Review

SCIENCE PROGRESS

Science Progress
2024, Vol. 107(1) 1–31

DOI: 10.1177/00368504241236557
journals.sagepub.com/home/sci

A bibliometric literature review of stock price forecasting: From statistical model to deep learning approach

Pham Hoang Vuong[1,2], Lam Hung Phu[1], Tran Hong Van Nguyen[3], Le Nhat Duy[2], Pham The Bao[1] and Tan Dat Trinh[1]

[1]Information Science Faculty, Sai Gon University, Ho Chi Minh City, Vietnam
[2]Faculty of Information Technology, Industrial University of Ho Chi Minh City, Ho Chi Minh City, Vietnam
[3]Faculty of Finance and Banking, Ton Duc Thang University, Ho Chi Minh City, Vietnam

Abstract

We introduce a comprehensive analysis of several approaches used in stock price forecasting, including statistical, machine learning, and deep learning models. The advantages and limitations of these models are discussed to provide an insight into stock price forecasting. Traditional statistical methods, such as the autoregressive integrated moving average and its variants, are recognized for their efficiency, but they also have some limitations in addressing non-linear problems and providing long-term forecasts. Machine learning approaches, including algorithms such as artificial neural networks and random forests, are praised for their ability to grasp non-linear information without depending on stochastic data or economic theory. Moreover, deep learning approaches, such as convolutional neural networks and recurrent neural networks, can deal with complex patterns in stock prices. Additionally, this study further investigates hybrid models, combining various approaches to explore their strengths and counterbalance individual weaknesses, thereby enhancing predictive accuracy. By presenting a detailed review of various studies and methods, this study illuminates the direction of stock price forecasting and highlights potential approaches for further studies refining the stock price forecasting models.

Corresponding authors:
Pham The Bao, Information Science Faculty, Sai Gon University, Ho Chi Minh City, Vietnam.
Email: ptbao@sgu.edu.vn

Tan Dat Trinh, Information Science Faculty, Sai Gon University, Ho Chi Minh City, Vietnam.
Email: trinhtandat@sgu.edu.vn

Keywords
Autoregressive integrated moving average, machine learning, deep learning, hybrid model, stock price forecasting

Introduction

The problem of stock price forecasting (SPF) has always been one of the most widely studied issues, involving a comprehensive approach that focuses on the analysis of historical prices, price movements, or trends to forecast future prices.[1] Numerous models and predictions of stock prices have been proposed.[2] Because stock prices move in a random walk,[3] researchers claim that the financial information of the company will be systematically reflected in the current price. According to the Efficient Market Hypothesis (EMH), an efficient market is one where prices always reflect all available information,[4] and it is categorized into three forms of market efficiency: Weak form, semi-strong form, and strong form. In practice, investors and financial practitioners have commonly employed technical analysis and fundamental analysis for SPF or trading decision-making.[5] According to the research in the work of Nti et al.,[6] fundamental analysis is the study of factors influencing supply and demand. Important data used for fundamental analysis include company data such as financial reports, annual company reports, and balance sheets.

A widely used method is time series analysis, which involves techniques for analyzing time series data to extract meaningful statistical attributes and characteristics of the data. The initial approach is decomposing the series, and commonly used methods are the Holt-Winters method[7] or the Census II X-11 method.[8] The autoregressive integrated moving average (ARIMA) approach is a widely used statistical method for analyzing and forecasting time series data.[9,10] While ARIMA has demonstrated its utility in capturing short- to medium-term price trends, it can be difficult to handle the complex dynamics and non-linear patterns often observed in stock markets. To address the shortcomings of conventional SPF systems based on ARIMA approaches, a learning-based approach using machine learning (ML)[11–13] and deep learning (DL) techniques was introduced.[14,15] The ML approaches have shown significant promise in understanding the complexities of financial markets, characterized by dynamic interactions among various elements that influence stock prices. During the 2000s, in comparison to conventional probabilistic or ML approaches, Zuo and Kita[16] employed a Bayesian network (BN) to predict the price-earnings ratio (P/E ratio). A study by Adebiyi et al. investigated the performance of both ARIMA and artificial neural network (ANN) using stock data from the New York Stock Exchange.[17]

Recently, DL approaches were applied to predicting stock prices. DL models can capture complex temporal dependencies and non-linear patterns that are prevalent in stock price movements. Modern models like the convolutional neural network (CNN), the long short-term memory (LSTM) network, and the bidirectional LSTM (BiLSTM) network utilize the approximation of a continuous function and adapt data with fewer assumptions, thereby achieving higher accuracy and efficiency in solving nonlinear issues. Furthermore, a hybrid model for SPF typically integrates multiple predictive modeling techniques to enhance the accuracy and reliability of stock price predictions.[18–20]

Performing a literature review on SPF is a necessary preliminary step before conducting the study or making decisions within this field. This review investigates the development of SPF techniques, ranging from traditional ARIMA methods to advanced DL methods. When starting with a literature review about SPF, we first take advantage of the capabilities of well-known platforms such as Scopus and Google Scholar. Our initial step involves entering targeted keywords into the Scopus database and combining various search terms. These include "stock price forecasting," "stock price prediction," "stock price forecasting using ARIMA," "machine learning in stock prediction," "deep learning model in stock forecasting," "CNN, LSTM in stock price forecasting," "GAN in stock price forecasting," "transformer for stock price forecasting," "graph-CNN in stock price forecasting," and "hybrid models in stock price forecasting." Through the application of specific filters, the search is fine-tuned to align with preferred publication dates, reputable journals, and subject areas. Essential keywords are used to ensure that the reviewed papers are relevant to the topic. Employing a parallel approach with Google Scholar, the same terms are entered, with particular emphasis on the "cited by" feature. This process leads to subsequent research papers that cite foundational works. The focus of our review was on specific forecasting models such as ARIMA, traditional ML, DL, and hybrid models for historical data or a particular stock market. We give priority to selecting papers for review that are primarily from peer-reviewed, reputable journals and conferences. At the same time, we exclude articles published in workshop publications or technical reports. Additionally, the keywords we select for the subject area are mainly related to computer science, engineering, economics, econometrics and finance, business, management and accounting, and decision sciences. Consequently, a total of 110 studies (only English-language papers) were identified, and those published in conferences or book chapters were removed by subject area, or they would be published in articles ($n = 15$).

During the process of accessing these resources, summaries, key findings, methodologies, and significant conclusions are systematically extracted. Our literature review is based on the use of these carefully chosen findings, ensuring it is both comprehensive and based on the most recent developments in SPF. Furthermore, it is important to mention that we exclusively assessed papers introducing new models for time series data forecasting, with a specific focus on forecasting methods for traditional financial data, such as historical data and indicators. We also exclude methods applied to non-traditional sources, such as social media trends, news updates, or news sentiment. The concept of sentiment analysis and the implementation of existing models were not considered in this review. After reading the abstract (and, as needed, other sections), the implementation of inclusion and exclusion criteria yielded 73 papers. Figure 1 shows the search strategy in our study.

Related works

In this section, we present a literature review on several common approaches that have been applied for SPF. Figure 2 shows a flowchart illustrating the common approaches to stock price forecasting outlined in the study.

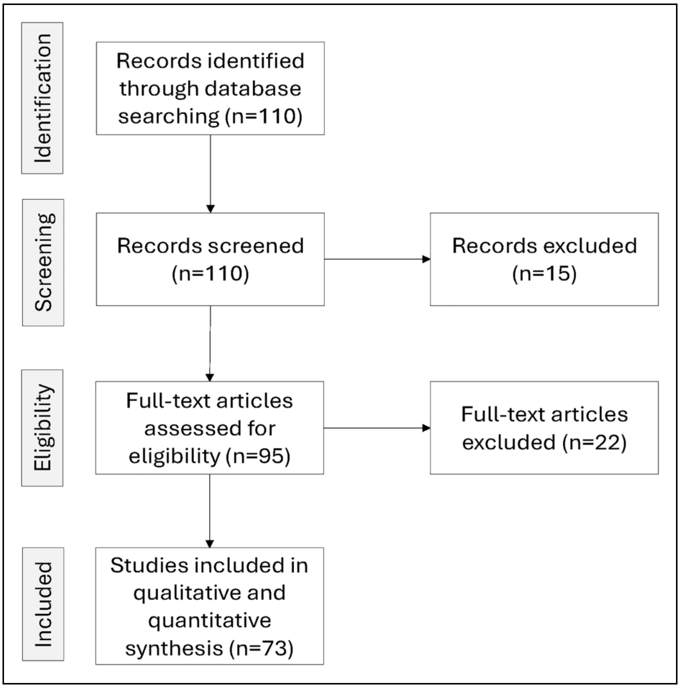

Figure 1. Search strategy for the selection method of the relevant studies.

ARIMA approaches

The well-known traditional statistics time series forecasting methods, such as ARIMA and its variants[17,21–29] are still used a lot because of their efficiency level. Table 1 presents a summary of ARIMA-based approaches for SPF. For the articles reviewed, we summarize the methods used, comparison methods, datasets, target outputs, input features, metrics evaluations, and briefly discuss performance results. Low and Sakk[22] examined the performance of two forecasting models, ARIMA and LSTM, for predicting stock prices. ARIMA is combined with LSTM to determine which is superior in terms of forecasting accuracy. The models were applied to data from ten different stock tickers, specifically exchange-traded funds from various market sectors. The results suggest that ARIMA shows comparable accuracy to LSTM's long-term prediction capabilities. Wahyudi[23] employed the ARIMA model to predict the volatility of Indonesian stock prices. The best ARIMA model is determined using the Akaike information criterion (AIC) criteria. The results indicate that the ARIMA model can compete well with existing techniques for stock price prediction, especially in the short term. Pulungan et al.[24] applied a combination of autoregressive (AR) and moving average (MA) methods. The data needs to be stationary for ARIMA to be applied efficiently. The most fitting ARIMA model for this data was identified as ARIMA (31,1).

