Information Fusion 65 (2021) 95–107

Contents lists available at ScienceDirect

Information Fusion

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

Full length article

Fusion in stock market prediction: A decade survey on the necessity, recent developments, and potential future directions

Ankit Thakkar, Kinjal Chaudhari *

Department of Computer Science and Engineering, Institute of Technology, Nirma University, Ahmedabad 382 481, Gujarat, India

ARTICLE INFO

Keywords:
Stock market prediction
Information fusion
Feature fusion
Model fusion
Machine learning
Deep learning

ABSTRACT

Investment in a financial market is aimed at getting higher benefits; this complex market is influenced by a large number of events wherein the prediction of future market dynamics is challenging. The investors' etiquettes towards stock market may demand the need of studying various associated factors and extract the useful information for reliable forecasting. Fusion can be considered as an approach to integrate data or characteristics, in general, and enhance the prediction based on the combinational approach that can aid each other. We conduct a systematic approach to present a survey for the years 2011–2020 by considering articles that have used fusion techniques for various stock market applications and broadly categorize them into information fusion, feature fusion, and model fusion. The major applications of stock market include stock price and trend prediction, risk analysis and return forecasting, index prediction, as well as portfolio management. We also provide an infographic overview of fusion in stock market prediction and extend our survey for other finely addressed financial prediction problems. Based on our surveyed articles, we provide potential future directions and concluding remarks on the significance of applying fusion in stock market.

1. Introduction

The financial market is an attractive field of study; it offers a variety of opportunities to investors, market analysts, as well as researchers from various disciplines. The market participation perspectives may differ among individuals, for example, learning the market behaviours, deriving influential aspects, trading through stocks, predicting future market trend, recommending assets for portfolio management, etc., however, the lack of financial literacy and knowledge of fundamental economic principles can critically affect the investment returns [1]. Therefore, an individual's understanding and approach towards a financial market can determine the type as well as extent of information required to study this domain. The economic market can be considered as a combination of financial investments, transactions, potential earning and/or losing, and several other actions that are performed at a massive level. It has extensively reached to a large number of fields and consequently, it also gets affected by numerous events; with this regard, the financial market can be interpreted as a model of complex systems [2].

In financial markets, various instruments such as stocks, bonds, commodities, derivatives, currencies, etc. are explored, investigated, studied, and traded in an exhaustive manner; such trading is based on buying and selling of instruments. Stock market is a financial market where the new issues of stocks, i.e., initial public offerings (IPOs), are created and sold at the primary market whereas the succeeding buying and selling are carried out at the secondary market [3]. The primary motivation behind investing in a stock market is to gain potential benefits of the investment [4]; while careful tradings can earn higher returns, the associated risk may sometimes result into loss of valuables. Such markets are non-linear, highly volatile, and chaotic in nature; they can get influenced by various events and hence, experience fluctuations. Therefore, it raises the demand of market valuation and several analytical evaluation to study the market behaviours. Based on the knowledge and expertise of stock market, a fundamental analysis can be carried out by examining the dominating factors to primarily derive long-term predictions whereas using the historical stock price data, derived information can be integrated for a technical analysis [5]; such characteristics can be fused using different strategies to prepare a reliable forecasting model [6].

Fusion can be considered as a transformation of single or multiple aspects with an aim to derive effective representation; it can be understood as the process of combining various factors that can improve the performance and provide useful results. The diversity of data, features, methods, parameters, etc. play a vital role in exploring as well as exploiting its environment; such information can be fused with each other to get performance benefits. While data fusion is the

* Corresponding author.
E-mail addresses: ankit.thakkar@nirmauni.ac.in (A. Thakkar), 17ftphde21@nirmauni.ac.in (K. Chaudhari).

https://doi.org/10.1016/j.inffus.2020.08.019
Received 10 June 2020; Received in revised form 27 July 2020; Accepted 20 August 2020
Available online 26 August 2020
1566-2535/

fusion applied on the raw form of data, information fusion is concerned with the processed and/or derived data [7]. Fusion technique(s) can be applied on raw or processed data, features, methods, to name a few. Such fusion methods are also utilized in ensemble as well as hybrid models. In some cases, these two models are considered equivalent, however, in other cases, they primarily differ on the characteristics of associated models; while an ensemble approach combines two or more homogeneous models to overcome the weakness of each other, a hybrid approach combines two or more heterogeneous models for expanding the models' learning abilities [8]. The fusion techniques are widely applied to diverse fields of research in order to prepare a favourable model. The financial market is one of the complex real-world applications of fusion where highly non-linear data are determined and processed to forecast the market behaviours. With an increase in the demand of exploring and forecasting economic systems, fusion-based techniques are noteworthy to study their significance in stock market. In this article, we take this opportunity to perceive the necessity of fusion in stock market prediction and conduct a comprehensive survey on the recent developments; we also discuss potential future directions to provide a state-of-the-art survey on fusion-based stock market prediction.

