Data Science in Finance and Economics

Review

Research on stock price prediction from a data fusion perspective

Aihua Li[1],*,†, Qinyan Wei[1],†, Yong Shi[2,3] and Zhidong Liu[1]

[1] School of Management Science and Engineering, Central University of Finance and Economics, Beijing, 102206, China

[2] College of Information Science and Technology, University of Nebraska at Omaha, Omaha, NE 68182, USA

[3] Key Laboratory of Big Data Mining and Knowledge Management, Chinese Academy of Sciences, Haidian District, Beijing, 100190, China

* **Correspondence:** Email: aihuali@cufe.edu.cn; Tel:86-10-62288622; Fax: 86-10-62288622.

† First author and second author contributed equally to this work

Abstract: Due to external factors such as political influences, specific events and sentiment information, stock prices exhibit randomness, high volatility and non-linear characteristics, making accurate predictions of future stock prices based solely on historical stock price data difficult. Consequently, data fusion methods have been increasingly applied to stock price prediction to extract comprehensive stock-related information by integrating multi-source heterogeneous stock data and fusing multiple decision results. Although data fusion plays a crucial role in stock price prediction, its application in this field lacks comprehensive and systematic summaries. Therefore, this paper explores the theoretical models used in each level of data fusion (data-level, feature-level and decision-level fusion) to review the development of stock price prediction from a data fusion perspective and provide an overall view. The research indicates that data fusion methods have been widely and effectively used in the field of stock price prediction. Additionally, future directions are proposed. For better performance of data fusion in the field of stock price prediction, future work can broaden the scope of stock-related data types used and explore new algorithms such as natural language processing (NLP) and generative adversarial networks (GAN) for text information processing.

JEL Codes: G10, C53

1. Introduction

The stock market, regarded as the economy's barometer, has been a key research area of interest for academics and industry. On the one hand, the stock market creates a favorable financing environment for companies. On the other hand, investors could earn potential returns from the stock market by executing investment decisions such as capital allocation, stock selection and timing (Keller and Siegrist, 2006). Correspondingly, investors have to be exposed to the investment risks associated with the investment behavior. As a result, stock price prediction is significant for the investment decision of investors to balance their return and risk of investment. Stock prices are highly volatile, non-linear and stochastic due to external factors such as political events and corporate development (Nti et al., 2021), making it difficult to achieve accurate stock price prediction by relying on the previous stock price data. Thus, it is critical for stock price prediction to capture both the previous stock price and other factors (Zhang et al., 2022b).

In stock price prediction research, the raw data can be divided into quantitative and qualitative data, which are characterized by multiple sources of heterogeneity, uncertainty and large scale (Zhang and Zhang, 2022). Before using these data for analysis, certain methods need to be used to process, integrate and utilize them to form a data warehouse for subsequent prediction studies (Evans et al., 2018).

Existing studies have used single algorithms to mine the stock price-related information and make single predictions to make the investment decision (Ariyo et al., 2014; Brogaard and Zareei, 2022). However, the generalization ability of investment decisions derived from a single model is weak and unreliable (Zhang et al., 2018). Therefore, the introduction of multiple models with different structures or the same structure but different initializations for stock price prediction and the integration of multiple predictions as the basis for the final investment decision are significant for improving the interpretability and reliability of stock price prediction models (Gandhmal and Kumar, 2019). Given the aforementioned difficulties, the data fusion (DF) method has been applied to stock price prediction (Thakkar and Chaudhari, 2021). Data fusion refers to the process of integrating the data, features and knowledge with multiple sources and/or different structures and modalities to obtain fused data, fused features and fused knowledge for the further analysis (Li et al., 2022a, 2022b). Previous studies have shown that fusing heterogeneous data from multiple sources and using multiple models for stock price prediction are more accurate and have better generalization performance compared to using single-source data and a single model (Thakkar and Chaudhari, 2021).

Although data fusion has been emphasized in stock price prediction, previous studies only focused on one aspect of the stock price prediction using data fusion. Furthermore, there is a lack of a complete and systematic summary of studies on the application of data fusion in stock price prediction. Thus, this study reviews the relevant research on data fusion in stock price prediction from three main levels of data fusion (data-level, feature-level and decision-level fusion), presenting a complete understanding of the research problem (Guo et al.,2014; Lee et al., 2023; Zhang et al., 2022). It is worth noting that only all types of stock price prediction problems are addressed, and no distinction is made between timeframes.

The article unfolds as follows. Section 2 briefly introduces the stock price prediction methods applied in the existing studies. Section 3 provides a statistical overview of the literature on existing

studies. Sections 4, 5 and 6 provide an overview on stock price prediction using data fusion at the data, feature and decision levels, respectively. Finally, this study concludes and highlights the directions for future studies.

2. Related studies

In recent years, extensive traditional time series-based prediction methods have been widely used in stock market prediction (Figure 1). The autoregressive (AR), moving average (MA) and autoregressive moving average (ARMA) models are basic models for stock price time-series prediction. It should be noted that the stock prices are supposed to be pre-processed and reach a stationary state (Shi et al., 2012). The autoregressive integrated moving average (ARIMA) model, developed based on the ARMA model, is one of the most representative statistical models for time-series data analysis and can be used to fit models to non-stationary stock price time series data (Ariyo et al., 2014). However, the above four models are only suitable for predicting stationary time-series data under certain conditions, and their variances are supposed to be constant, while the stock market is non-stationary. Autoregressive conditional heteroskedasticity (ARCH) and generalized autoregressive conditional heteroskedasticity (GARCH) could address the problems caused by the traditional econometric second assumption (constant variance) for time-series variables and are more widely used in stock price prediction (Jeantheau, 2004; Liu et al., 2021b). Traditional stock price prediction methods mainly use the linear regression to predict stock trends based on stationary linear historical data. However, stock prices are non-stationary, influenced not only by historical stock trading data but also by non-linear factors such as political factors, investment sentiment and specific events.

Machine learning algorithms with excellent non-linear regression performance are becoming popular methods for stock market prediction, including logistic regression, support vector machine (SVM), decision tree and ensemble learning (Brogaard and Zareei, 2022; Cheng et al., 2021; Xiao et al., 2019; Yang et al., 2022). Recently, machine learning algorithms have also been gradually applied to quantitative stock trading, and research results have shown that machine learning-based quantitative stock trading strategies are better than traditional simple trading strategies (Wang and Yan, 2023; Yan et al., 2023). In addition, deep learning algorithms, such as deep neural network (DNN), convolution neural network (CNN) and recurrent neural network (RNN) can extract potential features of highly unstructured data and explore complex intrinsic patterns of stock price movements based on time series data, and they have been used to predict stock market trends (Hu et al., 2021; Liu et al., 2021a; Lu and Lu, 2021). Among all RNN-based models, long short-term memory (LSTM) could be the most effective model for time series prediction. LSTM uses a set of memory cells with gate structure to replace hidden neurons of RNN. As such, through the gate structure feature, the information is retained and persistently updated in the following training iterations. So, it has the advantage of solving the gradient explosion and gradient disappearance problems in neural network algorithms. Several articles have shown that it outperforms RNNs and traditional machine learning algorithms in stock predictions based on time series data (Liu et al., 2021; Brogaard and Zareei, 2022).

Although machine learning and deep learning have the advantage of non-linear prediction performance in stock price prediction, most of the studies only utilize a single historical stock data and a single prediction model to realize stock price prediction, which has limitations in terms of interpretability and generalization performance (Zhang et al., 2018). Extending stock-related data

sources, fusing multiple sources of heterogeneous stock data and fusing multi-algorithm prediction results thus become the key to improving prediction performance.