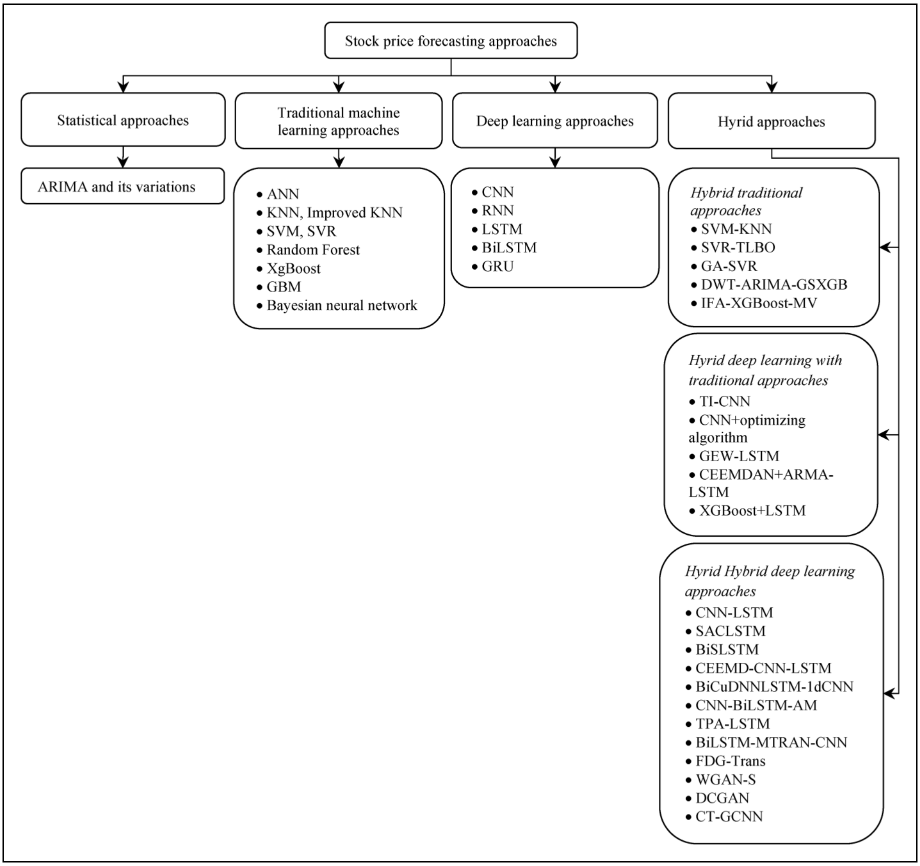

Figure 2. A flowchart covering the various approaches outlined in the study.

Machine learning approaches

ML techniques can capture nonlinear information in time series data without relying on stochastic data or economic knowledge. Thus, ML approaches can be used to build high-performance SPF systems without expert knowledge. The traditional ML algorithms, such as ANNs,[11,30,31] k-nearest neighbors (KNN),[32,33] support vector machine (SVM),[34–40] ensemble models,[41–47] and BN,[48,49] have been successfully and widely used in SPF systems. Table 2 presents articles on SPF based on ML approaches.

Sigo[30] explored the nonlinear movement patterns of three leading stocks on the Bombay Stock Exchange (BSE) in India. ANN is employed to analyze data spanning from 2008 to 2017. The results of the study aim to guide investors in making informed investment decisions and maximizing their returns by focusing on the most valuable stocks. Selvamuthu et al.[31] addressed the challenge of predicting stock prices in the Indian stock market. Recognizing that stock price data is inherently difficult to predict due to its dynamic nature, the authors explored the efficiency of ANN.

When applied to SPF, KNN is used to predict a stock's future price based on its past values. Yunneng[32] presented an enhanced version of the KNN algorithm for stock price predictions. This improvement aims to provide more accurate predictions of stock prices. Lin et al.[33] presented a novel method for improving the accuracy of stock time series forecasting using a multidimensional KNN algorithm. The results showed that the proposed method outperformed the other models in predicting stock prices, proving to be a more reliable and effective forecasting system.

SVMs have been primarily developed for classification problems, but their application has been extended to regression problems known as support vector regression (SVR). SVR can be applied to SPF. Xiao et al.[34] introduced a novel methodology for stock price analysis and forecasting, combining singular spectrum analysis (SSA) and SVM. Ismail et al.[35] aimed to predict the direction of stock price movement. The study introduced a hybrid method that combines various ML techniques—namely logistic regression (LR), ANN, SVM, and random forest (RF)—to enhance prediction accuracy.

Developing an ensemble model for SPF involves aggregating the predictions from multiple models to improve the accuracy and robustness of the predictions. Syukur and Istiawan[42] investigated the prediction of the LQ45 index on the Indonesia Stock Exchange (ISX) using various ML techniques. RF was found to obtain the best performance in predicting the LQ45 index compared to C4.5, SVM, LR, Naïve Bayes (NB), and linear discriminant analysis (LDA).

In the context of SPF, Bayesian neural networks (BNNs) enable the prediction of the likelihood of various stock prices, based on given evidence or observed variables. Chandra and He[48] explored the utilization of BNNs for forecasting stock prices. Malagrino et al.[49] explored the potential of BNNs to understand the influence of global stock market indices on iBOVESPA, the primary index of the São Paulo Stock Exchange in Brazil. The objective is to forecast the closing direction of iBOVESPA the next day. The BNN models were able to achieve a mean accuracy of around 71%, with a peak accuracy of nearly 78%, in predicting the daily closing direction of iBOVESPA.

Deep learning approaches

A DL model can effectively outperform traditional SPF systems in terms of accuracy. CNNs[50–54] and recurrent neural network (RNN) such as LSTM or gated recurrent unit (GRU),[55–60] and BiLSTM[61–64] are extensively employed for SPF systems. Table 3 presents a summary of DL-based approaches for SPF.

CNNs are commonly used for image and video processing; however, their proficiency in identifying hierarchical patterns can extend to time series forecasting as well. Gunduz et al.[50] applied a CNN model that utilized specially ordered features derived from various indicators, prices, and temporal information appropriate to stocks in the Borsa Istanbul 100. Wen et al.[51] presented an approach to forecasting stock market trends utilizing financial time series data, exemplified by the S&P 500. CNN explored to distinguish the spatial structure inherent in the time series.

RNNs have been proposed for SPF systems. Results have demonstrated that methodologies based on RNN can outperform classic ML techniques. RNNs can handle

Table 1. Summary of the existing ARIMA-based stock price forecasting approaches.

Reference No.	Method	Comparisons	Dataset	Targets	Input features	Metrics	Results
21	ARIMA	LSTM, SARIMAX	Stock exchange market data from Yahoo Finance	Closing values of stock prices	Index, Open, Close, Adj Close, High, Low, Volume and Close USD	MAE	Error for ARIMA is less as compared to SARIMAX
22	ARIMA	LSTM	Ten different stock tickers comprising exchange-traded funds	Stock price prediction	Closing price	MSE	ARIMA was found to be more accurate in making point predictions
23	ARIMA	ARIMA (0, 1, 1) ARIMA (1, 1, 0) ARIMA (1, 1, 1)	Daily Indonesia CSPI	Daily movement of stock prices	Closing price	MAPE	ARIMA (01,1): 0.8431
24	ARIMA	ARIMA (31,1)	Socially Responsible Investment Index (SRI-KEHATI) on the Indonesia Stock Exchange	Daily closure of the SRI-KEHATI Index data	Closing price	Ljung-Box Q statistical test	ARIMA (31,1) had a significant effect on the SRI-KEHATI Index

ARIMA: autoregressive integrated moving average; MAPE: mean absolute percent error; MSE: mean squared error; MAE: mean absolute error; LSTM: long short-term memory; CSPI: composite stock price index.

sequences of variable length, offering flexibility in managing time series data of diverse lengths. To overcome the vanishing gradient problem intrinsic to RNNs, LSTM networks, GRU, and their variants were developed. The utilization of LSTMs has been substantiated as effective in accurately forecasting stock prices. Ghosh et al.[55] demonstrated the efficacy of employing both LSTM networks and RFs to forecast directional movements of stock prices from the S&P 500 index for intraday trading. Authors[56] proposed an optimized approach for predicting stock prices using advanced DL techniques, such as LSTM and GRU models. The authors employed DL LASSO and principal component analysis (PCA) for dimensionality reduction, focusing on various factors influencing stock prices.

BiLSTM is often used for sequence-to-sequence learning tasks, like SPF. The BiLSTM allows the model to capture both past and future information around a specific time step, potentially enhancing the model's ability to understand the underlying patterns in the sequence. Xu et al.[62] focused on utilizing a stacked DL structure for stock market predictions, specifically aiming to predict the stock price of the subsequent day. This model employs historical stock price data sourced from Yahoo Finance and integrates several methodologies, including the wavelet transform technique, stacked autoencoder, RNN, and BiLSTM. Liu et al.[64] employed an auto-encoder (AE) technique to extract stock price series data, recognizing its proficiency in managing the non-smooth and nonlinear characteristics inherent in the data. The core structure of the AE incorporates a BiLSTM module, which allows the model to efficiently extract substantial historical and prospective information from stock price series data.

Hybrid approaches

In SPF, hybrid models refer to combinations of different models aiming to leverage the strengths and reduce the drawbacks of individual methods. Hybrid models can achieve higher predictive accuracy than single models. However, they also come with challenges, such as increased model complexity, potential difficulties in model interpretation, and the requirement for extensive tuning and validation. These hybrid models can generally be categorized into two main types: hybrid traditional approaches and hybrid DL approaches. Table 4 presents a summary of hybrid-based approaches for SPF.