1.1. Necessity of fusion in stock market

The stock market accumulates the buying and selling of stocks based on the ownership of traders. To analyse such markets, the historical stock market data as well as various factors can be combined to derive market patterns. The efficient market hypothesis (EMH) assumes that the available market information is incorporated within prices [9], however, the validation proof of EMH has been controversial over a period of time. On the other hand, the complexity levels of economic markets are defined to be the time-series nature, high cross-correlation with various entities, and collective market behaviour during extreme influencing events [2].

The behavioural finance has been studied with a primary focus on how psychology influences the behaviours of financial practitioners and its consequences with respect to the stock markets [10,11]. Such psychological biases can play a critical role in the market efficiency; therefore, an alternative theory namely, adaptive market hypothesis (AMH), has been proposed in contrast to the point of view where EMH assumes a frictionless market without imperfections [12]. AMH is developed based on the principles of evolutionary biology, i.e., competition, mutation, reproduction, and natural selection [13]; such an evolutionary perspective of AMH is a reconciliation which argues that the financial markets evolve and its efficiency varies with time [12]. Even being an essentially qualitative and descriptive hypothesis, the arguments of AMH have been widely supported with return predictability of the stock market [14,15] as well as foreign exchange rates [16]; also, its potential implications have been evaluated for the digital currency [12] as well as cryptocurrency market [17]. This emphasizes on the understanding that amongst different factors which can influence the stock market, a careful selection of complementing and/or contradicting aspects may be fused to enhance the collected information and thus, improve the predictions.

The fusion techniques consist of different data sources, processed information, derived features, and/or prediction methods that are combined under appropriate scenarios. For a nebulous stock market, fusion can aid to extract knowledge base from a diverse set of data sources; it can extend the strength of a model to overcome weakness of the other model; it can exploit the search space to derive effective solutions. Also, the dependency on specific market concepts can be explored with fusion. Therefore, to address and demonstrate the necessity of fusion in stock market prediction, this article presents a focused survey on the recent advances of fusion-based stock market prediction. The primary motivation behind this survey is to identify how limitations of one or more entities are overcome with the others and to denote the importance of fusion for real-world stock market applications. To the best of our knowledge, our review work is the foremost survey that is devoted to fusion in stock market prediction and the same can serve as the foundation in this field.

1.2. Survey strategy

In this article, we briefly categorize fusion into information fusion, feature fusion, and model fusion. With each fusion category, we explore the existing research work having primary applications of stock price and/or trend prediction, portfolio management, risk/return forecasting, as well as other stock market concepts; though the reviewed articles are mainly based on fusion techniques, some of them might not have explicitly referred their work as a fusion approach.

For collecting research articles for this survey, we applied a systematic strategy as follows. Using Google Scholar search engine, we conducted an initial search with terms "stock market prediction" and "fusion"; for a major concentration on the recent studies, we restricted our search within years 2011 – 2020. We extended the search using additional terms such as "information fusion", "feature fusion", and "model fusion" whereas the potential applications related to stock markets were gathered using the terms "stock price prediction", "stock trend prediction", "portfolio management", "risk prediction", "return prediction", and replacing "prediction" by "forecasting" in each of these search operations, as well as using other financial market terms; to ensure that these terms were closely related to the actual work of the articles, we restricted our search such that articles having such terms in their titles were put together. We also set the exclusion criteria for articles resulting with the considered search operations; in case of having the above-mentioned terms mainly occurring in the reference section, the articles were examined to find if the terms were being referred for the related work and not for the actual proposed approach; in such scenario, the articles were excluded from being considered in our survey. To minimize the redundancy, we eliminate similar articles from being reviewed, if found. On the other hand, it was found that though specific fusion methods were applied in a set of articles, "fusion" term had not been specifically used to describe the same; therefore, we closely studied a large number of articles to identify whether they represented fusion at any level; thus, for an exhaustive survey, we advanced our collection by including such fusion-based articles as well as the articles related to the ones downloaded using the aforementioned search strategies.

1.3. Related work

It is observed that majority of the existing review articles have not precisely focused on fusion in stock market, however, they might have referred to fusion-based articles as a part of the survey. Therefore, we compare our survey with the existing surveys even if they are partially related to the central focus of our survey.