In subsequent studies, external factors such as macroeconomic indicators, financial network news and specific events have been gradually combined with stock price data and incorporated into stock price prediction models (Lee et al., 2023). Additionally, among the factors affecting stock prices, market and investor sentiment offers important information. Some studies have emphasized the explanation of sentiment information of stock prices, and the results showed a positive correlation between sentiment information and stock market trading volume (Long et al., 2020; Shields et al., 2021; Zhang et al., 2017a). With the development of graph mining techniques, graph data such as candlestick charts have also been considered as one of the information sources for stock price prediction models (Kim and Kim, 2019; Liu et al., 2022; Wang et al., 2019).
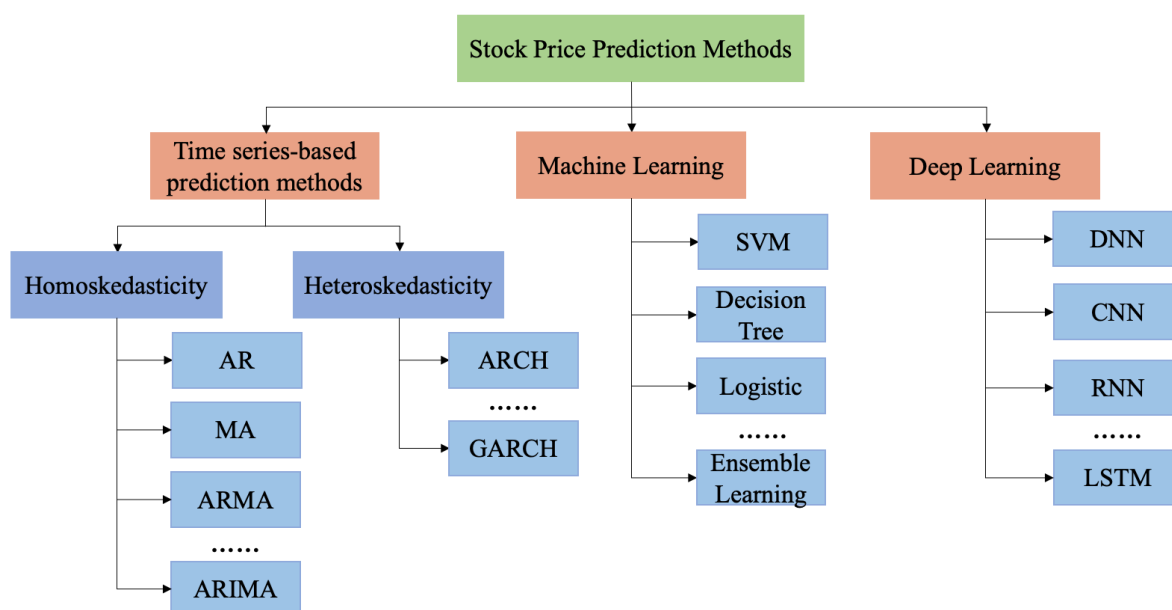

Figure 1. Stock price prediction methods.

The stock market has a huge variety of data sources, including the internet, databases, emails, social networking sites, news media, etc. These sources generate vast amounts of stock-related data daily, typically in terabytes or gigabytes (Nti et al., 2021). However, the ubiquity of such data and the impact of factors such as public sentiment and economic indicators on stock prices make the integration and utilization of data from multiple sources challenging in stock price prediction. Within this context, data fusion is emerging as a critical area of research in stock price prediction. The primary objective is to combine data from various sources into a new dataset or feature set that provides comprehensive knowledge of the factors influencing stock price movement. The resulting dataset is then used as input for prediction models to generate more accurate predictions and facilitate better investment decisions. Previous research has shown that the application of data fusion methods in predicting stock prices results in models with higher accuracy and better generalization performance (Stoean et al., 2019; Thakkar and Chaudhari, 2021; Zhou et al., 2022). Thus, data fusion has become a crucial subject of exploration for researchers, with the potential to provide investors with more reliable insights to inform their investment decisions.

3. Overviews

To provide a clear overview of data fusion development in stock price prediction, we extensively searched the SCI literature database within the Web of Science platform. We used advanced search parameters, including the first-level subject term "Data Fusion" and the second-level subject term "Stock". The search was limited to articles published between 2003 and 2023 and was subjected to the "Web of Science Core Collection" inclusion criteria. Our search returned a total of 379 records.

Using the search results, we created Figure 2 to display the trend in the number of articles published over time. As observed from the graph, the number of relevant studies generally increased over time and significantly spiked between 2015 and 2022.
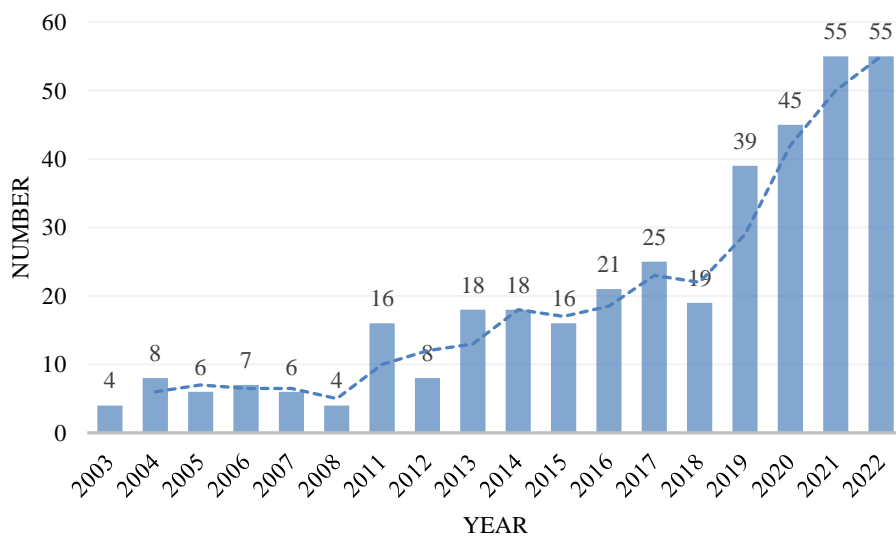

Figure 2. The trend in the number of articles published over time.
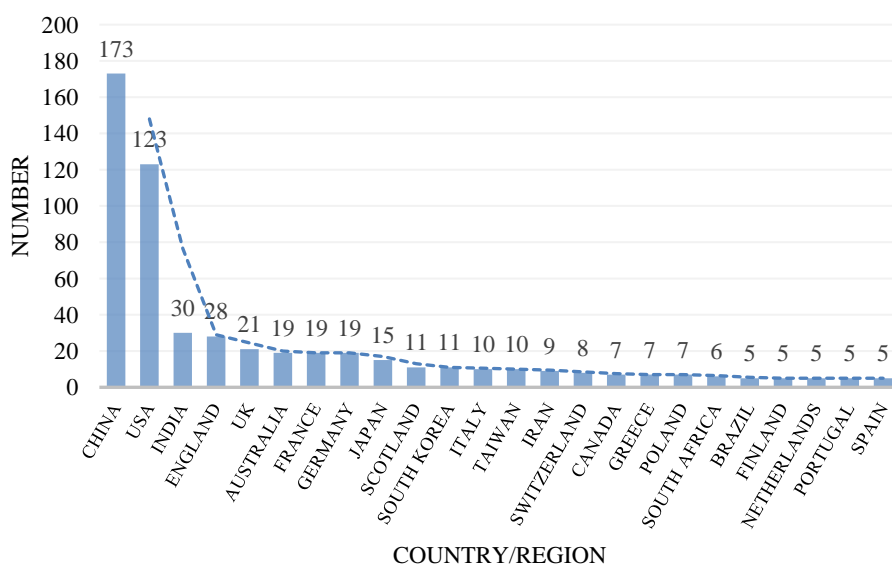

Figure 3. Distribution of relevant literature by country/region.

Research interest in applying data fusion to stock price prediction varies globally. Figure 3 displays the distribution of relevant literature by country/region using the results of our literature search, indicating that China and the USA have the most relevant studies.

We further counted the research directions of the screened literature to construct the top ten research directions based on the number of publications, as displayed in Figure 4. As observed in the figure, most of the relevant literature is published in the fields of computer science, mathematics, engineering and business economics.

Figure 4. Distribution of relevant literature by research directions.

The above literature statistical visualizations demonstrate that the number of studies on stock price prediction through data fusion is increasing, indicating potential future research implications in this field.

Additionally, analysis of the existing literature reveals that research on stock price prediction with data fusion is primarily conducted at three levels: data-level, feature-level and decision-level fusion.