Hybrid traditional approaches[65–70] typically combine traditional statistical methods with ML techniques, or they combine various ML approaches with each other. Nayak et al.[65] introduced a hybrid model that integrates both the SVM and KNN techniques for predicting Indian stock market indices. The model's performance was evaluated using the mean squared error (MSE), and it was found that the SVM-KNN model outperformed several baseline models. Siddique and Panda[66] compared various hybrid ML models for prediction. These models utilized dimension reduction techniques such as orthogonal forward selection (OFS) and kernel PCA (KPCA). They were combined with SVR and teaching-learning-based optimization (TLBO). The study concluded that the model incorporating KPCA (KPCA-SVR-TLBO) outperformed and was more feasible than the model employing OFS (OFS-SVR-TLBO).

Hybrid DL approaches frequently combine DL techniques with traditional methods[71–75] or DL architectures with each other, such as CNN-LSTM, LSTM or BiLSTM with

Table 2. Summary of the existing ML-based stock price forecasting approaches.

Reference No.	Method	Comparisons	Dataset	Targets	Input features	Metrics	Results
30	ANN	SPSS statistics tool	Bombay Stock Exchange Limited	Future direction of the stock price movements	Opening price, high price, low price, and closing price	AAE, MAE, RMSE	ANNs provide higher accuracy
31	ANN	ANN_SCG, ANN_LM, ANN_BR	Reliance Private Limite from Thomson Reuter Eikon	Stock prices and movements	Tick Data, and 15-min Data	MAPE, MSE	ANN_SCG obtained best performance
32	KNN	Baseline KNN, regression prediction	Historical data of stock Neimengyiji	History	High, low, open, and close	Standard error	Improved KNN yielded the best result
33	EEMD–MKNN–TSPI	EEMD–MKNN, MKNN–TSPI	NAS, DJI, S&P 500, Russell 2000; and stock data from 04 regions	Opening and closing price	Opening and closing prices	MAPE, MASE, NMSE	EEMD–MKNN–TSPI model outperforms the EEMD–MKNN and MKNN–TSPI models
34	SSA–SVM	ANFIS, SVM, EEMD–ANFIS, EEMD–SVM, and SSA–ANFIS	Shanghai Stock Exchange Composite Index	Daily closing price	Closing price	MSE, MAPE, DS, R^2	SSA–SVM model exhibiting the best prediction performance
35	SVM	LR, ANN, RF	Kuala Lumpur Composite Index, Kuala Lumpur Stock Exchange Industrial, Kuala Lumpur Stock Exchange Technology	Next day movement	Stock returns, technical indicators, connected components, Holes	Average of the prediction performances	Support vector machine with persistent homology generates the best outcome
42	Random Forest	LR, LDA, NB, KNN, K*, C4.5, CART, ANN, SVM	Indonesia Stock Exchange	Prediction of the LQ45 index	15 variables (volum, value, …)	Accuracy, recall, precision	RF had the best performance
43	Random Forest	XGBoost, Bagging Classifier, AdaBoost, Extra Trees Classifier, Voting Classifier	NYSE, NASDAQ, NSE	Direction of stock price movement	40 technical indicators and the OHLCV variables	Accuracy, precision, f1-score, specificity, and AUC	Extra Trees classifier outperformed the other models

(Continued)

Table 2. (continued)

Reference No.	Method	Comparisons	Dataset	Targets	Input features	Metrics	Results
[48]	Bayesian neural network	FNN-Adam, FNN-SGD	3 M Company, China Spacesat Company Limited, Commonwealth Bank of Australia, Daimler AG	Future trends of stock price	Adjusted closing price	RMSE	Bayesian neural network provided one of the best performances
[49]	Bayesian neural network	ANN, SVM, KNN, decision tree, NB, …	12 indices: Nasdaq Composite, NYSE Composite, Dow Jones, …	Next day closing direction	Closing direction	Accuracy	The mean accuracy was around 71%

ANN: artificial neural network; KNN: k-nearest neighbor; SVM: support vector machine; SSA: singular spectrum analysis; RF: random forest; NB: Naïve Bayes; LDA: linear discriminant analysis; MAPE: mean absolute percent error; RMSE: root mean squared error; MSE: mean squared error; MAE: mean absolute error; AUC: area under the receiver operating characteristic curve; SPSS: statistical product and services solutions.

attention mechanisms (AMs), transformer models, and graph convolutional neural network (GraphCNN).[76–90] These hybrid DL models prove to be efficient in identifying complex patterns and relationships in data due to the high capacity and adaptability of DL architectures, especially in applications like SPF. Chandar[71] proposed a new method for stock trading by combining technical indicators and CNNs, termed TI-CNN. The model uses ten technical indicators derived from historical stock data, converts them into an image using gramian angular field, and then inputs this into the CNN. Korade and Zuber[72] explored the usage of CNN for SPF and aim at optimizing the CNN hyperparameters using different optimization techniques. The authors employ the firefly algorithm (FF), particle swarm optimization, and random search for optimizing the hyperparameters, comparing their performance based on different evaluation metrics applied to training and testing datasets.

The study by Lu et al.[76] proposed a method for forecasting stock prices utilizing a hybrid CNN-LSTM model. This model utilizes CNN for efficient feature extraction from historical data and LSTM to analyze relationships in time-series data, subsequently predicting stock prices. Wang et al.[78] aimed to predict the closing price of stocks using a composite model called CNN-BiSLSTM. Here, the BiSLSTM represents bidirectional special LSTM.

The integration of AMs with LSTMs in SPF models presents the possibility of improved prediction accuracy and reliability. An attention-LSTM model can analyze historical stock prices and potentially other relevant information to predict future stock prices. Lu et al.[81] discussed a combined approach using CNN, BiLSTM, and attention mechanism for predicting stock prices. The results showed that the CNN-BiLSTM-AM method outperforms seven other methods in accuracy. The study referenced by Chen et al.[82] introduced a novel model for predicting stock prices, utilizing a CNN, a BiLSTM, and an efficient channel attention (ECA) module.

Transformers were developed to reduce the limitations inherent to AMs and recurrent models like RNNs. Specifically, they address the challenges brought about by the inherent sequential processing of RNNs and the high computational demands of AMs, allowing for more efficient and scalable modeling of sequential data. When employed for SPF, transformer models are good at identifying complex patterns within time series data and understanding the long-term dependencies existing between various time steps. Wang[85] introduced a novel method named BiLSTM-MTRAN-TCN for predicting stock prices. This method used BiLSTM, an improved transformer model (MTRAN-TCN), and TCN (temporal convolutional network), aiming to explore the individual benefits of each model. Li and Qian[86] introduced a novel hybrid neural network—the FDG-transformer—specifically developed for predicting stock prices.

Recently, generative adversarial networks (GANs) and GraphCNN have been applied for SPF, often achieving high accuracy. GANs can be used to generate synthetic time-series data that mimics real stock price movements. This synthetic data can help augment the training data, allowing models to generalize better to unseen data and potentially leading to more accurate forecasts. The Wu et al.[87] introduced a novel framework that combines GAN with piecewise linear representation for predicting stock market trading actions such as buying, selling, and holding. Staffini[88] proposed a novel approach to predicting stock prices using a deep convolutional GAN (DCGAN). The generator model of the GAN learns to generate data like real stock prices. The discriminator model learns to distinguish between real and generated stock prices. The results show that the proposed DCGAN model outperformed standard, widely used tools for forecasting stock prices.

GraphCNNs extend convolutional operations from regular grids to irregular graphs. This advancement allows models to effectively capture the relational structures and dependencies between different entities. GraphCNNs are particularly applicable to SPF, where they can model the relationships between different stocks or between different features of a single stock. The work of Wang et al.[89] proposed a new model for stock price prediction, integrating a knowledge graph, GraphCNN, and community detection. This model aims to overcome the limitations of existing models, which often neglect deeper influencing factors and rely on small-scale stock datasets. Fuping[90] concentrated on predicting stock price movements using a novel method called the conceptual-temporal graph CNN (CT-GCNN) model. This model explores stock price movements in both time and concept dimensions, accounting for the linkage effect of price movement among stocks within the same conceptual segment.

Dataset and metric evaluation

Statistics of selected papers

This subsection investigates a statistical analysis of 73 papers, specifically selected for this review on SPF, utilizing analyze tools from the Scopus platform. Figure 3 illustrates the annual publication count, revealing a conspicuous upward trend in papers on SPF from 2014 to 2023. The number of papers increased from a single publication in 2014 to 18 in 2023.

Figure 4 illustrates the distribution of published papers, categorized by source (compare the document counts for the top 10 sources). Research on SPF has been published in various journals and conferences. Particularly notable is the number of studies in the 'Expert Systems with Applications' journal. The variety of journals, spanning specialized and multidisciplinary fields such as computing, economics, and finance, underscores the interdisciplinary essence of the research.

Research on SPF is primarily distributed through articles (61), indicating a robust presence in academic journals, followed by conference papers (11), which suggest active discussions and explorations in conference settings, and is minimally represented in book chapters (1), as shown in Figure 5. The dominance of articles points towards depth and precision in the field, while conference papers highlight ongoing, dynamic discussions and potential collaborations among researchers.

Figure 6 illustrates the number of published papers by authors from various countries or territories (compare the document counts for the top 15 countries or territories). This shows a geographical diversity in the research within this field. China, with 24 publications, and India, with 17, lead in stock market research, reflecting their strong economic interests and growth. Other countries also show global interest in SPF. This figure shows that this research topic has widespread relevance and interest across different economic contexts.

Figure 7 illustrates the number of published papers by authors, categorized by affiliation (compare the document counts for the top 15 affiliations), with leading contributions from Capital University of EcoNomics and Business and Hebei University of Science and Technology at three publications each. Several other universities from around the world have also made notable contributions. This figure implies a robust and diverse global effort in the study of SPF, reflecting both academic and industry interests in the area.

Dataset

Many studies have used stock market data for their analyses, as shown in Tables 1–4. This research examines a wide range of datasets sourced from various stock indices and companies. These datasets have been used in multiple studies that focus on SPF. They include noticeable indices such as the FTSE MIB Index and A-share market stock prices in China, as well as specific companies like Amazon, Apple, Google, Tesla, IBM, and Oracle. The data for these studies is often sourced from platforms like Yahoo Finance.