The computational approaches include various techniques that are adopted to predict stock market; artificial neural network (ANN) is one of the machine learning approaches that have been largely applied to stock market prediction. Some of the ANN-based models including a fusion model, as well as other techniques, were studied in Ref. [18]. On the other hand, a classifier fusion-based survey was carried out in study [19] where authors discussed the financial applications wherein the multiple classifier systems were integrated. A detailed study of fusion in opinion mining was conducted in review article [20]; along with various applications of opinion mining, stock market prediction using fusion of techniques was discussed. In Ref. [21], stock market prediction approaches were reviewed based on the soft computing techniques. One of the recent surveys on data fusion in smart city applications [22], authors discussed about smart economics based on fusion approaches whereas the survey on multi-source knowledge fusion [23], stock market prediction was discussed in the multi-modal-based aspect.

Table 1
A comparative analysis of our survey with the existing surveys under various criteria: C1 — Financial market conceptualization, C2 — Information fusion, C3 — Feature fusion, C4 — Model fusion, C5 — Stock price prediction, C6 — Stock trend prediction, C7 — Portfolio management, C8 — Risk/Return forecasting, C9 — Other stock market applications.

Criteria (→) Reference (↓)	C1	C2	C3	C4	C5	C6	C7	C8	C9
[18]				✓	✓	✓		✓	
[19]		✓		✓					✓
[20]				✓				✓	
[21]				✓	✓	✓			
[22]	✓	✓		✓	✓				
[23]		✓				✓			
[24]	✓			✓	✓				
[25]	✓	✓				✓			
Our survey	✓	✓	✓	✓	✓	✓	✓	✓	✓

For the quantitative financial applications, various machine learning techniques were reviewed in Ref. [24]. Subsequently, a fusion-based approach in financial text mining applications was covered under the survey on deep learning for financial applications [25]. Hence, we prepare a comparative analysis of our survey with the existing surveys under various criteria associated with fusion and stock market as shown in Table 1. It can be noted that amongst the existing surveys, a limited work has been regulated around fusion in stock market. In the presented survey, we take primitive steps of studying fusion perspectives based on stock market applications.

The fusion techniques can be classified using various criteria including relationships between input data sources, the type and nature of input/output data, the level of abstraction, the level of data processing, and the type of architecture [7]. To provide an infographic overview of fusion in stock market prediction, we present the classification criteria of fusion techniques and corresponding categories [7], the point of consideration to fuse information, feature, and model, as well as potential applications of fusion in stock market, as shown in Fig. 1; our survey is primarily focused on the three fusion approaches and associated stock market prediction techniques. This presentation is a general overview of fusion in stock market prediction and the same can be extended as well as modified as per the requirements. In our survey, the presented infographic is aimed to direct the readers of this survey through the wider scope of potential work based on fusion in the field of stock market; it can be considered as a walk-through summary to integrate fusion at various levels of operations within a stock-related prediction procedure. It can be mentioned here that each approach may not explicitly specify the adopted fusion criteria, however, for the completeness of understanding how specific fusion may be incorporated to an approach, the readers would be directed to refer to the study given by Ref. [7].

The organization of the remaining article can be summarized as follows: for the thorough coverage of all the criteria (C1–C9) mentioned in Table 1, Section 2 reviews information fusion in stock market which includes fusion techniques based on raw data as well as processed data; Section 3 describes feature fusion in stock market and elaborates its potential applications; Section 4 presents model fusion in stock market including learning models as well as evolutionary models; Section 5 discusses various aspects of fusion in stock market and the potential future directions; Section 6 provides the concluding remarks on the presented survey.

2. Information fusion in stock market

The time-series stock market data presents the market tradings; in its raw form, the historical data generally includes open, close, high, low, and volume information of the trading day; depending on the extent of details required, such data can be derived for various trading frequencies, for example, minute-wise, on hourly basis, daily, etc. While open and close indicate the opening and closing prices, respectively, high and low indicate the highest and the lowest prices attained in the trading frequency duration, respectively; volume is the representation of number of shares traded in the given duration. This data can be further processed to derive technical indicators; also, it can be appended with data from other sources to expand the coverage of corresponding information about the stock. The merger and/or expansion of data from multiple sources raise an opportunity of data as well as information fusion; here, information fusion can be considered one step ahead of the raw data fusion.

One of the classifications categorizes data fusion based on the relationship between corresponding input data sources [7,26]. It includes complementary, redundant, and cooperative criteria that represent data on different aspects, different viewpoints, and different modules, respectively [7]. In stock market, the historical data can be diversified using information fusion and the rich information can be employed for reliable prediction. To provide an overall perspective of information fusion in stock market prediction, we prepare a graphical representation as shown in Fig. 2. The major data sources include quantitative and crowd-sourced data along with the derived information based on fundamental and technical analyses; these sources further represent various categories that define specific information associated with the targeted stock. The given representation can be referred to study which combination of information sources is applied in respective work under information fusion; it can be treated as a general overview to analyse how different fusion aspects are acquired and utilized for different stock-related forecasting applications as shown in Fig. 1; this section reviews various approaches associated with information fusion in stock market.