4. Stock price prediction based on data-level fusion

Figure 5 illustrates that stock-related data can be categorized into two main types: qualitative and quantitative data. Quantitative data mainly comprises numerical information such as historical stock data, corporate financial data, macroeconomic data and other stock-related index data. Traditional stock price prediction models primarily utilize historical stock data to predict trends, as described in Section 2 of this paper. However, the efficient market hypothesis (EMH) and the stochastic wandering hypothesis (SWH) suggest that using historical stock data alone may not predict future stock price trends effectively (Malkiel and Fama, 1970; Malkiel, 2015). Recent studies, such as Stoean et al. (2019), used LSTM-based prediction models to predict the closing price of 25 stocks in the Bucharest Stock Exchange using only historical stock prices and achieved significant results. They also suggested that combining multi-indicator data can further improve the predictive power of their models.
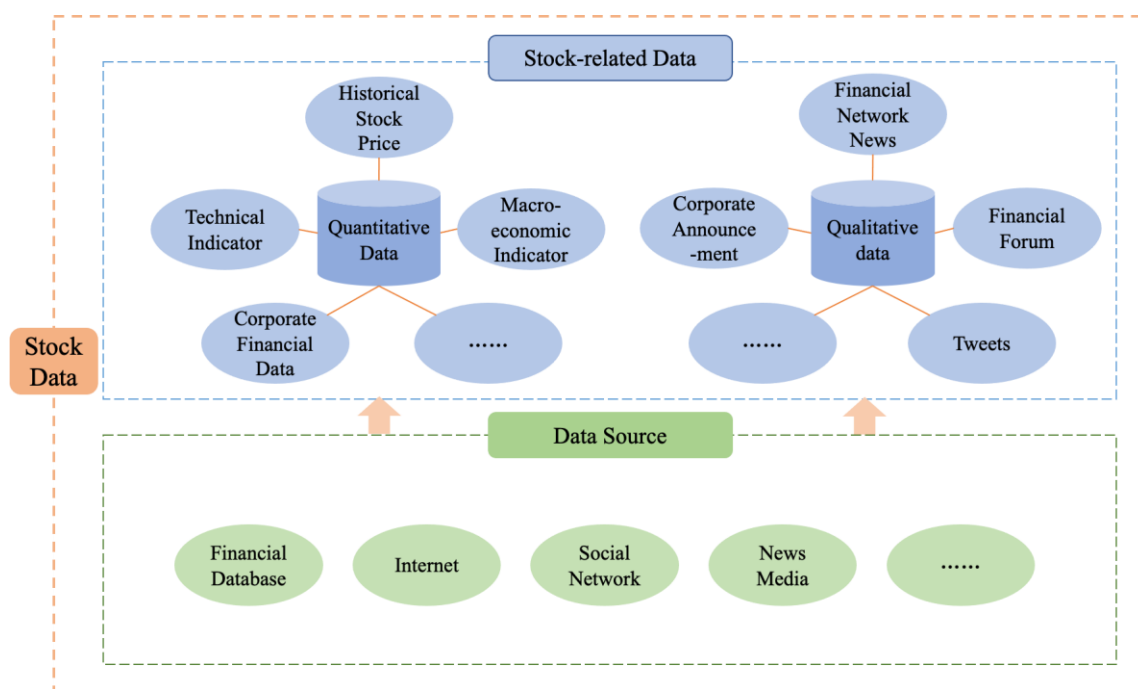
Stock-related Data

Stock Data

Historical Stock Price

Technical Indicator

Quantitative Data

Macro-economic Indicator

Corporate Financial Data

......

Financial Network News

Corporate Announce-ment

Qualitative data

Financial Forum

......

Tweets

Data Source

Financial Database

Internet

Social Network

News Media

......

Figure 5. Overview of stock price data.

Apart from quantitative data, qualitative factors that impact stock price are also gaining attention in recent research, as indicated in (García-Medina et al., 2018). One such factor is event-specific information from online media or Web news which has high correlations with stock price, with studies showing that various events, including financial network news and firm-specific announcements, can impact stock price trends (Shi et al., 2022; Thakkar and Chaudhari, 2021). However, information on stock price-related events collected from the Web is very sparse. Although Web news is incrementally available, events are usually presented as unstructured text, so the number of events that can be extracted from Web news remains limited. In addition, the same event may be described differently on different websites in different ways and thus be treated as different events, leading to event sparsity. So, using event information solely for stock price prediction is not enough (Zhang et al., 2017). Additionally, from behavioral finance theory, emotions also play a significant role in decision making, and investors' own investment emotions may influence their investment decisions (Zhou et al., 2018). For example, the collective level of optimism or pessimism in society can affect investor decisions (Nofsinger, 2005). Due to the recent advances in natural language processing (NLP) techniques, sentiment-driven stock prediction techniques have also been proposed by extracting indicators of public mood from social media, where positive mood for a stock will probably indicate a rising trend in the price, and negative mood will more likely mean a decreasing trend (Bollen et al., 2011). Therefore, sentiment information from social media, such as investor sentiment information in financial forum discussions and tweets, can also be used as a complement to quantitative data.

Several studies have added sentiment information to their prediction models, with improved predictive power as a result (Li et al., 2020). For example, Chiong et al. (2018) developed a sentiment analysis method based on financial news disclosure, extracting sentiment-related features as input for the stock price prediction model. Compared with the prediction models without the inclusion of sentiment-related features, their proposed SVM and particle swarm optimization (PSO)-based model

with sentiment feature extraction performed well in terms of accuracy and time. Their results showed a positive correlation between public sentiment and future stock prices. However, relying on the sentiments alone is not sufficient for prediction either. For example, during holidays, people's mood tends to be positive, yet it may not really reflect their investment opinions.

To tackle the above problems, data-level fusion and feature-level fusion are being introduced in stock price prediction research to achieve integration of data from different sources and provide more informative inputs for the subsequent prediction model training.

Data-level fusion and feature-level fusion can solve the challenge of fusing multi-source stock data, but they differ in their approaches. Multi-source stock data includes two types of data: multi-source homogeneous data and multi-source heterogeneous data. Data-level fusion can be applied to both types of data, while feature-level fusion is more suitable for heterogeneous data.

In data-level fusion, raw stock-related data from various sources are merged before being further processed. Homogeneous stock data fusion involves the straightforward merger of stock data with the same structure, such as the direct integration of different dimensions of stock trading data provided by Bloomberg and Wind (Zhang et al., 2017b). However, the small number of homogeneous stock data categories and the simple fusion process limit the interpretability of the final fused data on stock prices. Therefore, some scholars performed data-level fusion on heterogeneous data. For instance, Nti et al. (2021) proposed a novel multisource information-fusion stock price movement prediction framework based on a hybrid deep neural network architecture (CNN and LSTM) named IKN-ConvLSTM and fused stock-related information from six heterogeneous data sources using data preprocessing and record matching approaches. The empirical evaluation of their model was carried out with stock data (January 3, 2017, to January 31, 2020) from the Ghana Stock Exchange (GSE), and the results showed good prediction accuracy (98.31%), specificity (0.9975), sensitivity (0.8939%) and F-score (0.9672) of the amalgamated dataset compared with the distinct dataset. Considering the limitation of using only stock price data to predict stock prices, Lee et al. (2023) introduced two types of modal data, macroeconomic indicators and month/week, into a stock direction classification model based on historical stock price data and achieved data-level fusion by modally linking the three types of data. The test results showed better performance for the fusion model compared with the comparison models and achieved statistically significant results. Specifically, 27 out of 50 stocks achieved higher classification accuracy than the comparative model. However, the manual merging required to integrate heterogeneous data can be time-consuming and labor-intensive.

Reviewing the data-level fusion for multi-source stock data, it can be found that data fusion in this level has certain shortcomings in terms of efficiency, effectiveness and stock price interpretability. More research is therefore distributed in the areas of feature-level fusion and decision-level fusion.

5. Stock price prediction based on feature-level fusion

While the raw data contains rich information, it also has a significant amount of noise. Direct data-level fusion may not yield effective information. Therefore, feature extraction, feature selection and feature fusion of the raw data are required before training prediction models with the data. In feature-level fusion, the input data can be heterogeneous, and it is usual to extract statistical features or signals from the heterogeneous raw data and select or connect these features before further analysis.