Earlier studies have utilized a diverse set of data types. These range from closing prices, opening prices, highs, lows, and trading volumes to other technical indicators. Some even incorporate historical and tick data. Data from major global stock exchanges, including the Shanghai Stock Exchange, Kuala Lumpur Stock Exchange, and BSE Limited, plays an important role in these studies. Some studies specifically aim to predict the direction or price of the next day's close. Table 5 presents some stock datasets and their source links.

Metric evaluation

From Tables 1–4 in the "Related works" section, we realize that various metrics are utilized to evaluate the performance of different models in SPF. These include mean absolute percent error (MAPE), mean squared error (MSE), root mean square error (RMSE), mean absolute error (MAE), accuracy, R^2 (coefficient of determination), mean bias error (MBE), precision, recall, F1-score, and area under the receiver operating characteristic curve (ROC-AUC), among others.

- Metrics for evaluating prediction accuracy: MAPE, MSE, RMSE, and MAE are widely employed across various models and approaches to assess the accuracy and performance of stock price predictions.
- Metrics for classification models: Accuracy, precision, recall, F1-score, and ROC-AUC are typically used to evaluate models where classification, such as predicting upward or downward trends, is involved.
- Understanding model predictions and biases: R^2 is utilized to comprehend how well the model's predictions align with actual outcomes, providing insights into the explanatory power of the model by understanding the proportion of variability in the dependent variable that is predictable from the independent variable(s).

Discussion and future directions

Method-Based discussion

ARIMA approaches. A review of ARIMA models reveals that both ARIMA and its enhanced variants[21,22] have been widely utilized in SPF. ARIMA often demonstrates a high degree of accuracy in short-term forecasts and is particularly efficient for linear and stationary series, making it apt for numerous financial time series data. The models facilitate the use of evaluative parameters, such as the AIC[23,25,27] and the Bayesian information criterion (BIC),[17,27] to select the most fitting models, thereby ensuring optimal performance. However, ARIMA models are unsuitable for non-linear and non-stationary data, restricting their application since financial time series often display non-linear behaviors.[17] They may not effectively capture evolving trends and patterns in the long term. Moreover, identifying the correct order of differencing and the appropriate number of AR and MA terms (p, d, q parameters)[27,75] can be challenging and necessitates expertise, making it less accessible to non-experts.

Machine learning approaches. ML can discern and model nonlinear relationships in stock prices, which poses challenges for ARIMA methods.[30,34] Many ML techniques, including SVM, KNN, ANN, and RF, provide a variety of strategies for learning. Each model has its own strengths in predicting stock prices.[31,33,34,41] However, it is important to note that due to the complexity of some ML models, they may fit the training data too closely and fail to generalize well to new, unseen data.[69,70] ML models can be highly sensitive to noise in the data, leading to inaccurate predictions. Training and optimizing ML models can be computationally intensive, demanding substantial resources and time, especially for large datasets.[45,49]

DL approaches. DL models have demonstrated their ability to predict stock prices with high accuracy and often outperform traditional models.[50,55,59,63] These models can deal with the inherent non-linear and non-smooth features of stock price data.[61,64] Their versatility enables them to handle various data types and structures, utilizing diverse variable sets from different markets.[19] Moreover, they provide flexibility in exploring time series data of varying lengths, which is particularly beneficial for stocks with inconsistent trading histories.[53,63,89]

However, DL models require significant computational power and resources for training, which may not be accessible to all.[51,61] The efficacy of these models heavily depends on the quality and quantity of the data; insufficient data can reduce their performance. The complexity of DL models can result in overfitting, particularly when the model is very complex relative to the simplicity of the task or the volume of available data. Some inherent issues, like the vanishing gradient problem in basic RNNs, necessitate the use of more advanced variants,[62,64] increasing the complexity of model development and implementation.

Hybrid approaches. Hybrid models frequently outperform single models by combining the strengths of multiple approaches. By utilizing a variety of models, hybrid models can generalize more effectively to unseen data, thereby reducing the risk of overfitting.[68,76,78] Especially hybrid models that incorporate ML and DL can explore non-linear relationships in data, a common occurrence in financial markets. DL hybrids, such as CNN-LSTM, can extract hierarchical features and comprehend sequential dependencies, making them suitable for time-series forecasting like stock prices.[76,77] AMs[81,82] and transformers[85,86] have demonstrated promise in investigating sequential data by allocating varying levels of importance to different time steps, which is pivotal in financial time-series data.

However, integrating multiple models can lead to increased complexity, making them computationally expensive and more challenging to manage.[69,79,88] These models often necessitate tuning and validation, which can be resource intensive. DL hybrids might

Table 3. Summary of the existing DL-based stock price forecasting approaches.

Reference No.	Method	Comparisons	Dataset	Targets	Input features	Metrics	Results
50	CNN	LR, CNN-Rand, CNN-Corr, LR With FS	BIST 100 Index	Hourly stock price direction	25 technical indicators with different time lags	Macro-Averaged F-Measure	CNN-Corr classifier yielded the best performance
51	CNN + frequent patterns	ARIMA, Wavelet + ARIMA, HMM, LSTM, SFM	*S&P* 500 and 07 individual stocks	Trend of stock price	Closed value	Accuracy, recall, precision, f1-score	Proposed method outperformed the others with a 4%–7% accuracy improvement
55	LSTM	Random Forest	S&P 500	Directional movements of stock price	Adjusted closing prices and opening prices	Various metric (mean, std error, sharpe ratio, …)	LSTM outperforms random forests
56	LSTM	LASSO-LSTM, PCA-LSTM, LASSO-GRU, PCA-GRU	Shanghai Composite Index	Stock price trend	Open, high, low, trading volume, and other technical indicators	RMSE, MAE	LSTM and GRU with LASSO yielded better accuracy than models with PCA
62	BiLSTM	WAE-BLSTM, W-BLSTM, W-LSTM, BLSTM, LSTM	S&P500	Next day closing price	Open, high, low, close (OHLC), 08 technical indicators	MAE, RMSE, R^2	WAE-BLSTM model outperformed the other models. MAE (0.0211), RMSE (0.0272), and R^2 (0.8934)
64	AE-BiLSTM-ECA	CNN, LSTM, BiLSTM, CNN-LSTM, AE-LSTM, CNN-BiLSTM, AE-BiLSTM, BiLSTM-ECA,	Shanghai Stock Composite Index (SSCI) and CSI 300	Closing price	Seven characteristics such as closing, high, open, low, previous day's closing price, up or down amount and and	MSE, RMSE, MAE, MAPE	AE-BiLSTM-ECA obtain the best accuracy. CSI 300 stock data: MSE: 3158.452 RMSE: 56.200 MAE: 36.681

(Continued)

Table 3. (continued)

Reference No.	Method	Comparisons	Dataset	Targets	Input features	Metrics	Results
		AE-LSTM-ECA, CNN-LSTM-ECA			up or down rate		MAPE: 1.020 SSCI stock data MSE: 1935.398 RMSE: 43.993 MAE: 28.940 MAPE: 1.019

ARIMA: autoregressive integrated moving average; BiLSTM: bidirectional long short-term memory; LR: logistic regression; GRU: gated recurrent unit; PCA: principal component analysis; AE: auto-encoder; ECA: efficient channel attention; MAPE: mean absolute percent error; MSE: mean squared error; RMSE: root mean squared error; MAE: mean absolute error; DL: deep learning.

Table 4. Summary of the existing hybrid-based stock price forecasting approaches.

Reference No.	Method	Comparisons	Dataset	Targets	Input features	Metrics	Results
Hybrid traditional approaches							
65	SVM-KNN	CEFLANN, FLIT2FNS	Bombay Stock Exchange (BSE Sensex) and CNX Nifty	Trends, volatility, and momentum of stock indices	Open, low, high, closing, and technical indicators	MAPE, MSFE, RMSFE	SVM-KNN has better performance than CEFLANN and FLIT2FNS
66	SVR-TLBO	OFS-SVR-TLBO, KPCA-SVR-TLBO	Tata Steel from Bombay Stock Exchange	Closing price	07 features	MAE, RMSE, MAPE	KPCA-SVR-TLBO performed better than OFS-SVR-TLBO
Hyrid deep learning with traditional approaches							
71	TI-CNN	CNN-TA, 1D CNN, CNN-LSTM	NASDAQ, NYSE	Stock movement, Buying and selling points	10 technical indicators	Accuracy, f1-score	TI-CNN achieves high prediction accuracy
72	CNN + optimizing algorithm	CNN, RS-CNN, FF-CNN, PSO-CNN	Tata Motors from Yahoo Finance	Closing price	date, open, low, close, high, volume, and adjacent close	MSE, MAE, RMSE, MAPE	FF-CNN outperformced the others
Hybrid deep learning approaches							
76	CNN-LSTM	MLP, CNN, RNN, LSTM, CNN-RNN	Shanghai Composite Index	Next day closing price	Open, high, low, closing price, volume, turnover, ups and downs, and change of the stock data	MAE, RMSE, R^2	CNN-LSTM obtained the best performance MAE(27.564), RMSE(39.688), R2(0.9646)

(Continued)

Table 4. (continued)