2.1. Stock price/trend prediction

One of the primary concerns towards the stock market is prediction of stock price and its movement direction; a considerable amount of work has been carried out that analyses market etiquettes and predicts the future price and/or trend. A highly volatile stock market experiences large fluctuations and thus, forecasting the price values and estimating whether the market would move upwards, downwards, or remain steady is difficult. Such predictions require study of various influential data as well as derivation of the previous patterns of price movements. Hence, information fusion can be integrated with the existing techniques to build reliable knowledge bases for improving stock market prediction.

The major concern around stock market analysis includes fundamental as well as technical analyses; while the concerned features of such analyses differ, their fusion can be beneficial for a stronger information source while predicting the stock market. One of the data fusion techniques proposed to integrate economic and financial parameters for the fundamental analysis with the stock price change over time-based technical analysis [27]. As the stock price was proven to have a lognormal distribution [28], it was demonstrated that the logarithm of a stock price at the given time followed normal distribution with mean and variance values; such lognormal models were developed for financial assets such as stock prices [29]. Subsequently, authors in Ref. [27] inferred from the histogram of stock prices for the companies of Iran that the stock price change followed normal distribution. An extended Kalman filter (EKF) was adopted to combine the analyses information; the predicted next-day stock price trend using data fusion as well as change in error using EKF structure indicated performance improvements as compared to regression and ANN models [27,30].

For different stock market-related events, various discussions and opinions take place where experts may provide their insights, suggestions, as well as predictions. Such events are explored from diverse perspectives in order to obtain inherent patterns to understand the market behaviour. While historical time-series data can provide primary

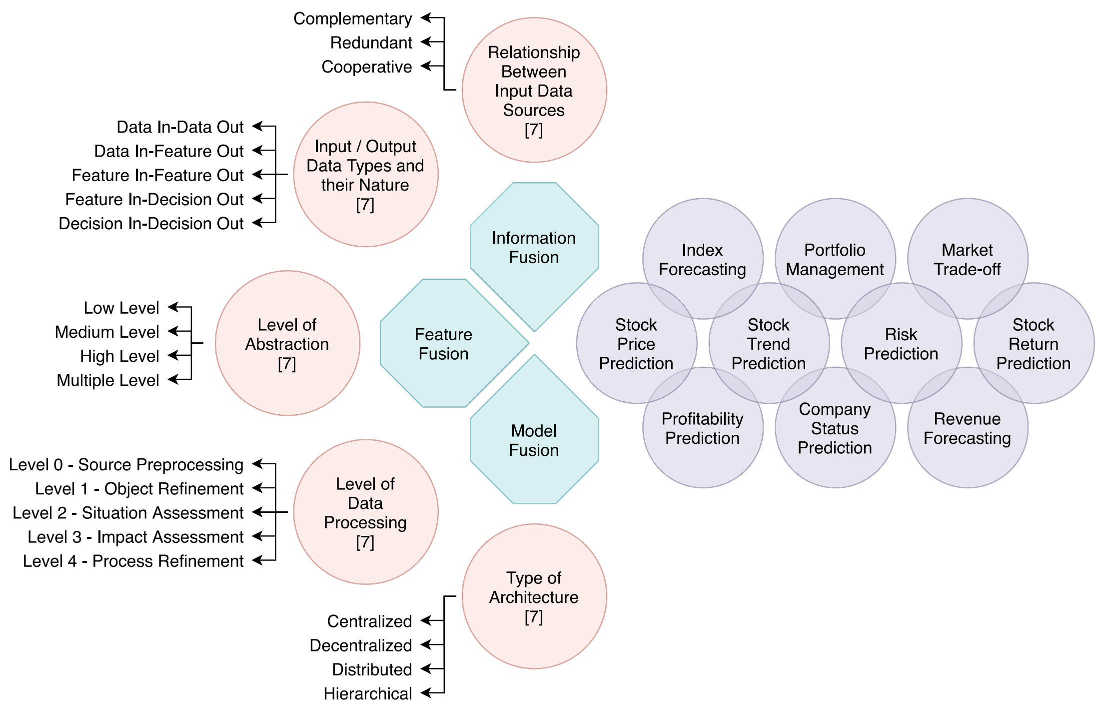

Fig. 1. An infographic overview of fusion in stock market prediction.