Regarding heterogeneous stock-related data, historical stock data provides fundamental stock trading information, and technical indicators derived from it can serve as features for learning specific

interpretations. Specific event attributes or market behavior based on a particular context can also be considered for potential features since stock markets are affected by numerous specific events. Moreover, investor or market sentiment can significantly impact stock market volatility, making sentiment features obtained using sentiment analysis another useful input for stock price prediction models. While individual features provide valuable information about the stock market, combining different aspects of these features through feature fusion is crucial to extract the intrinsic characteristics of the stock market. Feature fusion is performed in a manner that can be viewed as feature-level abstraction or object refinement of the processed data (Chiong et al., 2018). The fused features can be applied to all levels of models for subsequent stock market analysis.

Through analysis, it has been discovered that obtaining information related to stock price fluctuations and considering the influencing features related to stock price comprehensively are crucial to improve the accuracy of prediction models. Generally speaking, there are four types of features that may affect stock prices: quantitative features (stock price features, macroeconomic features, financial status features), sentiment features (financial forum features, tweet features, social media features), specific event features (network news features, corporate announcement features) and chart features (candlestick chart features). Different feature extraction and fusion methods are required for different features.

Regarding fusion paths, there are two types of paths: serial fusion and parallel fusion. The specific paths for feature-level fusion are shown in Figure 6. The former path has only one fusion step, which is "raw data - data merging - feature extraction - feature fusion - model training". The latter path has multiple parallel processing steps, which are "raw data - feature extraction - feature fusion - model training".
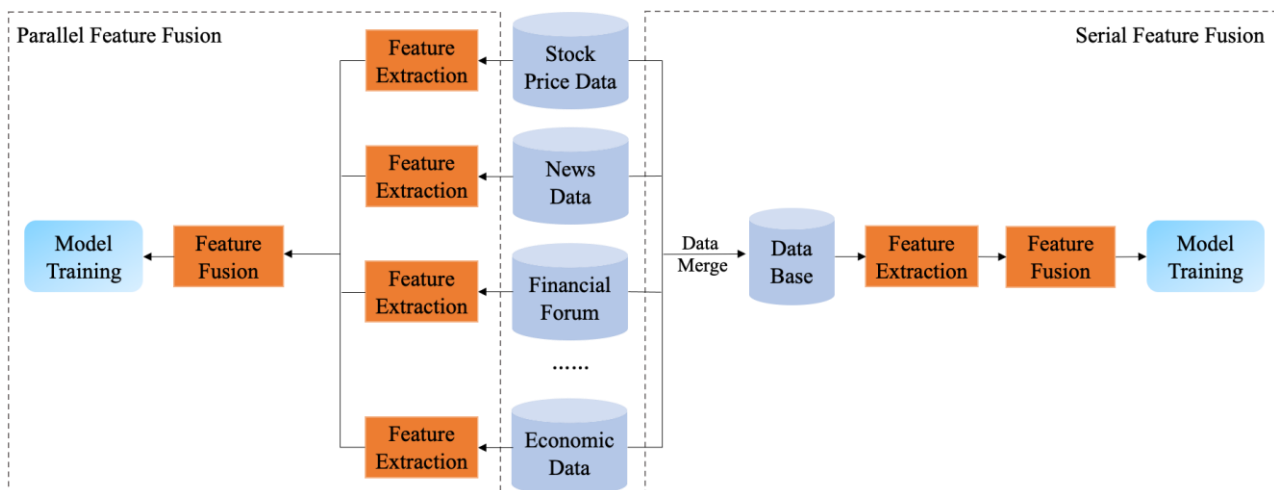

Figure 6. Serial/Parallel feature fusion paths.

5.1. Serial feature fusion

In the serial fusion path, feature selection and fusion are based on the merged dataset. Nti et al. (2021) combined six heterogeneous data sources containing quantitative features, sentiment features and specific event features into a data warehouse using record matching, followed by training a CNN and stacked LSTM model on the resulting data warehouse to achieve feature selection and fusion. Lee

et al. (2023) used modal linking to merge stock price data, macroeconomic indicators and date data in quantitative features before feature-level fusion to generate merged data and use them as input to the multi-headed attention layer for feature fusion. Due to differences in dimensionality and structure among the raw stock-related data, direct merging of raw data is time and operationally expensive. Moreover, serial fusion is prone to sparsity issues, so its application in stock price prediction problems has been limited.

5.2. Parallel feature fusion

Unlike serial feature fusion, the parallel fusion path employs various models to perform feature extraction and fusion for data with different categories of features, solving the dimensional inconsistency issue that results from fusing features with variable dimensions. The most fundamental feature in stock price prediction is the quantitative feature represented by historical stock data, containing rich stock-related information. Considering the non-stationary, stochastic and non-linear characteristics of stock price time-series data, some scholars utilize a combination of multiscale convolutional feature fusion (MCFF) and LSTM to achieve the fusion of different representation features. They employ the wavelet transform and normalization to denoise the financial time series data and achieve the fusion of eight features, including close, open, volume, etc. In this way, the stock price history data are represented as the form of an eight-dimensional time series and further used as the input for an LSTM-based model. This method successfully extracts and merges diverse features from stock price time-series data, thus improving the generalization performance of the prediction model (Zhang and Zhang, 2022). Furthermore, independent component analysis (ICA) can be adopted to extract stock price features and technical indicator features from historical prices, fusing both features using canonical correlation analysis (CCA) based feature fusion methods (Guo et al., 2014).

However, using only stock price features or technical indicator features for stock price prediction has its limitations; therefore, incorporating stock price related features such as event-specific features, sentiment features and graph features should be considered to establish a multi-class feature fusion model. Zhang et al. (2017b) explored the joint effect of event-specific features and investor sentiment on stock price prediction in their study. They accomplished the fusion of quantitative stock features, sentiment features and event-specific features through the framework of coupled matrix and tensor factorization and introduced a stock correlation matrix to tackle the data sparsity issue. The analytical frameworks of Sun et al. (2021) and Daradkeh (2022) were based on CNN and bidirectional long short term memory (Bi-LSTM) networks and using two algorithms to extract event features from text data and emotional polarity features based on backward and forward contextual information, respectively. They then implemented feature fusion through a fully connected network. Differently, Daradkeh further incorporated another quantitative feature, i.e., macroeconomic features, into their stock trend prediction framework. Both studies suggested that combining quantitative stock data with non-quantitative stock-related information from event-specific features, sentiment polarity, etc. can enhance stock price prediction performance. Moreover, the candlestick chart, strongly associated with stock price features, is considered to be the optimal chart feature for stock price prediction, and thus it was also incorporated in the studies of Kim and Kim (2019) and Liu et al. (2022). In the feature extraction and fusion framework of the candlestick chart, the most prominent algorithm used is CNN+Bi-LSTM, which enables extraction of graph features and transforms them into time-series

features for further fusion with stock price features and technical features as inputs to the final prediction model.

Although the aforementioned studies used feature-level fusion to unveil more intrinsic information about stocks and achieved robust prediction results, most of them used only a single prediction model such as SVM, LSTM or reinforcement learning (RL) to obtain a single prediction result as an investment basis for decision making (Daradkeh, 2022; Guo et al., 2014; Liu et al., 2022). However, the prediction results of a single prediction model are vulnerable to interference from different factors, resulting in unstable investment decisions. Consequently, it is indispensable to enhance the reliability of the prediction outcomes by considering the prediction results from multiple prediction models at the final decision level.