Reference No.	Method	Comparisons	Dataset	Targets	Input features	Metrics	Results
77	SACLSTM	SVM, CNN-cor, CNNpred, ANN	10 stocks from American market and Taiwan	Direction of the stock market (rise and fall)	Historical data, futures, and options	Accuracy	SACLSTM performs relatively well compared to the others
78	BiSLSTM	MLP, RNN, LSTM, BiLSTM, CNN-LSTM, CNN-BiLSTM	Shenzhen Component Index	Closing price	Open, high, low, closing, volume, turnover, ups and downs, and change	MAE, RMSE, R^2	CNN-BiSLSTM has optimal values for MAE (113.47137), RMSE (162.53164), R2 (0.98634)
81	CNN-BiLSTM-AM	MLP, CNN, RNN, LSTM, BiLSTM, CNN-LSTM, CNN- BiLSTM, BiLSTM-AM	Shanghai Composite Index	Next day stock closing price	Open, high, low, closing, volume, turnover, ups and downs, and change	MAE, RMSE, R^2	CNN-BiLSTM-AM yielded the best results MAE(21.952), RMSE(31.694), R2(0.9804)
82	CNN-BiLSTM-ECA	CNN, LSTM, BiLSTM, CNN-LSTM, CNN-BiLSTM, BiLSTM-ECA, CNN-LSTM-ECA, CNN-BiLSTM-ECA	Shanghai Composite Index, China Unicom, CSI 300	Next day closing price	Closing, high, low, open, previous day's closing price, change, ups and downs, and other time series data	MSE, RMSE, MAE	CNN-BiLSTM-ECA obtained the best performance
85	BiLSTM-MTRAN-CNN	BiLSTM-SA-TCN, CNN-BiLSTM, CNN-BiLSTM-AM, BiLSTM	A-share Index, Shanghai Composite Index, Shenzhen Component Index, CSI 300 and	Next day closing price	Trading data and technical indexes data	MAE, MSE, RMSE, R^2	BiLSTM-MTRAN-TCN outperforms the other methods

(Continued)

Table 4. (continued)

Reference No.	Method	Comparisons	Dataset	Targets	Input features	Metrics	Results
			Growth Enterprise Board Index				
86	FDG-Trans	DeepLOB, DeepAtt, MHF	Limit order book (LOB), CSI-300	Price movements.	LOB information	R^2, MSE, MAE	The FDG-Trans has less error compared to the other models
87	WGAN-S	H-LSTM, GAN, GAN-S, LSGAN, LSGAN-S, WGAN	Taiwan Stock Exchange Capitalization Weighted Stock Index	Three trading actions: buying, selling, and holding	Opening, closing, highest, lowest, trading volume, and technical indices	Cumulative return on investment, the Sharpe ratio, and winning percentage.	GAN outperform LSTM
88	DCGAN	ARIMAX-SVR, RF regressor, LSTM, GAN	FTSE MIB Index	Closing price	Technical indicators	RMSE, MAE, MAPE	DCGAN obtained the best performance
89	Improve GCNN	MOM, MR, LSTM, DARNN, SFM, GCN, TGC, HATS, STHGCN	A-share market stock prices in China	Trend of stock price	Open, high, low, close, and trading volume	Accuracy, recall, precision, f1-score, AUC	Proposed model achieves the best accuracy,
90	CT-GCNN	GNN, LSTM, RNN, CNN, BP	A-share market stock prices in China	Stock price movement	Open, close, exchange rate, high, low, trading volume	MSE, MAPE	CT-GCNN model demonstrated stability and superiority

CT-GCNN: conceptual-temporal graph convolutional neural network; BiLSTM: bidirectional long short-term memory; KNN: k-nearest neighbor; SVM: support vector machine; SVR: support vector regression; RF: random forest; RNN: recurrent neural network; KPCA: Kernel principal component analysis; OFS: orthogonal forward selection; teaching-learning-based optimization; FF: firefly algorithm; PSO: particle swarm optimization; RS: random search; ECA: efficient channel attention; DCGAN: deep convolutional generative adversarial network; GAN: generative adversarial network; MAPE: mean absolute percent error; MSE: mean squared error; RMSE: root mean square error; MAE: mean absolute error; AUC: area under the receiver operating characteristic curve

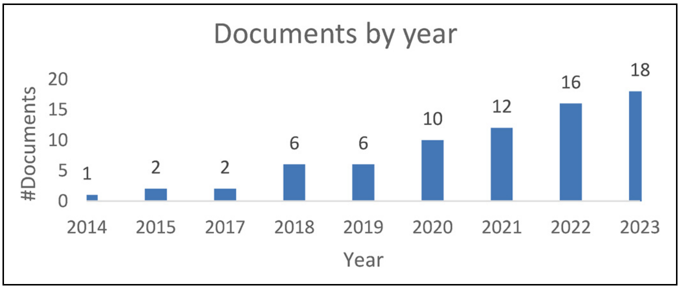

Figure 3. Number of published papers by year.

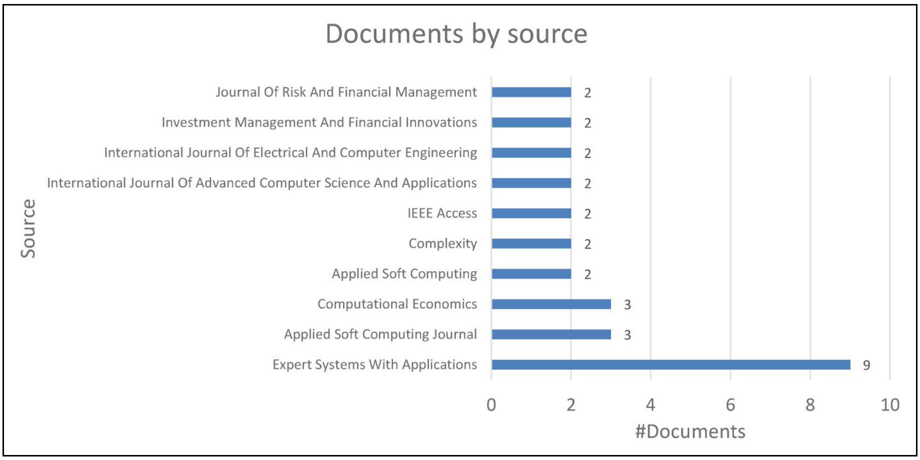

Figure 4. An example of number of published papers by source.

require substantial amounts of data for effective training, which might pose a limitation in certain scenarios.

Research limitations

The main focus of our review was on ARIMA, traditional ML, DL, and hybrid models for historical data for time series analysis in SPF. However, alternative approaches such as the EMH, fundamental analysis, a combination of technical and fundamental analysis, and sentiment and social data analysis were not given prominence in our investigation. Our review specifically evaluates the effectiveness of methods utilizing traditional financial data, such as historical data and indicators. The exclusion of non-traditional sources, such as social media trends, news updates, sentiment analysis, or real-time market indicators, presents limitations. It overlooks important factors influencing investor behavior and market trends. In addition, the study overlooks advanced computational approaches such as quantum algorithms and reinforcement learning techniques. Finally, the study faced several limitations in selecting

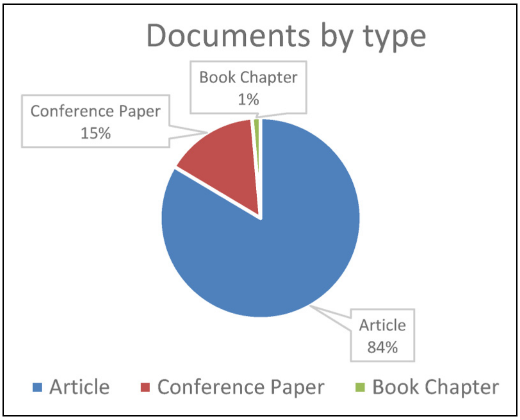

Figure 5. Number of published papers by type.

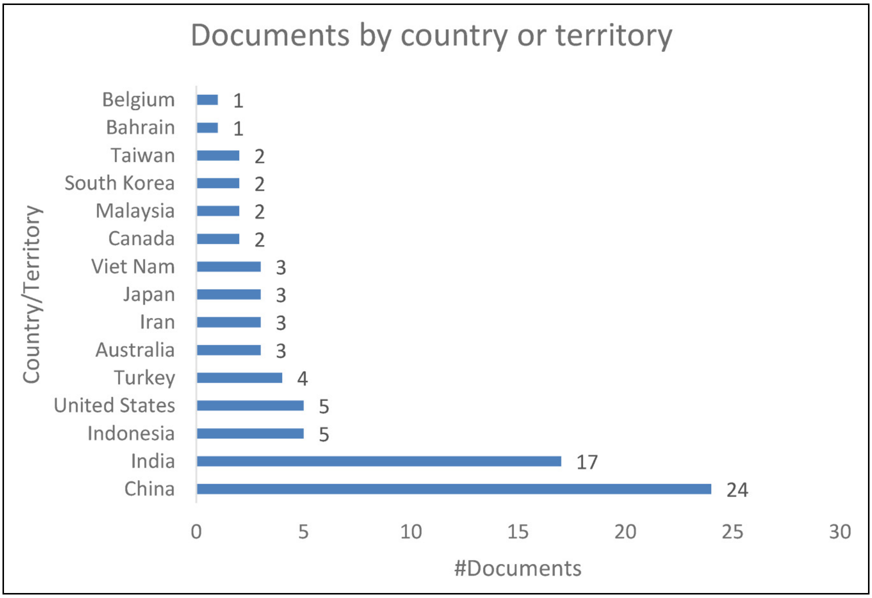

Figure 6. An example of the number of published papers by authors, categorized by country or territory.

papers and excluding irrelevant documents from the large number of returned results from Scopus. We primarily focused on journal papers, excluding other sources like conference papers, workshop publications, book chapters, and technical reports, which might have led to missing relevant articles due to the narrow scope of sources. We recognize that the search strategy used in our study may have led to the omission of relevant articles.