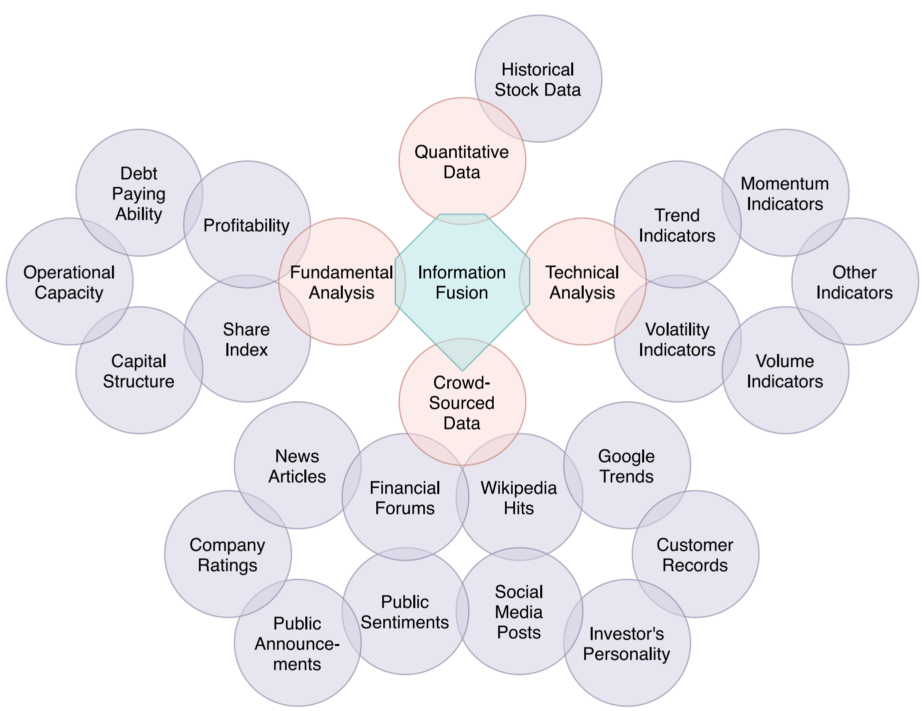

Fig. 2. Graphical representation of information fusion in stock market prediction.

information, investors may follow social opinions as well; such opinions can be collected using market analysis, online polls, discussion forums, etc. With the help of crowd-sourced knowledge bases, authors proposed to predict one-day-ahead stock price movement in Ref. [31]. Authors considered Google and Wikipedia platforms as the non-traditional experts; the historical stock market data were collected and technical indicators were derived, followed by integration of Wikipedia hits and Google news data to prepare a rich knowledge base. Using the combined data, authors generated features and utilized three machine learning models namely, decision tree (DT), support vector machine (SVM), and ANN for stock trend prediction. Also, an intelligent trading-based expert tool was presented to assist the decision-making process on various instruments [31].

On the other hand, the intuition that stocks with higher correlation could be affected by the same event was considered and an information fusion approach was applied to predict the stock trend in study [32]. Authors proposed to fuse quantitative, event-specific, and sentiment information, i.e., historical stock prices, financial Web news-based events, and financial discussion board-based user sentiments, respectively. The collected data were categorized based on similar events and further combined from such heterogeneous sources to derive their internal relations. The correlations enhanced the shared knowledge-based learning and hence, improved the prediction performance. Based on the collected data from various sources, another fusion approach was proposed to predict the fluctuation trend of the given stock [33]; using a tensor-based sub-mode coordinate (SMC) algorithm, the subspace dimensions were reduced according to the stock similarity. Authors proposed to fuse data collected from financial Web news, sentiments of the investors extracted from social media platform, and quantitative data; the enhanced features were further given to long short-term memory (LSTM) model for the prediction. Hence, the diverse data sources can help for potentially fusing information to build a knowledge base that can be utilized to derive inherent market patterns.

Another way of approaching the market sentiments is through social media such as Twitter. An information fusion approach was proposed to determine the relationship between Twitter sentiments and stock price and the fused information was used to predict one-day-ahead stock price change in Ref. [34]. Authors considered that authoritative Twitter users could be derived using the number of followers and the count of retweets; using this information along with the number of positive and negative tweets, authors calculated the sentiment weighting factors for each day for the Twitter users. The granger causality test and predictions using vector autoregression (VAR) improved the prediction performance; also, the authors concluded that a minimum of 2500 daily tweets were required for such prediction [34]. Considering that a specific lag period represents strong correlation between the social sentiments and market prices, an SVM-based non-linear granger causality approach was proposed in study [35]. It fused the sentiments with historical stock market data for extracting a higher level of statistical significance at the particular lag period. The optimal lag periods were evaluated and daily stock prices were predicted for four companies. The improved results indicated the impact of public sentiments and hence, the need of information fusion at various stages to understand the target in a broader way.

As opposed to the public data, another useful approach may be the trading behaviours of investors for the specific company stocks; while the market sentiments present public moods and may be affected by certain events, an investor's transaction records can aid the knowledge base for understanding the actual dominance for trading. Based on such transaction records as well as public market information, a stock trend prediction approach was proposed in study [36]; authors explored potential fusion of such private and public data and derived relevant stocks using correlation-based knowledge graph. To maintain the privacy of investors, their identities were desensitized and features based on the transactions as well as volume information were further selected. Authors adopted an attention-based bidirectional LSTM model for predicting stock price movement direction. Hence, the obtained results signified the usefulness of higher level of data fusion.