6. Stock price prediction based on decision-level fusion

Generally, if the prediction models have different structures or the same structure but with random initialization, they may generate distinct stock price prediction outcomes. Additionally, different prediction models vary in their capabilities to learn different types of data or features, which in turn affects the final prediction outcomes. In this case, better prediction can often be obtained by considering the results of multiple prediction models rather than relying on a single prediction outcome from a single model (Sun et al., 2021). The learning result of a model can be deemed the probability of belonging to a specific category under the influence of a certain feature, and the fusion of outcomes can improve the correlation between features to a certain extent (Lai et al., 2021). This strategy of aggregating "group intelligence" is also known as ensemble learning in the data mining field, which aims to balance the limitations of a single model by the strengths of each base learner. The ensemble learning embodies the idea of decision-level fusion, which means that the decision output from multiple base learners is combined into a single prediction result about the stock price to obtain a more stable investment decision than a single model. Hence, this combination of multiple predictions provided by multiple learners is also known as decision-level fusion (Ho et al., 1994; Tulyakov et al., 2008).

In terms of structural similarities and differences between the base learners, decision-level fusion can be classified into decision-level fusion of homogeneous base learners and decision-level fusion of heterogeneous base learners. The variance lies in the application of prediction models with distinct parameters of the same algorithm or different algorithms.

Prediction studies based on homogeneous algorithms utilize the same algorithm for training and predicting stock price related data. Among these algorithms that are listed in Table 1 in terms of popularity, the artificial neural network (ANN) is the most frequently used base learner for stock price prediction, followed by LSTM, decision tree and SVM. Precisely, ANN enables us to design useful nonlinear systems accepting large numbers of inputs, with the design based solely on instances of input-output relationships. However, ANNs are based on the empirical risk minimization principle, which may run the risk of model over-fitting and local minimums (Giacomel et al., 2015; Lahmiri, 2018). The LSTM model is one kind of recurrent neural network which can theoretically store an infinite amount of time information and avoids the negative influence of vanishing gradient and exploding gradient through the control of input gate, output gate and forget gate (Xie et al., 2018; Yang et al., 2020). Additionally, probabilistic neural network (PNN), extreme learning machine (ELM) and deep belief network (DBN) have also been applied as base learners in a few studies. In contrast,

heterogeneous algorithms for stock price prediction combine several different algorithms for model training and prediction, ensuring the diversity of model prediction.

Table 1. Homogeneous base learner applications.

Base Learner	Source	Algorithm
Homogeneous	Giacomel et al., 2015; Lahmiri, 2018;	ANN
	Lahmiri and Boukadoum, 2015; Nezhad and Bidgoli, 2019	
	Xie et al., 2018; Yang et al., 2020	LSTM
	Zhou et al., 2022	Decision Tree
	Qiu et al., 2017	SVM
	Chandrasekara et al., 2019	PNN
	Khuwaja et al., 2020	ELM
	Wang et al., 2018	DBN
	Zhong et al., 2017	Random Forest
	Lin et al., 2021	KNN
	Melin et al., 2012	ANFIS[1]

Table 2 summarizes the stock price prediction literature based on heterogeneous algorithms. Most of these studies use integration between different machine algorithms. For instance, Abraham and Auyeung (2009) investigated how the seemingly chaotic behavior of stock markets could be well-represented using an ensemble of intelligent paradigms which includes ANN, SVM, neuro-fuzzy system (NFS) and a difference boosting neural network (DBNN). Experimental results reveal that the ensemble techniques performed better than the individual methods. Similarly, some other scholars have used different combinations of algorithms for stock price prediction studies, such as Back Propagation Neural Network (BPNN)-RNN-SVR, CNN-LSTM and ANN-DecisionTree-KNN. Non-machine learning algorithms such as user knowledge, expert knowledge, speech and text encoder, and Delphi method is also used as part of ensemble algorithms for stock price prediction.

Table 2. Heterogeneous base learner applications.

Base Learner	Source	Algorithms
Heterogeneous	Abraham and Auyeung, 2009	ANN, DBNN, NFS, SVM
	Alhnaity and Abbod, 2020	BPNN, RNN, SVR
	Chong et al., 2020	CNN, CNN-LSTM, LSTM
	Qian and Rasheed, 2007	ANN, Decision Tree, KNN
	Barak et al., 2017	ANN, Decision Tree, Rule-Based Algorithm, SVM
	Dash et al., 2019	MLP[2], Random Forest, Naïve Bayes, SVM
	Lee and Kim, 1995	ID3, Expert Knowledge, User Knowledge
	Sawhney et al., 2020	MTL[3], Speech and Text Encoder, SVM
	Kuo et al., 1996	ANN, Delphi Method
	Kristjanpoller and Michell, 2018	ANFIS, GARCH

[1] Adaptive Neuro Fuzzy Inference System
[2] Multi-Layer Perceptron
[3] Multi-Task Learning

After summarizing the characteristics of the base learners in the study, we analyze the model structure of the existing literature with reference to Zhang's (2022a) study and can conclude that the decision fusion model design can be divided into three categories, which are traditional decision-level fusion, auxiliary model-based decision-level fusion and two-stage decision-level fusion.

6.1. Traditional decision-level fusion

Traditional decision-level fusion involves multiple base learners (e.g., Figure 7), which fuse the prediction results of each base learner to provide a combined prediction of the corresponding analysis variables. Stock price prediction using stock time-series data is one of these cases. It decomposes the stock time-series features into multiple parts according to the slide window and trains the time-series features of each part independently to provide partial predictions of the corresponding analysis variables. The partial predictions are then fused. Carta et al. (2021) proposed a multi-level and multi-integrated stock trading model based on deep learning and deep reinforcement learning to address the low performance of a single supervised classifier in predicting future market behavior. Their base learners were selected based on a deep double-Q network (DQN) that used raw stock price data to generate multiple stock trading signals through different iterations. Finally, given such outputs from the previous meta-learners, they proposed that the final agent works in an ensemble fashion, which considers majority voting of decisions to generate the final trading signal with respect to long, idle or short positions. Similarly, Liu et al. (2021a) segmented the stock time-series data into datasets based on a 20-day time interval. They proposed three-branch structure based on CNN and LSTM with different parameter settings to complete feature fusion and stock price prediction for each segmented dataset, respectively. The prediction results of the three-branch model were then fused in the LSTM's cell model as input to the dense layer used for denormalization to achieve the final decision fusion. Precisely, the final output of their designed model is the predicted value of the (t+1)-th day's closing stock price using the previous t days' feature data.

In contrast, the idea of Shi et al. (2022) used two different algorithms, multiple linear regression (MLR) and LightBoost, to train the stock price dataset separately. They then introduced the LSTM as the fusion algorithm to fuse the prediction results and formed the final prediction. Similar studies have also used a combination of logistic regression, random forest and XGBoost as the base learners for the first level of the basic prediction model, followed by a new XGBoost as the second level of the prediction model to achieve decision-level fusion (Zhang and Lu, 2021).
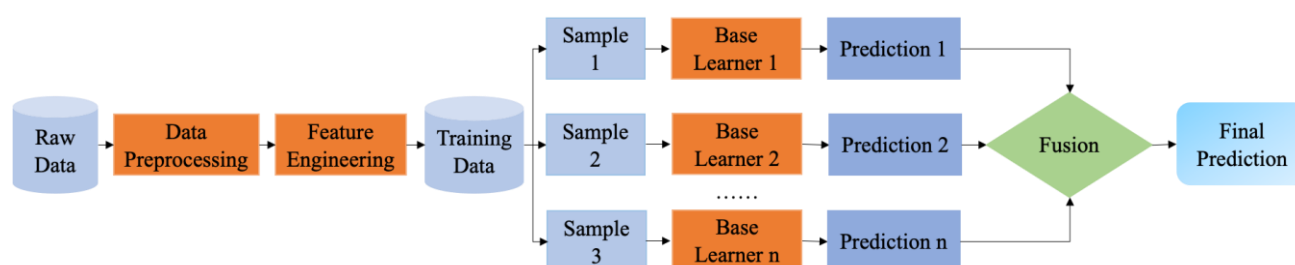

Figure 7. Traditional decision-level fusion.