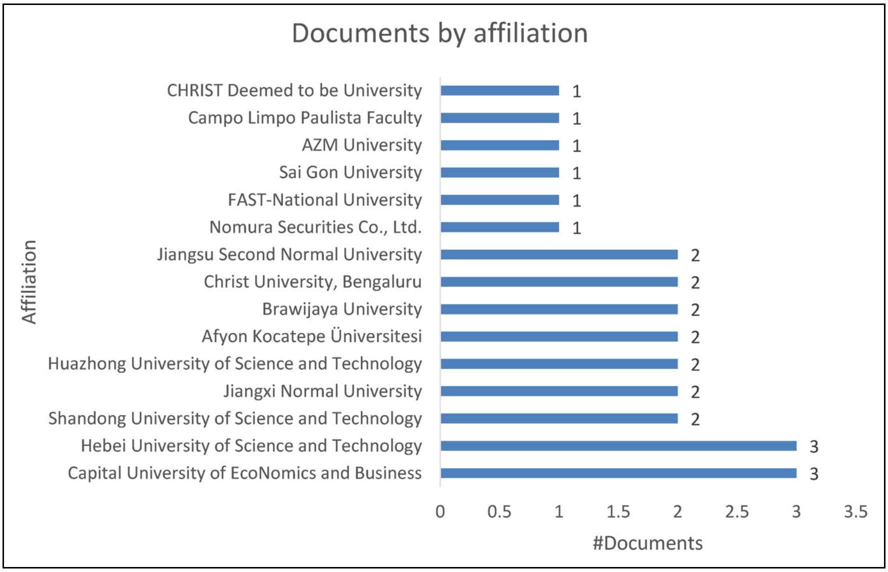

Figure 7. An example of the number of published papers by authors, categorized by affiliation.

Future directions

Forecasting stock prices is a complex task due to the myriad of factors influencing market dynamics, presenting numerous challenges. One significant challenge is managing market noise,[19] as prices are influenced by both relevant information and irrelevant or random data, making the differentiation between impactful and non-impactful information complex. The accuracy of predictions is also relied on the quality and completeness of the data, with the handling of missing data, outliers, and incorrect data posing persistent challenges.

The challenges within stock price prediction have catalyzed the exploration of new methodologies and approaches, all aimed at enhancing predictive accuracy and reliability. Here are some future directions for SPF:

+ In addition to employing time series data for forecasting, it is important to explore and use alternative data sources for stock price prediction. Specifically, exploiting sentiments extracted from various platforms, such as social media and news outlets[91–94] can significantly enhance our ability to comprehend and accurately predict market movements.

+ The combination of traditional financial theories with AI algorithms—including transformers,[85,95] graphCNN,[77,96] reinforcement learning,[97,98] meta-learning,[99,100] leads to enhanced performance in SPF.

+ Real-time analysis for SPF provides a range of advantages.[101,102] It enables timely decision-making by offering up-to-the-minute data, allowing traders to react quickly to market changes and news events.

+ As quantum computing technology continues to advance, researchers can explore the application of quantum algorithms to optimize trading strategies and enhance the efficiency of SPF models.[103–106]

Table 5. Some stock datasets and their source links.

Dataset/Market capitalization	Source	Dataset/Market capitalization	Source
Stock exchange data	https://www.kaggle.com/datasets/mattiuzc/stock-exchange-data https://finance.yahoo.com/	3M Company	https://finance.yahoo.com/quote/MMM/
Exchange Traded Funds	https://www.sectorspdrs.com/	China Spacesat Co., Ltd (600118.SS)	https://finance.yahoo.com/quote/600118.SS/
Indonesia Stock Market (JCI), Composite Stock Price Index (CSPI)	https://tradingeconomics.com/indonesia/stock-market	BIST 100 (XU100.IS)	https://finance.yahoo.com/quote/XU100.IS/
SRI-KEHATI Index	https://www.bloomberg.com/quote/BNPSRIK:IJ	CSI 300 Index	https://www.bloomberg.com/quote/SHSZ300:IND
Bombay Stock Exchange (BSE Sensex)	https://www.bseindia.com/	NIFTY 50	https://g.co/finance/NIFTY_50:INDEXNSE
Thomson Reuter Eikon	https://eikon.refinitiv.com/	Tata Motors Limited	https://finance.yahoo.com/quote/TATAMOTORS.BO/
Norwegian Air Shuttle ASA (NAS)	https://finance.yahoo.com/quote/NAS.OL/	China Unicom Hong Kong Ltd	https://www.bloomberg.com/quote/762:HK
Dow Jones Industrial Average (DJIA)	https://finance.yahoo.com/quote/%5EDJI/	Shanghai SE A Share	https://www.investing.com/indices/shanghai-se-a-share
Standard & Poor's 500 Stock Index (S&P 500)	https://www.marketwatch.com/investing/index/spx	Growth Enterprise Market	http://www.aastocks.com/en/stocks/market/index/hk-index-con.aspx?index = GEM
Russell 2000 Index	https://www.cnbc.com/quotes/.RUT	Taiwan Stock Exchange Weighted Index	https://www.bloomberg.com/quote/TWSE:IND
Shanghai Stock Exchange (SSE) Composite Index	https://www.bloomberg.com/quote/SHCOMP:IND	FTSE MIB Index	https://finance.yahoo.com/quote/FTSEMIB.MI/
Kuala Lumpur Composite Index (KLCI)	https://www.bloomberg.com/quote/FBMKLCI:IND	Borsa Istanbul	https://borsaistanbul.com/en
LQ45 Index	https://finance.yahoo.	Nikkei 225	https://finance.yahoo.

(Continued)

Table 5. (continued)

Dataset/Market capitalization	Source	Dataset/Market capitalization	Source
	com/quote/%5EJKLQ45/		com/quote/%5EN225/
National Stock Exchange of India Ltd (NSE)	https://www.nseindia.com/	DAX	https://g.co/finance/DAX:INDEXDB
National Association of Securities Dealers Automated Quotations System (NASDAQ)	https://www.nasdaq.com/	KOSPI 200 Index	https://finance.yahoo.com/quote/%5EKS200/
New York Stock Exchange (NYSE)	https://www.nyse.com/index	Mercedes-Benz Group AG	https://finance.yahoo.com/quote/MBG.DE/
Commonwealth Bank of Australia (CBA.AX)	https://finance.yahoo.com/quote/CBA.AX/	iBOVESPA	https://finance.yahoo.com/quote/%5EBVSP/

Conclusion

SPF remains a topic of interest among investors, analysts, and researchers. This field is dedicated to predicting the future price of a stock, leveraging historical data and various influential factors. The study introduces a literature review and bibliometric analysis of papers on SPF, analyzing various methods, datasets, and metric evaluations, and summarizing results. The published papers were collected from the Scopus database, covering a range of methods, from ARIMA to DL approaches. We briefly provided the concepts and applicability of these approaches in SPF. The number of papers has increased from 2014 to 2023. We found a significant rise in the use of DL models and hybrid DL models from 2020 to 2023. This trend indicates that the research topic is attracting more attention from researchers.

These statistical models, such as ARIMA and its variations, are effective in linear conditions but are often affected by the nonlinear complexity of financial markets. The application of ML and DL in this field has led to the introduction of models that can represent complex patterns and nonlinear relationships in stock data. Based on this analysis, we realize that DL models such as LSTMs, convolutional LSTMs, transformers, and GANs are much more effective than traditional approaches, demonstrating promising performances. In recent years, the combination of various DL approaches has shown promising potential to enhance the performance of SPF. However, it might face challenges such as the need for large datasets for training, computational costs, and the risk of overfitting.

Future research in this domain should investigate various data types to improve prediction accuracy. This includes not only traditional financial data such as historical stock prices and volume but also technical indicators and non-traditional data like financial news, news sentiments, and social media sentiments. Moreover, exploring and optimizing novel hybrid models is necessary for enhancing the performance of forecasting systems.

Acknowledgements

This work was partly supported by Saigon University and Industrial University of Ho Chi Minh City.

Author Contributions

P.H.V contributed to the writing and data analysis; L.H.P contributed to the data collection; T.H.V.N and L.N.D contributed to the correction and data analysis; P.T.B contributed to the correction and supervision; T.D.T contributed to the writing, data analysis, correction and supervision.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Tan Dat Trinh https://orcid.org/0000-0003-0443-612X

References

1. Sonkavde G, Dharrao DS, Bongale AM, et al. Forecasting stock market prices using machine learning and deep learning models: a systematic review, performance analysis and discussion of implications. *Int J Finan Stud* 2023; 11: 94.
2. Petropoulos F, Apiletti D, Assimakopoulos V, et al. Forecasting: theory and practice. *Int J Forecast* 2022; 38: 705–871. DOI: https://doi.org/10.1016/j.ijforecast.2021.11.001
3. Obeidat S. Examining the random walk hypothesis in the Amman stock exchange: an analytical study. *Accounting* 2021; 7: 137–142.
4. Guo Y, Yao S, Cheng H, et al. China's copper futures market efficiency analysis: based on nonlinear Granger causality and multifractal methods. *Res Pol* 2020; 68: 101716. DOI: https://doi.org/10.1016/j.resourpol.2020.101716
5. Beyaz E, Tekiner F, Zeng X-j, et al. Comparing technical and fundamental indicators in stock price forecasting. In: 2018 *IEEE 20th International Conference on High Performance Computing and Communications; IEEE 16th International Conference on Smart City; IEEE 4th International Conference on Data Science and Systems (HPCC/SmartCity/DSS)* 2018, pp.1607-1613. IEEE.
6. Nti IK, Adekoya AF and Weyori BA. A systematic review of fundamental and technical analysis of stock market predictions. *Artif Intell Rev* 2020; 53: 3007–3057.
7. Ahmar AS, Singh PK, Ruliana R, et al. Comparison of ARIMA, SutteARIMA, and holt-winters, and NNAR models to predict food grain in India. *Forecasting* 2023; 5: 138–152.
8. Dudek G. STD: a seasonal-trend-dispersion decomposition of time series. *IEEE Trans Knowl Data Eng* 2023; 35: 10339–10350. DOI: 10.1109/TKDE.2023.3268125