2.2. Risk/return forecasting

A critical trade-off in the financial market is risk-return trade-off which indicates a rise in potential return with respect to an increase of risk. In other words, the higher expected returns are associated with the higher levels of risk; such trade-off is considered to be a fundamental aspect in finance [37]. Though some investors may choose to keep a safe trading and therefore, accept lower return on behalf of low risk, investors having risk-averse nature may desire higher return and hence, take high risk with the trading. Thus, it is important to forecast the potential return, inherent risk, as well as relation between them during stock valuation. Also, a considerable number of studies have been carried out to demonstrate the distributions of market returns such as fat-tailed and skewed return distribution [38,39]. Such financial statistics can be utilized by the learning models in order to infer the characteristics of stock market assets; the role of information fusion can be examined to identify data distribution and corresponding predictions can be evaluated.

One of the primary concerns of an investment is to make an early-warning that can be helpful in preventing the financial crisis of an investor as well as the investing firm; an information fusion approach was proposed for such financial early-warning prediction in [40]. Authors adopted the Dempster–Shafer theory (DST) that inferred based on an aggregation of independent information from different data sources to obtain a higher degree of belief. A total of 25 variables were selected to derive the financial crisis information and the prediction results were calculated using SVM as well as logistic models; these data were fused using DST to prepare the final rating of each company's probability of having an early-warning.

The market trends can be visualized using historical stock data; similarly, the market news can reflect various events, their impact on stock prices, discussions on corresponding analyses, as well as market behaviours. Such data from diverse perspectives can be fused to develop broader viewpoints that can be utilized for predicting stock return. An integrated approach was proposed in Ref. [41] by applying information fusion of market news and stock prices to forecast intra-day stock return. The inherent information from these sources were extracted using two sub-kernels of a multi-kernel support vector regression (MKSVR) system; the proposed information fusion outperformed models with single information source. Subsequently, sensor data that provide timely measurements of various events, along with social media data, were considered to derive informative features of the given stock [42]; for characterization of such time-series, the internal dependences were approximated using copula theory. Also, the social media data were assisted using Google trends and fusion was applied to infer events for stock return prediction.

While the investments in stock market are inherently risky, the logistics as well as lenders may provide a loan to the customers and their firms; the concern raises with the possibility of unability of the borrower to repay the lended money. Hence, it arises the potential credit risk; in order to prevent the large possibilities of facing such issues, identification of various factors associated with the credit risk must be carried out before making an investment. An information fusion approach was proposed to forecast credit risk of the customers [43]; the internal history of the customers such as financial and operational data, credit history, as well as customer characteristics information were collected along with the external dynamics such as industry index as well as relevant news. Such data were prepared into features and aligned for risk prediction. The improved results indicated the potential fusion of internal and external information for credit risk prediction and hence, for possible decision-making.

A daily stock return was predicted using three models namely, adaptive neuro-fuzzy inference system (ANFIS), ANN, and SVM in Ref. [44]; here, an information fusion-based sensitivity analysis was carried out to evaluate relative importance of individual variables given as input to the prediction model so as to reduce uncertainty of models and increase the extracted knowledge.

2.3. Stock index forecasting

From an investment perspective, various stock indices are studied and evaluated; the market valuation can be helpful to predict potential benefits of the investment. Before taking any decision on trading one or more indices, market liquidity, associated risk, potential returns, and other information are likely to be explored using fundamental aspects, governed regulations, official statements, as well as other directly or indirectly associated firms and their information. This can help in determining the expected growth of an index, its steadiness over time, investment strategies, and other useful factors.

Considering a decision support required for investment in financial markets, future indices on short-term as well as long-term horizons were predicted using regulatory disclosures in Ref. [45]. The proposed text mining approach collected ad-hoc announcements from the publishing service provider and integrated machine learning techniques under the decision support system. The German prime index (DAX), German composite index (CDAX), and STOXX Europe 600 (STOXX) indices were targeted and their stock data were fused with the derived sentiments from timely disclosures in both languages, English and German; the results indicated the significance of financial decision-making as well as decision support for index forecasting of exchange-traded funds (ETFs) [45].

2.4. Other applications

While the information fusion is significantly applied to the price, trend, and index forecasting, as well as risk and return prediction techniques, it has sparsely covered other financial applications.

In stock market, a large number of events can have influence on the market etiquettes; apart from the direct correlations between similar or dependent companies, the indirect associations among various firms can also affect the stock market. Thus, identification and study of such market events can be a critical task. A knowledge-based complex event processing system was proposed in Ref. [46] to monitor the stock market based on the available real-time stock events and background knowledge; here, the companies having joint-stocks were considered and their publicly available data were fused to prepare a knowledge base. Initially, this approach was employed for addressing user queries by extracting the company knowledge base and fusing it with event data streams [47].