6.2. Auxiliary model-based decision-level fusion

Auxiliary model-based decision-level fusion consists of two parallel prediction paths (e.g., Figure 8), the original prediction model based on stock data and the prediction model based on the effects of external factors. The latter usually predicts the effects of external factors using machine learning algorithms or non-machine learning algorithms, and the former fine-tunes the prediction results by combining the latter in the decision-level fusion to form the final fusion decision (Kuo et al., 1996; Kristjanpoller and Michell, 2018). Lai et al. (2021) considered the effect of news features on stock prices in their stock price prediction model. They built two models, the original stock price prediction model and the news feature prediction model, using two algorithms, Bi-GRU and TextCNN, to learn news text data and time-series data, respectively. The prediction results of the two models were then fused into a final single prediction by assigning different weights to different inputs well, so as to tell the discriminant network which inputs are more important using the weighted sum method and attention mechanism.
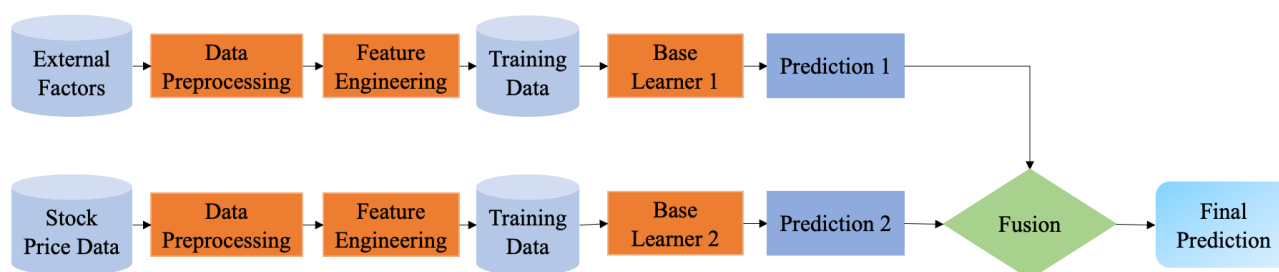

Figure 8. Auxiliary model-based decision-level fusion.

6.3. Two-stage decision-level fusion

In the two-stage decision-level fusion, the prediction in the second stage is based on the fusion prediction result of the first stage as input (e.g., Figure 9). To improve on the traditional method of using the technical indicators on day t as the input to the model to predict the closing price of stocks on day t+n, Zhang et al. (2021) proposed a two-stage ensemble learning algorithm based on a support vector regression – ensemble adaptive neuro fuzzy inference system (SVR-ENANFIS). The SVR was used in the first stage of the model to predict the day t+n values of several technical indicators such as moving average (MA) and exponential moving average (EMA), which were used as inputs for the second stage of the model. In the second stage, the ENANFIS ensemble algorithm, which contains 10 basic ANFIS models, was used to train and obtain predictions for the input data obtained in the previous stage. Finally, a weighted average method was used to fuse the 10 outputs to obtain the stock closing price prediction for day t+n. Comparison results indicated that this two-stage model performed better than the single-stage ENANFIS, two-stage SVR-linear and two-stage SVR-SVR models.
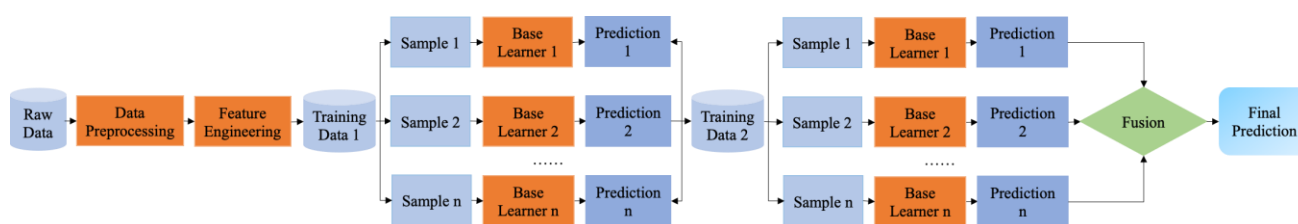
Figure 9. Two-stage decision-level fusion.

As seen from the summary of the three types of decision fusion model design, traditional decision fusion is a more applied decision fusion framework with certain advantages in terms of comprehensibility and operability. However, this framework only considers a single stock data source and overlooks external factors such as sentiment information and news events that may impact stock prices. Most prediction models use the relevant features on day t to predict the stock price on day t+n. Hence, there is research value and significance in exploring auxiliary model-based decision-level fusion and two-stage decision-level fusion as alternative approaches.

Again, from the point of view of the techniques fused at the decision level, the main methods applied in stock price prediction research are Voting, Weighted sum, Averaging, Stacking, LSTM, and XGBoost, with the Majority voting being the most applied Voting method, and the Weighted averaging being the one that is applied more in Averaging method. In addition, tree-based algorithms, fuzzy algorithms, sorting and selection methods are also applied (Zhang et al., 2021a).

7. Conclusions

This paper presented a systematic review of research on stock price prediction from a data fusion perspective, summarizing three important data fusion levels in the field of stock price prediction: stock price prediction based on data-level fusion, stock price prediction based on feature-level fusion and stock price prediction based on decision-level fusion.

The results highlight that data fusion has been widely used and has achieved some success in the field of stock price prediction. Meanwhile, the application of data fusion in the field of stock price prediction provides some insights. The types of stock-related data available in the study vary, including mainly historical stock data, macroeconomic data, specific events, user sentiment information, chart information, etc. To improve the accuracy of prediction models, future studies could focus on broadening the scope of data collection and obtaining stock-related information as much as possible. For instance, as a new type of additional event information, the domain knowledge graph contains some stock-related knowledge. Introducing knowledge graph information can well enrich event descriptions and reduce sparsity, thus making better use of the event information and improving the accuracy of stock price prediction. In addition, considering the specificity of text information for the collection and processing of text information, natural language processing (NLP) techniques such as generative adversarial network (GAN) and autoencoder can be used to accomplish text classification, text information extraction and other tasks to improve the ability of acquiring stock-related text information.

This paper has certain content limitations that can be addressed in future work. One limitation is that the summary of the paper is not extensive enough and lacks necessary detail. For example, the stock price prediction problem can be further categorized into short-term prediction and long-term

classification. Future studies can broaden the scope and quantity of literature collection to gain a better understanding of research progress in this field over various periods. Another limitation is that the issues addressed in this paper only cover a small portion of the finance field, and the stock price prediction field has relatively mature research. Future studies can focus on cutting-edge research topics in the current financial market, such as anomaly detection, quantitative trading and other emerging research domains.

Acknowledgments

Use of AI tools declaration

The authors declare they have not used artificial intelligence (AI) tools in the creation of this article.

Conflict of interest

The authors declare that there is no conflict of interest.

References

Abraham A, Auyeung A (2009) Integrating Ensemble of Intelligent Systems for Modeling Stock Indices. In: Mira, J., Álvarez, J.R., *Artificial Neural Nets Problem Solving Methods,* Eds., Berlin: Springer, 774–781. https://doi.org/10.1007/3-540-44869-1_98

Alhnaity B, Abbod MF (2020) A new hybrid financial time series prediction model. *Eng Appl Artif Intel* 95: 103873. https://doi.org/10.1016/j.engappai.2020.103873

Ariyo AA, Adewumi AO, Ayo CK (2014) Stock Price Prediction Using the ARIMA Model. *2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation*, 106-112. https://doi.org/10.1109/uksim.2014.67

Barak S, Arjmand A, Ortobelli S (2017) Fusion of multiple diverse predictors in stock market. *Inform Fusion* 36: 90–102. https://doi.org/10.1016/j.inffus.2016.11.006

Bollen J, Mao H, Zeng X (2011) Twitter mood predicts the stock market. *J Comput Sci* 2: 1–8. https://doi.org/10.1016/j.jocs.2010.12.007

Brogaard J, Zareei A (2022) Machine Learning and the Stock Market. *J Financ Quant Anal* 58: 1431–1472. https://doi.org/10.1017/s0022109022001120

Carta S, Corriga A, Ferreira A, et al. (2021) A multi-layer and multi-ensemble stock trader using deep learning and deep reinforcement learning. *Appl Intell* 51: 889–905. https://doi.org/10.1007/s10489-020-01839-5

Chandrasekara V, Tilakaratne CD, Mammadov M (2019) An Improved Probabilistic Neural Network Model for Directional Prediction of a Stock Market Index. *Appl Sci* 9: 5334. https://doi.org/10.3390/app9245334