9. Zhang GP. Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing* 2003; 50: 159–175.
10. Babu CN and Reddy BE. A moving-average filter based hybrid ARIMA–ANN model for forecasting time series data. *Appl Soft Comput* 2014; 23: 27–38.
11. Vijh M, Chandola D, Tikkiwal VA, et al. Stock closing price prediction using machine learning techniques. *Procedia Comput Sci* 2020; 167: 599–606.
12. Hoque KE and Aljamaan H. Impact of hyperparameter tuning on machine learning models in stock price forecasting. *IEEE Access* 2021; 9: 163815–163830.
13. Mintarya LN, Halim JN, Angie C, et al. Machine learning approaches in stock market prediction: a systematic literature review. *Procedia Comput Sci* 2023; 216: 96–102.
14. Li AW and Bastos GS. Stock market forecasting using deep learning and technical analysis: a systematic review. *IEEE Access* 2020; 8: 185232–185242.
15. Hu Z, Zhao Y and Khushi M. A survey of forex and stock price prediction using deep learning. *Appl Syst Innov* 2021; 4: 9.
16. Zuo Y and Kita E. Stock price forecast using Bayesian network. *Expert Syst Appl* 2012; 39: 6729–6737.
17. Adebiyi AA, Adewumi AO and Ayo CK. Comparison of ARIMA and artificial neural networks models for stock price prediction. *J Appl Math* 2014; 2014: 1–7.
18. Rezaei H, Faaljou H and Mansourfar G. Stock price prediction using deep learning and frequency decomposition. *Expert Syst Appl* 2021; 169: 114332.
19. Rouf N, Malik MB, Arif T, et al. Stock market prediction using machine learning techniques: a decade survey on methodologies, recent developments, and future directions. *Electronics (Basel)* 2021; 10: 2717.
20. Ren S, Wang X, Zhou X, et al. A novel hybrid model for stock price forecasting integrating encoder forest and informer. *Expert Syst Appl* 2023; 234: 121080.
21. Prasad VK, Savaliya D, Sanghavi S, et al. Stock Price Prediction for Market Forecasting Using Machine Learning Analysis. In: *International Conference on Computing, Communications, and Cyber-Security* 2022, pp.477-492. Springer.
22. Low PR and Sakk E. Comparison between autoregressive integrated moving average and long short term memory models for stock price prediction. *IAES Int J Artif Intell* 2023; 12: 1828–1835. Article. DOI: 10.11591/ijai.v12.i4.pp1828-1835.
23. Wahyudi ST. The ARIMA model for the Indonesia stock price. *Int J Econ Manage* 2017; 11: 223–236.
24. Pulungan DP, Wahyudi S, Suharnomo S, et al. Technical analysis testing in forecasting socially responsible investment index in Indonesia stock exchange. *Invest Manag Financ Innov* 2018; 15: 135–143.
25. Meher BK, Hawaldar IT, Spulbar CM, et al. Forecasting stock market prices using mixed ARIMA model: a case study of Indian pharmaceutical companies. *Invest Manag Financ Innov* 2021; 18: 42–54.
26. Kobiela D, Krefta D, Król W, et al. ARIMA Vs LSTM on NASDAQ stock exchange data. *Procedia Comput Sci* 2022; 207: 3836–3845.
27. Dong H. Stock Price Prediction Using ARIMA and LSTM. In: 2022 *6th Annual International Conference on Data Science and Business Analytics (ICDSBA)* 2022, pp.195-201. IEEE.
28. Subakkar A, Graceline Jasmine S, Jani Anbarasi L, et al. An Analysis on Tesla's Stock Price Forecasting Using ARIMA Model. In: *Proceedings of the International Conference on Cognitive and Intelligent Computing: ICCIC 2021, Volume 2* 2023, pp.83–89. Springer.
29. Suripto S. Decision-making model to predict auto-rejection: an implementation of ARIMA for accurate forecasting of stock price volatility during the COVID-19. *Decis Sci Lett* 2023; 12: 107–116.

30. Sigo MO. Big data analytics-application of artificial neural network in forecasting stock price trends in India. *Acad Account Financ Stud* 2018; 22: 1–13.
31. Selvamuthu D, Kumar V and Mishra A. Indian Stock market prediction using artificial neural networks on tick data. *Financ Innov* 2019; 5: 1–12.
32. Yunneng Q. A new stock price prediction model based on improved KNN. In: 2020 *7th International Conference on Information Science and Control Engineering (ICISCE)* 2020, pp.77–80. IEEE.
33. Lin G, Lin A and Cao J. Multidimensional KNN algorithm based on EEMD and complexity measures in financial time series forecasting. *Expert Syst Appl* 2021; 168: 114443.
34. Xiao J, Zhu X, Huang C, et al. A new approach for stock price analysis and prediction based on SSA and SVM. *Int J Inf Technol Decis Mak* 2019; 18: 287–310.
35. Ismail MS, Noorani MSM, Ismail M, et al. Predicting next day direction of stock price movement using machine learning methods with persistent homology: evidence from Kuala Lumpur stock exchange. *Appl Soft Comput* 2020; 93: 106422.
36. Zhang J, Li L and Chen W. Predicting stock price using two-stage machine learning techniques. *Comput Econ* 2021; 57: 1237–1261.
37. Tas E and Atli AH. A comparison of SVR and NARX in financial time series forecasting. *Int J Computat Econ Economet* 2022; 12: 303–320.
38. Tas E and Atli AH. Stock price ranking by learning pairwise preferences. *Comput Econ* 2022; 63: 513–528.
39. Khoa BT and Huynh TT. Forecasting stock price movement direction by machine learning algorithm. *Int J Electric Comput Eng* 2022; 12: 6625.
40. Bazrkar MJ and Hosseini S. Predict stock prices using supervised learning algorithms and particle swarm optimization algorithm. *Comput Econ* 2023; 62: 165–186.
41. Ballings M, Van den Poel D, Hespeels N, et al. Evaluating multiple classifiers for stock price direction prediction. *Expert Syst Appl* 2015; 42: 7046–7056.
42. Syukur A and Istiawan D. Prediction of LQ45 index in Indonesia stock exchange: A comparative study of machine learning techniques. *Int J Intell Eng Syst* 2021; 14: 453–463.
43. Ampomah EK, Qin Z and Nyame G. Evaluation of tree-based ensemble machine learning models in predicting stock price direction of movement. *Information* 2020; 11: 332.
44. Sadorsky P. A random forests approach to predicting clean energy stock prices. *J Risk Financ Manag* 2021; 14: 48.
45. Abraham R, Samad ME, Bakhach AM, et al. Forecasting a stock trend using genetic algorithm and random forest. *J Risk Financ Manag* 2022; 15: 188.
46. Basak S, Kar S, Saha S, et al. Predicting the direction of stock market prices using tree-based classifiers. *North Am J Econ Finan* 2019; 47: 552–567.
47. Noh H, Jang H and Yang CW. Forecasting Korean stock returns with machine learning. *Asia-Pacific J Finan Stud* 2023; 52: 193–241.
48. Chandra R and He Y. Bayesian Neural networks for stock price forecasting before and during COVID-19 pandemic. *Plos One* 2021; 16: e0253217.
49. Malagrino LS, Roman NT and Monteiro AM. Forecasting stock market index daily direction: A Bayesian network approach. *Expert Syst Appl* 2018; 105: 11–22.
50. Gunduz H, Yaslan Y and Cataltepe Z. Intraday prediction of Borsa istanbul using convolutional neural networks and feature correlations. *Knowl Based Syst* 2017; 137: 138–148.
51. Wen M, Li P, Zhang L, et al. Stock market trend prediction using high-order information of time series. *IEEE Access* 2019; 7: 28299–28308.
52. Sezer OB and Ozbayoglu AM. Algorithmic financial trading with deep convolutional neural networks: time series to image conversion approach. *Appl Soft Comput* 2018; 70: 525–538.
53. Hoseinzade E and Haratizadeh S. CNNpred: CNN-based stock market prediction using a diverse set of variables. *Expert Syst Appl* 2019; 129: 273–285.

54. Khanna V, Joshua J and Pramila R. Implementation of Supervised Pre-Training Methods for Univariate Time Series Forecasting. In: 2023 *2nd International Conference for Innovation in Technology (INOCON)* 2023, pp.1–8. IEEE.
55. Ghosh P, Neufeld A and Sahoo JK. Forecasting directional movements of stock prices for intraday trading using LSTM and random forests. *Finan Res Lett* 2022; 46: 102280.
56. Gao Y, Wang R and Zhou E. Stock prediction based on optimized LSTM and GRU models. *Sci Program* 2021; 2021: 1–8.
57. Fathali Z, Kodia Z and Ben Said L. Stock market prediction of Nifty 50 index applying machine learning techniques. *Appl Artif Intell* 2022; 36: 2111134.
58. Satria D. Predicting banking stock prices using RNN, LSTM, and GRU approach. *Appl Comput Sci* 2023; 19: 82–94.
59. Khan R and Baloch S. Forecasting the Stability of COVID-19 Vaccine Companies Stock Market using LSTM and Time-series Models. In: 2023 *International Conference on Communication, Computing and Digital Systems (C-CODE)* 2023, pp.1–6. IEEE.
60. Singh B, Henge SK, Mandal SK, et al. Auto-regressive integrated moving average threshold influence techniques for stock data analysis. *Int J Adv Comp Sci Appl* 2023; 14: 446–455.
61. Liu B, Yu Z, Wang Q, et al. Prediction of SSE Shanghai enterprises index based on bidirectional LSTM model of air pollutants. *Expert Syst Appl* 2022; 204: 117600.
62. Xu Y, Chhim L, Zheng B, et al. Stacked deep learning structure with bidirectional long-short term memory for stock market prediction. *Commun Comp Inf Sci* 2020; 1265: 447–460.
63. Mootha S, Sridhar S, Seetharaman R, et al. Stock price prediction using bi-directional LSTM based sequence to sequence modeling and multitask learning. In: 2020 *11th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)* 2020, 0078-0086. IEEE.
64. Liu M, Sheng H, Zhang N, et al. A New Deep Network Model for Stock Price Prediction. In: *International Conference on Machine Learning for Cyber Security* 2022, pp.413-426. Springer.
65. Nayak RK, Mishra D and Rath AK. A naïve SVM-KNN based stock market trend reversal analysis for Indian benchmark indices. *Appl Soft Comput* 2015; 35: 670–680.
66. Siddique M and Panda D. Prediction of stock index of tata steel using hybrid machine learning based optimization techniques. *Int J Rec Technol Eng* 2019; 8: 3186–3193. Article. DOI: 10.35940/ijrte.B3223.078219.
67. Tsai M-C, Cheng C-H, Tsai M-I, et al. Forecasting leading industry stock prices based on a hybrid time-series forecast model. *PloS One* 2018; 13: e0209922.
68. Manjunath C, Marimuthu B and Ghosh B. Analysis of Nifty 50 index stock market trends using hybrid machine learning model in quantum finance. *Int J Electr Comput Eng (IJECE)* 2023; 13: 3549–3560.
69. Chen W, Zhang H, Mehlawat MK, et al. Mean–variance portfolio optimization using machine learning-based stock price prediction. *Appl Soft Comput* 2021; 100: 106943.
70. Wang Y and Guo Y. Forecasting method of stock market volatility in time series data based on mixed model of ARIMA and XGBoost. *China Commun* 2020; 17: 205–221.
71. Chandar SK. Convolutional neural network for stock trading using technical indicators. *Autom Softw Eng* 2022; 29: 1–14.
72. Korade NB and Zuber M. Stock price forecasting using convolutional neural networks and optimization techniques. *Int J Adv Comp Sci Appl* 2022; 13: 378–385.
73. Kim HY and Won CH. Forecasting the volatility of stock price index: a hybrid model integrating LSTM with multiple GARCH-type models. *Expert Syst Appl* 2018; 103: 25–37.
74. Lv P, Wu Q, Xu J, et al. Stock index prediction based on time series decomposition and hybrid model. *Entropy* 2022; 24: 146.