Apart from various predictions about the future stock market price trend, the market status of a company can also be a significant interest of the traders. Ref. [48] was proposed to predict the one-month-ahead status about how promising an industry would be; considering that the securities companies provided ratings to various industries, authors adopted the DST and presented the degree of belief regarding the investment value of an industry. Based on the historical ratings, predictions were made for each trading day to determine the most promising industries for the next month; also, the higher average rise rate was observed for the predicted industries.

It can be observed that various factors affect the market dynamics of a stock index; hence, development of a correlation among such influential factors can be helpful for future predictions. Ref. [49] was proposed with a focus on the Bayesian networks to predict future stock price movement direction; while the graphical representation of these networks could be useful to visualize market etiquette, the multivariate nodes could also indicate associated risks; such beneficial approach was employed to predict and analyse the S&P market index [49]. The network topology was prepared using several factors such as interest rate, consumer price index, unemployment rate, money supply, housing start, and implied volatility index; based on their correlation with the future directions, some of the factors were eliminated as well. The conditional probability tables could derive the directions; these predictions were further utilized for trading strategies of S&P index futures (i.e., long or short) as well as options (i.e., put or call). Hence, the fusion of domain knowledge and historical data resulted into higher equity profits.

2.5. Summary

In this section, we review the stock-related forecasting applications that are addressed using information fusion. As observed in the reviewed articles, the historical stock data can be appended, combined, integrated, as well as fused with the collected and/or derived information and further applied for the prediction task.

The significance of adopting information fusion can be perceived from the wider range of applications covered by the same. It indicates an advantageous fusion where information diversity plays a vital role in enhancing the market knowledge and hence, improving the prediction performance. However, the selection of specific data sources for information fusion can be challenging; while some approaches fuse the crowd-sourced information, others choose to integrate derived information using various indicators. Information fusion of different data sources may introduce data sparsity as for each trading day, data from each of the sources may not be available; hence, pre-processing such data by eliminating days with missing value(s) may result into loss of information. The reliability of analysis-based data as well as the impartiality of public sentiment-related data may be questionable; here, a well-established methodology can be followed to extract specific information.

In the view towards improving the prediction performance using information fusion, the selection of supporting features, as well as prediction model, plays a critical role. It can be noticed that information fusion can primarily provide a collection of data, however, selection of an appropriate set of features requires a deeper understanding of the data characteristics as well as their suitability for the corresponding prediction model. Subsequently, the limitations of a prediction model can also be responsible for performance degradation. For example, difficulty in determining noise covariance matrixes and in satisfying all situations using fixed noise covariance matrixes while using EKF [50]; instability of DT [51]; lack of transparency in outputs of SVM [52]; slow convergence and being trapped in local optima using ANN [53]. Therefore, it is essential to identify the appropriateness of a specific approach for the target stock market application. Also, the motivation to fuse specific data as well as the rationale behind integrating diversity may require higher attention. On the other hand, identification of the amount of information diversity required to obtain desired prediction performance is critical; one of the potential limitations may be identification of an appropriate arrangement of the available data [54] such that the application of information fusion should enhance the prediction performance. It can be treated as a potential research scope to analyse the application requirement(s) and combine the required type of data sources. The fused information can also provide valuable features which can increase the prediction model's learning capabilities; hence, the steps following information fusion must also be carefully selected. Such research directives can be addressed to improve forecasting models.

3. Feature fusion in stock market

In a learning-based approach, features play a key role in expressing the given data from various perspectives as well as retrieving detailed information about different attributes. From the available list of features for the given problem, a set of features are constructed, selected, and/or extracted for preparing them suitable for the prediction model. The historical stock market data present basic stock trading information; technical indicators are derived as features to learn specific interpretation. On the other hand, because the market gets influenced by numerous events, corresponding attributes or market's behaviour based on a specific context can also be considered as potential features. Though individual features may be capable of revealing partial information about the market, feature fusion can be carried out to build a different facet of these features and hence, exploit the inherent characteristics of stock market. It may be performed as a characteristic-level abstraction as well as an object refinement of the processed data [7] and the fused features can be applied to various stock market applications.