Cheng K, Huang M, Fu C, et al. (2021) Establishing a Multiple-Criteria Decision-Making Model for Stock Investment Decisions Using Data Mining Techniques. *Sustainability* 13: 3100. https://doi.org/10.3390/su13063100

Chiong R, Fan Z, Hu Z, et al. (2018) A sentiment analysis-based machine learning approach for financial market prediction via news disclosures. In *Proceedings of the Genetic and Evolutionary Computation Conference Companion*. https://doi.org/10.1145/3205651.3205682

Chong L, Lim KG, Lee CC (2020) Stock Market Prediction using Ensemble of Deep Neural Networks. *2020 IEEE 2nd International Conference on Artificial Intelligence in Engineering and Technology (IICAIET),* 1–5. https://doi.org/10.1109/iicaiet49801.2020.9257864

Daradkeh MK (2022) A Hybrid Data Analytics Framework with Sentiment Convergence and Multi-Feature Fusion for Stock Trend Prediction. *Electronics* 11: 250. https://doi.org/10.3390/electronics11020250

Dash R, Samal S, Dash R, et al. (2019) An integrated TOPSIS crow search based classifier ensemble: In application to stock index price movement prediction. *Appl Soft Comput* 85: 105784. https://doi.org/10.1016/j.asoc.2019.105784

Evans L, Owda M, Crockett K, et al. (2018) Big Data Fusion Model for Heterogeneous Financial Market Data (FinDf). In *Springer eBooks*, 1085–1101. https://doi.org/10.1007/978-3-030-01054-6_75

Gandhmal DP, Kumar KS (2019) Systematic analysis and review of stock market prediction techniques. *Comput Sci Rev* 34: 100190. https://doi.org/10.1016/j.cosrev.2019.08.001

García-Medina A, Sandoval L, Junior Bañuelos EU, et al. (2018) Correlations and flow of information between the New York Times and stock markets. *Physica D* 502: 403–415. https://doi.org/10.1016/j.physa.2018.02.154

Giacomel FDS, Pereira ACM, Galante R (2015) Improving Financial Time Series Prediction Through Output Classification by a Neural Network Ensemble. In: Chen, Q., Hameurlain, A., Toumani, F., Wagner, R., Decker, H., *Database and Expert Systems Applications*, Eds., Cham: Springer, 331–338. https://doi.org/10.1007/978-3-319-22852-5_28

Guo Z, Wang H, Liu Q, et al. (2014) A Feature Fusion Based Forecasting Model for Financial Time Series. *PLOS ONE* 9: e101113. https://doi.org/10.1371/journal.pone.0101113

Ho TK, Hull JR, Srihari SN (1994) Decision combination in multiple classifier systems. *IEEE T Pattern Anal* 16: 66–75. https://doi.org/10.1109/34.273716

Hu Z, Zhao Y, Khushi M (2021) A Survey of Forex and Stock Price Prediction Using Deep Learning. *Appl Syst Innov* 4: 9. https://doi.org/10.3390/asi4010009

Jeantheau T (2004) A link between complete models with stochastic volatility and ARCH models. *Financ Stoch* 8: 111–131. https://doi.org/10.1007/s00780-003-0103-6

Keller C, Siegrist M (2006) Investing in stocks: The influence of financial risk attitude and values-related money and stock market attitudes. *J Econ Psychol* 27: 285–303. https://doi.org/10.1016/j.joep.2005.07.002

Khuwaja P, Khowaja SA, Khoso I, et al. (2020) Prediction of stock movement using phase space reconstruction and extreme learning machines. *J Exp Theor Artif Intell* 32: 59–79. https://doi.org/10.1080/0952813x.2019.1620870

Kim T, Kim H (2019) Forecasting stock prices with a feature fusion LSTM-CNN model using different representations of the same data. *PLOS ONE* 14: e0212320. https://doi.org/10.1371/journal.pone.0212320

Kristjanpoller RW, Michell VK (2018) A stock market risk forecasting model through integration of switching regime ANFIS and GARCH techniques. *Appl Soft Comput* 67: 106–116. https://doi.org/10.1016/j.asoc.2018.02.055

Kuo R, Lee LJ, Lee C (1996) Integration of artificial neural networks and fuzzy Delphi for stock market forecasting. *1996 IEEE International Conference on Systems, Man and Cybernetics. Information Intelligence and Systems* 2: 1073–1078. https://doi.org/10.1109/icsmc.1996.571232

Lahmiri S (2018) A Technical Analysis Information Fusion Approach for Stock Price Analysis and Modeling. *Fluct Noise Lett* 17: 1850007. https://doi.org/10.1142/s0219477518500074

Lahmiri S, Boukadoum M (2015) Intelligent Ensemble Forecasting System of Stock Market Fluctuations Based on Symetric and Asymetric Wavelet Functions. *Fluct Noise Lett* 14: 1550033. https://doi.org/10.1142/s0219477515500339

Lai S, Ye C, Zhou H (2021) Chinese stock trend prediction based on multi-feature learning and model fusion. *2021 IEEE International Conference on Smart Data Services (SMDS)*, 18–23. https://doi.org/10.1109/smds53860.2021.00013

Lee KC, Kim WH (1995) Integration of human knowledge and machine knowledge by using fuzzy post adjustment: its performance in stock market timing prediction. *Expert Syst* 12: 331–338. https://doi.org/10.1111/j.1468-0394.1995.tb00270.x

Lee T, Teisseyre P, Lee J (2023) Effective Exploitation of Macroeconomic Indicators for Stock Direction Classification Using the Multimodal Fusion Transformer. *IEEE Access* 11: 10275–10287. https://doi.org/10.1109/access.2023.3240422

Li X, Wu P, Wang W (2020) Incorporating stock prices and news sentiments for stock market prediction: A case of Hong Kong. *Inform Process Manag* 57: 102212. https://doi.org/10.1016/j.ipm.2020.102212

Li AH, Wang DW, Xu WJ, et al. (2022a) Anomaly Detection of Growth Enterprise Market Listed Companies with Financial Fraud Based on Data Fusion. *Data Analysis and Knowledge Discovery* 7: 33–47. Available from: http://kns.cnki.net/kcms/detail/10.1478.G2.20220920.1740.004.html

Li AH, Xu WJ, Shi Y (2022b) Framework of business intelligence and analysis based on data fusion. *Comput Sci* 49: 185–194. https://doi.org/10.11896/jsjkx.211100080

Lin G, Lin A, Cao J (2021) Multidimensional KNN algorithm based on EEMD and complexity measures in financial time series forecasting. *Expert Syst Appl* 168: 114443. https://doi.org/10.1016/j.eswa.2020.114443

Liu P, Zhang Y, Bao F, et al. (2022) Multi-type data fusion framework based on deep reinforcement learning for algorithmic trading. *Appl Intell* 53: 1683–1706. https://doi.org/10.1007/s10489-022-03321-w

Liu Y, Yu X, Wu Y, et al. (2021a) Forecasting Variation Trends of Stocks via Multiscale Feature Fusion and Long Short-Term Memory Learning. *Sci Programming* 1–9. https://doi.org/10.1155/2021/5113151

Liu Z, Huynh TLD, Dai P (2021b) The impact of COVID-19 on the stock market crash risk in China. *Res Int Bus Financ* 57: 101419. https://doi.org/10.1016/j.ribaf.2021.101419

Long J, Chen Z, He W, et al. (2020) An integrated framework of deep learning and knowledge graph for prediction of stock price trend: An application in Chinese stock exchange market. *Appl Soft Comput* 91: 106205. https://doi.org/10.1016/j.asoc.2020.106205

Lu R, Lu M (2021) Stock Trend Prediction Algorithm Based on Deep Recurrent Neural Network. *Wirel Commun Mob Com* 2021: 1–10. https://doi.org/10.1155/2021/5694975

Nofsinger JR (2005) Social Mood and Financial Economics. *J Behav Financ* 6: 144–160. https://doi.org/10.1207/s15427579jpfm0603_4