75. Vuong PH, Dat TT, Mai TK, et al. Stock-price forecasting based on XGBoost and LSTM. *Comp Syst Sci Eng* 2022; 40: 237–246.
76. Lu W, Li J, Li Y, et al. A CNN-LSTM-based model to forecast stock prices. *Complexity* 2020; 2020: 1–10.
77. Wu JM-T, Li Z, Herencsar N, et al. A graph-based CNN-LSTM stock price prediction algorithm with leading indicators. *Multimedia Syst* 2021; 29: 1751–1770.
78. Wang H, Wang J, Cao L, et al. A stock closing price prediction model based on CNN-BiSLSTM. *Complexity* 2021; 2021: 1–12.
79. Kanwal A, Lau MF, Ng SP, et al. BiCuDNNLSTM-1dCNN—A hybrid deep learning-based predictive model for stock price prediction. *Expert Syst Appl* 2022; 202: 117123.
80. Aldhyani TH and Alzahrani A. Framework for predicting and modeling stock market prices based on deep learning algorithms. *Electronics (Basel)* 2022; 11: 3149.
81. Lu W, Li J, Wang J, et al. A CNN-BiLSTM-AM method for stock price prediction. *Neural Comput Appl* 2021; 33: 4741–4753.
82. Chen Y, Fang R, Liang T, et al. Stock price forecast based on CNN-BiLSTM-ECA model. *Sci Program* 2021; 2021: 1–20.
83. Wei X, Lei B, Ouyang H, et al. Stock index prices prediction via temporal pattern attention and long-short-term memory. *Adv Multim* 2020; 2020: 1–7.
84. Xu F and Tan S. Deep learning with multiple scale attention and direction regularization for asset price prediction. *Expert Syst Appl* 2021; 186: 115796.
85. Wang S. A stock price prediction method based on BiLSTM and improved transformer. *IEEE Access* 2023; 11: 104211–104223.
86. Li C and Qian G. Stock price prediction using a frequency decomposition based GRU transformer neural network. *Appl Sci* 2022; 13: 222.
87. Wu J-L, Tang X-R and Hsu C-H. A prediction model of stock market trading actions using generative adversarial network and piecewise linear representation approaches. *Soft Comput* 2023; 27: 8209–8222.
88. Staffini A. Stock price forecasting by a deep convolutional generative adversarial network. *Front Artif Intell* 2022; 5: 837596.
89. Wang T, Guo J, Shan Y, et al. A knowledge graph–GCN–community detection integrated model for large-scale stock price prediction. *Appl Soft Comput* 2023; 145: 110595.
90. Fuping Z. Conceptual-temporal graph convolutional neural network model for stock price movement prediction and application. *Soft COmput* 2023; 27: 6329–6344.
91. Smith S and O'Hare A. Comparing traditional news and social media with stock price movements; which comes first, the news or the price change? *J Big Data* 2022; 9: 1–20.
92. Liu Q, Lee W-S, Huang M, et al. Synergy between stock prices and investor sentiment in social media. *Borsa Istanbul Rev* 2023; 23: 76–92.
93. Bouadjenek MR, Sanner S and Wu G. A user-centric analysis of social media for stock market prediction. *ACM Trans Web* 2023; 17: 1–22.
94. Ashtiani MN and Raahemi B. News-based intelligent prediction of financial markets using text mining and machine learning: a systematic literature review. *Expert Syst Appl* 2023; 217: 119509.
95. Haryono AT, Sarno R and Sungkono KR. Transformer-gated recurrent unit method for predicting stock price based on news sentiments and technical indicators. *IEEE Access* 2023; 11: 77132–77146.
96. Ma D, Yuan D, Huang M, et al. VGC-GAN: A multi-graph convolution adversarial network for stock price prediction. *Expert Syst Appl* 2024; 236: 121204.
97. Zou J, Lou J, Wang B, et al. A novel deep reinforcement learning based automated stock trading system using cascaded lstm networks. *Expert Syst Appl* 2024; 242: 122801.

98. Sun Q, Wei X and Yang X. GraphSAGE with deep reinforcement learning for financial portfolio optimization. *Expert Syst Appl* 2024; 238: 122027.
99. Ma S and Fildes R. Large-scale time series forecasting with meta-learning. In: *Forecasting with artificial intelligence: Theory and applications*. Cham: Springer Nature Switzerland, 2023, pp.221–250.
100. Hong A, Gao M, Gao Q, et al. Non-stationary financial time series forecasting based on meta-learning. *Electron Lett* 2023; 59: e12681.
101. Gajamannage K, Park Y and Jayathilake DI. Real-time forecasting of time series in financial markets using sequentially trained dual-LSTMs. *Expert Syst Appl* 2023; 223: 119879.
102. Melgar-García L, Gutiérrez-Avilés D, Rubio-Escudero C, et al. A novel distributed forecasting method based on information fusion and incremental learning for streaming time series. *Inf Fusion* 2023; 95: 163–173.
103. Qiu Y, Liu R and Lee RS. The design and implementation of a deep reinforcement learning and quantum finance theory-inspired portfolio investment management system. *Expert Syst Appl* 2024; 238: 122243.
104. Cao Y, Zhou X, Fei X, et al. Linear-layer-enhanced quantum long short-term memory for carbon price forecasting. *QuantMach Intell* 2023; 5: 26.
105. Srivastava N, Belekar G and Shahakar N. The Potential of Quantum Techniques for Stock Price Prediction. In: 2023 *IEEE International Conference on Recent Advances in Systems Science and Engineering (RASSE)* 2023, pp.1–7. IEEE.
106. How M-L and Cheah S-M. Business renaissance: opportunities and challenges at the Dawn of the quantum computing era. *Businesses* 2023; 3: 585–605.

Author biographies

Pham Hoang Vuong received a BSc degree in Mathematics and Computer Sciences from University of Education - National University of HCM City, Vietnam in 1992. He also received Engineering degree from Hanoi University of Science and Technology in 2000 and received master's degree from Ho Chi Minh City University of Science in 2001. He is currently a lecturer at Computer Science Department, Sai Gon University, Vietnam since 2008. His research areas of interest include stock-price forecasting and computer vision based on deep learning.

Lam Hung Phu received a bachelor's degree in information technology from Sai Gon University of HCM City, Vietnam in 2023. His research areas of interest include computer vision and data analysis.

Tran Hong Van Nguyen received a BSc degree in Finance and Control from Saxion University of Applied Sciences, Holland, and a master's degree in international management from University of Rennes 1, France in 2012 and 2014 respectively. Currently, she is a PhD student in Industrial Engineering and Engineering Management and an assistant researcher Decision Analysis Lab at National Tsing Hua University, Taiwan. Since 2016, she has joined as a lecturer at the Faculty of Finance and Banking, Ton Duc Thang University, Vietnam. Her current research interests include smart manufacturing, operation research, financial management, and technology management.

Le Nhat Duy received PhD degree in Computer Science from Moscow State Pedagogical University, Russia in 2013. He is currently a lecturer at Computer Science Department, Industrial University of Ho Chi Minh City, Vietnam since 2013. His research areas of interest include computational intelligence and cryptography.

Pham The Bao received his BSc degree in Algebra from University of Natural Science - National University of HCM City, Vietnam in 1995. He also received MSc degree in Mathematical Foundation of Computer Science and PhD degree in Computer Science from University of Natural Science - National University of HCM City, Vietnam in 2000 and 2009, respectively. He was a lecturer and professor in Department of Computer Science, Faculty of Mathematics Computer Science, University of Natural Science, Vietnam from 1995 to 2018. He is currently dean and professor at Computer Science Department, Sai Gon University, Vietnam since 2019. His research includes image processing, pattern recognition and intelligent computing.

Tan Dat Trinh received a BSc degree in Mathematics and Computer Sciences from University of Natural Science - National University of HCM City, Vietnam in 2010. He also received Master of Engineering and PhD degree in Electronics and Computer Engineering from Chonnam National University, Korea in 2013 and 2017, respectively. He is currently a lecturer at Computer Science Department, Sai Gon University, Vietnam since 2019. His research areas of interest include speech signal processing, computer vision and pattern recognition.