forecasting based on support vector machines, in: International conference on computational science, Springer, 444–451, 2006.
[104] C.-H. Wu, G.-H. Tzeng, Y.-J. Goo, W.-C. Fang, A real-valued genetic algorithm to optimize the parameters of support vector machine for predicting bankruptcy, Expert systems with applications 32 (2) (2007) 397–408.
[105] X. Zhiyun, G. Ming, Complicated financial data time series forecasting analysis based on least square support vector machine, Journal of Tsinghua University (Science and Technology), Beijing (2008) 82–84.
[106] H. Ince, T.B. Trafalis, Short term forecasting with support vector machines and application to stock price prediction, International Journal of General Systems 37 (6) (2008) 677–687.
[107] W. Zeng-min, W. Chong, Application of support vector regression method in stock market forecasting, in: 2010 International Conference on Management and Service Science, 2010.
[108] L. Breiman, J. Friedman, R. Olshen, C. Stone, Cart, Classification and Regression Trees; Wadsworth and Brooks/Cole: Monterey, CA, USA.
[109] L. Breiman, Bagging predictors, Machine learning 24 (2) (1996) 123–140.
[110] L. Breiman, Random forests, Machine learning 45 (1) (2001) 5–32.
[111] A. Booth, E. Gerding, F. McGroarty, Automated trading with performance weighted random forests and seasonality, Expert Systems with Applications 41 (8) (2014) 3651–3661.
[112] M. Nikou, G. Mansourfar, J. Bagherzadeh, Stock price prediction using DEEP learning algorithm and its comparison with machine learning algorithms, Intelligent Systems in Accounting, Finance and Management 26 (4) (2019) 164–174.
[113] R. Collobert, Deep learning for efficient discriminative parsing, in: Proceedings of the fourteenth international conference on artificial intelligence and statistics, JMLR Workshop and Conference Proceedings, 224–232, 2011.
[114] H.F. Golino, C.M.A. Gomes, D. Andrade, et al., Predicting academic achievement of high-school students using machine learning, Psychology 5 (18) (2014) 2046.
[115] W. Jiang, Applications of deep learning in stock market prediction: recent progress, Expert Systems with Applications 184 (2021) 115537.
[116] J. Schmidhuber, Deep learning in neural networks: An overview, Neural networks 61 (2015) 85–117.
[117] W. Bao, J. Yue, Y. Rao, A deep learning framework for financial time series using stacked autoencoders and long-short term memory, PloS one 12 (7) (2017) e0180944.
[118] T. Fischer, C. Krauss, Deep learning with long short-term memory networks for financial market predictions, European Journal of Operational Research 270 (2) (2018) 654–669.
[119] K. Greff, R.K. Srivastava, J. Koutník, B.R. Steunebrink, J. Schmidhuber, LSTM: A search space odyssey, IEEE transactions on neural networks and learning systems 28 (10) (2016) 2222–2232.
[120] I. Sutskever, O. Vinyals, Q.V. Le, Sequence to sequence learning with neural networks, Advances in neural information processing systems 27.
[121] S. Ji, W. Xu, M. Yang, K. Yu, 3D convolutional neural networks for human action recognition, IEEE transactions on pattern analysis and machine intelligence 35 (1) (2012) 221–231.
[122] C. Szegedy, A. Toshev, D. Erhan, Deep neural networks for object detection.
[123] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 3431–3440.
[124] I. Goodfellow, Y. Bengio, A. Courville, Deep learning, MIT press (2016).
[125] A. Tsantekidis, N. Passalis, A. Tefas, J. Kanniainen, M. Gabbouj, A. Iosifidis, Forecasting stock prices from the limit order book using convolutional neural networks, in: 2017 IEEE 19th conference on business informatics (CBI), vol. 1, IEEE, 7–12, 2017.
[126] O.B. Sezer, A.M. Ozbayoglu, Algorithmic financial trading with deep convolutional neural networks: Time series to image conversion approach, Applied Soft Computing 70 (2018) 525–538.
[127] C.-F. Huang, A hybrid stock selection model using genetic algorithms and support vector regression, Applied Soft Computing 12 (2) (2012) 807–818.
[128] W. Long, Z. Lu, L. Cui, Deep learning-based feature engineering for stock price movement prediction, Knowledge-Based Systems 164 (2019) 163–173.
[129] Y. Baek, H.Y. Kim, ModAugNet: A new forecasting framework for stock market index value with an overfitting prevention LSTM module and a prediction LSTM module, Expert Systems with Applications 113 (2018) 457–480.
[130] S.P. Das, S. Padhy, A novel hybrid model using teaching–learning-based optimization and a support vector machine for commodity futures index forecasting, International Journal of Machine Learning and Cybernetics 9 (1) (2018) 97–111.
[131] A.H. Bukhari, M.A.Z. Raja, M. Sulaiman, S. Islam, M. Shoaib, P. Kumam, Fractional neuro-sequential ARFIMA-LSTM for financial market forecasting, IEEE Access 8 (2020) 71326–71338.
[132] B. Weng, M.A. Ahmed, F.M. Megahed, Stock market one-day ahead movement prediction using disparate data sources, Expert Systems with Applications 79 (2017) 153–163.
[133] M.-W. Hsu, S. Lessmann, M.-C. Sung, T. Ma, J.E. Johnson, Bridging the divide in financial market forecasting: machine learners vs. financial economists, Expert Systems with Applications 61 (2016) 215–234.
[134] A. Picasso, S. Merello, Y. Ma, L. Oneto, E. Cambria, Technical analysis and sentiment embeddings for market trend prediction, Expert Systems with Applications 135 (2019) 60–70.
[135] C. Li, T.-W. Chiang, Complex neurofuzzy ARIMA forecasting–a new approach using complex fuzzy sets, IEEE Transactions on Fuzzy Systems 21 (3) (2012) 567–584.
[136] A. d. Silva Soares, M.S. Veludo de Paiva, C. José Coelho, Technical and Fundamental Analysis for the Forecast of Financial Scrip Quotation: An Approach Employing Artificial Neural Networks and Wavelet Transform, in: International Symposium on Neural Networks, Springer, 1024–1032, 2007.
[137] E.A. Gerlein, M. McGinnity, A. Belatreche, S. Coleman, Evaluating machine learning classification for financial trading: An empirical approach, Expert Systems with Applications 54 (2016) 193–207.