Malkiel BG, Fama EF (1970) EFFICIENT CAPITAL MARKETS: A REVIEW OF THEORY AND EMPIRICAL WORK. *J Financ* 25: 383–417. https://doi.org/10.1111/j.1540-6261.1970.tb00518.x

Malkiel EF (2015) A random walk down Wall Street: the time-tested strategy for successful investing. *Choice Reviews Online* 52: 52–6493. https://doi.org/10.5860/choice.191812

Melin P, Soto J, Castillo O, et al. (2012) A new approach for time series prediction using ensembles of ANFIS models. *Expert Syst Appl* 39: 3494–3506. https://doi.org/10.1016/j.eswa.2011.09.040

Nezhad MF, Bidgoli BM (2019) Development of an Ensemble Learning-based intelligent model for Stock Market Forecasting. *Sci Iran* 28: 395–411. https://doi.org/10.24200/sci.2019.50353.1654

Nti IK, Adekoya AF, Weyori BA (2021) A novel multi-source information-fusion predictive framework based on deep neural networks for accuracy enhancement in stock market prediction. *J Big Data* 8：1–28. https://doi.org/10.1186/s40537-020-00400-y

Qian B, Rasheed K (2007) Stock market prediction with multiple classifiers. *Appl Intell* 26: 25–33. https://doi.org/10.1007/s10489-006-0001-7

Qiu X, Zhu H, Suganthan PN, et al. (2017) Stock Price Forecasting with Empirical Mode Decomposition Based Ensemble ν-Support -Support Vector Regression Model. In: Mandal, J., Dutta, P., Mukhopadhyay, S., *Computational Intelligence, Communications, and Business Analytics*, Eds., Singapore: Springer 775: 22–34. https://doi.org/10.1007/978-981-10-6427-2_2

Sawhney R, Mathur P, Mangal A, et al. (2020) Multimodal Multi-Task Financial Risk Forecasting. *Proceedings of the 28th ACM International Conference on Multimedia*, 456–465. https://doi.org/10.1145/3394171.3413752

Shi S, Liu W, Jin M (2012) Stock price forecasting using a hybrid ARMA and BP neural network and Markov model. *2012 IEEE 14th International Conference on Communication Technology*, 981–985. https://doi.org/10.1109/icct.2012.6511341

Shi Z, Wu Z, Shi S, et al. (2022) High-Frequency Forecasting of Stock Volatility Based on Model Fusion and a Feature Reconstruction Neural Network. *Electronics* 11: 4057. https://doi.org/10.3390/electronics11234057

Shields R, Zein S, Brunet N (2021) An Analysis on the NASDAQ's Potential for Sustainable Investment Practices during the Financial Shock from COVID-19. *Sustainability* 13: 3748. https://doi.org/10.3390/su13073748

Stoean C, Paja W, Stoean R, et al. (2019) Deep architectures for long-term stock price prediction with a heuristic-based strategy for trading simulations. *PLOS ONE* 14: e0223593. https://doi.org/10.1371/journal.pone.0223593

Sun L, Xu W, Liu J (2021) Two-channel Attention Mechanism Fusion Model of Stock Price Prediction Based on CNN-LSTM. *ACM Transactions on Asian and Low-resource Language Information Processing* 20: 1–12. https://doi.org/10.1145/3453693

Thakkar A, Chaudhari K (2021) Fusion in stock market prediction: A decade survey on the necessity recent developments and potential future directions. *Inform Fusion* 65: 95–107. https://doi.org/10.1016/j.inffus.2020.08.019

Tulyakov S, Jaeger S, Govindaraju V, et al. (2008) Review of Classifier Combination Methods. In: Marinai, S., Fujisawa, H., *Machine Learning in Document Analysis and Recognition*. Eds., Berlin: Springer 90: 361–386. https://doi.org/10.1007/978-3-540-76280-5_14

Wang Q, Xu W, Zheng H (2018) Combining the wisdom of crowds and technical analysis for financial market prediction using deep random subspace ensembles. *Neurocomputing* 299: 51–61. https://doi.org/10.1016/j.neucom.2018.02.095

Wang Y, Liu H, Guo Q, et al. (2019) Stock Volatility Prediction by Hybrid Neural Network. *IEEE Access* 7: 154524–154534. https://doi.org/10.1109/access.2019.2949074

Wang Y, Yan K (2023) Application of Traditional Machine Learning Models for Quantitative Trading of Bitcoin. *Artif Intell Evol* 2023: 34–48. https://doi.org/10.37256/aie.4120232226

Xiao J, Zhu X, Huang C, et al. (2019) A New Approach for Stock Price Analysis and Prediction Based on SSA and SVM. *Intl J Inf Tech Decis Mak* 18: 287–310. https://doi.org/10.1142/s021962201841002x

Xie Q, Cheng G, Xu X, et al. (2018) Research Based on Stock Predicting Model of Neural Networks Ensemble Learning. *MATEC Web of Conferences* 232*:* 02029. https://doi.org/10.1051/matecconf/201823202029

Yang Y, Hu X, Jiang H (2022) Group penalized logistic regressions predict up and down trends for stock prices. *North Am J Econ Financ* 59: 101564. https://doi.org/10.1016/j.najef.2021.101564

Yan K, Wang Y, Li Y (2023) Enhanced Bollinger Band Stock Quantitative Trading Strategy Based on Random Forest. *Artif Intell Evol* 2023: 22–33. https://doi.org/10.37256/aie.4120231991

Yang YJ, Yang YM, Xiao JH (2020) A Hybrid Prediction Method for Stock Price Using LSTM and Ensemble EMD. *Complexity* 2020: 1–16. https://doi.org/10.1155/2020/6431712

Zhang C, Sjarif NNA, Ibrahim R (2022a) Decision Fusion for Stock Market Prediction: A Systematic Review. *IEEE Access* 10: 81364–81379. https://doi.org/10.1109/access.2022.3195942

Zhang G, Xu L, Xue Y (2017a) Model and forecast stock market behavior integrating investor sentiment analysis and transaction data. *Cluster Comput* 20: 789–803. https://doi.org/10.1007/s10586-017-0803-x

Zhang J, Li L, Chen W (2021) Predicting Stock Price Using Two-Stage Machine Learning Techniques. *Comput Econ* 57: 1237–1261. https://doi.org/10.1007/s10614-020-10013-5

Zhang Q, Qin C, Zhang Y, et al. (2022b) Transformer-based attention network for stock movement prediction. *Expert Syst Appl* 202: 117239. https://doi.org/10.1016/j.eswa.2022.117239

Zhang X, Qu S, Huang J, et al. (2018) Stock Market Prediction via Multi-Source Multiple Instance Learning. *IEEE Access* 6: 50720–50728. https://doi.org/10.1109/access.2018.2869735

Zhang X, Zhang L (2022) Forecasting Method of Stock Market Volatility Based on Multidimensional Data Fusion. *Wirel Commun Mob Comput* 1–14. https://doi.org/10.1155/2022/6344064

Zhang X, Zhang Y, Wang S, et al. (2017b) Improving stock market prediction via heterogeneous information fusion. *Knowl Based Syst* 143: 236–247. https://doi.org/10.1016/j.knosys.2017.12.025

Zhang Y, Lu S (2021) Multi-Model Fusion Method and its Application in Prediction of Stock Index Movements. *2021 6th International Conference on Machine Learning Technologies*, 58–64. https://doi.org/10.1145/3468891.3468900

Zhong Y, Zhao Q, Rao W (2017) Predicting stock market indexes with world news. *2017 4th International Conference on Systems and Informatics (ICSAI)*, 1535–1540. https://doi.org/10.1109/icsai.2017.8248528

Zhou F, Zhang Q, Zhu Y, et al. (2022) T2V_TF: An adaptive timing encoding mechanism based Transformer with multi-source heterogeneous information fusion for portfolio management: A case of the Chinese A50 stocks. *Expert Syst Appl* 213: 119020. https://doi.org/10.1016/j.eswa.2022.119020

Zhou Z, Xu K, Zhao J (2018) Tales of emotion and stock in China: volatility causality and prediction. *World Wide Web* 21: 1093–1116. https://doi.org/10.1007/s11280-017-0495-4