In Fig. 3, we provide a graphical representation of feature fusion in stock market prediction; here, the derived features can be fused using various actions. Such features can have different types; they can also be analysed based on various aspects. Also, a variety of operations can be performed on these features to fuse them together; these operations may be applied individually or combined with other factors. Selection of specific set of operations and/or analytical factors for

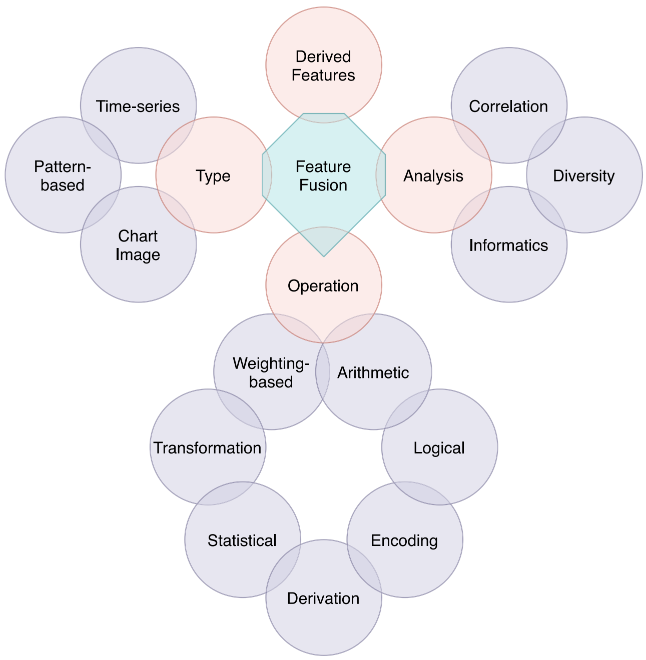

Fig. 3. Graphical representation of feature fusion in stock market prediction.

feature fusion may be dependent on the target application. The given pictorial representation can be considered as a generalized overview of feature fusion; it can be referred to observe which combination(s) of analysis and/or operation is adopted with a specific input type and applied to the stock-related forecasting application(s) as given by Fig. 1. This section reviews the stock-related forecasting applications based on feature fusion.

3.1. Stock price/trend prediction

The analysis of market trends is important for decision-making; it determines whether money should be invested to a specific stock as well as the length of duration to be considered for the same. The collected information about the given stock(s) can be integrated to derive useful features; to combine such features, various techniques can be performed with the aim to enhance the feature significance.

For one-day-ahead close price prediction, an SVR-based prediction model was proposed in Ref. [55]. Authors pre-processed the historical data and divided it into close price feature and a set of 39 derived technical indicators; individually, dimensions of these features were reduced using independent component analysis (ICA). Further, the extracted feature were fused using canonical correlation analysis (CCA) to derive union feature that was used for predicting next-day's close price. On the other hand, a minimal variation order weighted average (OWA) operator was employed to fuse price data periods in Ref. [56]; here, authors proposed to multiply features with their corresponding weights and summarize as a single aggregated factor; clustering and fuzzy rules were combined to train ANFIS and future stock prices were predicted.

While features of the similar representations can be combined, similar approach can be implemented with features having different descriptions. For stock price prediction, the temporal and graphical features were combined under feature fusion using LSTM and convolutional neural network (CNN)-based model in study [57]. Authors proposed to use the time-series stock market data along with stock chart images; the prepared stock charts were evaluated for price prediction and combined with volume information to create fusion chart images. The results indicated higher performance of fusion chart images as compared to stock chart images.

The combination of technical and fundamental analyses are generally observed while applying information fusion, however, the derived features from such analyses can be capable of providing useful features. In Ref. [58], authors considered technical indicators along with the sentiments derived from news articles and applied feature fusion to predict the stock market trend based on a portfolio of 20 stocks. The experiments based on fusing sentiment embeddings with the quantitative price data indicated performance enhancement and hence, the importance of feature fusion.

3.2. Portfolio management

Stock market investments are aimed to benefit the investors; as compared to investing in a single asset, a large number of investments are carried out on multiple assets. This becomes a useful approach where the invested money would not be fully dependent on the market performance of single asset. Selection of multiple assets also requires attention; assets belonging to the same group or having common characteristics would be similarly affected by specific events. To overcome the dominance of alike assets, diversification can be highly desirable; such assets, preferably diverse in nature, can be combined within a portfolio wherein the asset liquidity can be associated with risk aversion behaviour of an investor [59].

The maximization of the expected returns is a primary concern of portfolio management. For the diversified stocks of a portfolio, long-term management scheme was proposed in Ref. [60]; authors integrated a list of fundamental indicators as stock attributes and evaluated the cognitive diversity scores between attributes. Based on the diversity strength as well as performance strength, the selected attributes were fused using combinational methods and portfolios with different number of stocks were tested according to the potential returns. Hence, combinational fusion based on various attributes can also be extended for portfolio management.

3.3. Summary

In a prediction model, features are crucial factors that define specific aspect of the input data and help in deriving useful information for the model to learn; feature fusion can be expected to inherit such information and enhance the forecasting performance. Based on the reviewed articles, it can be observed that feature fusion is explored to a limited extent as compared to information fusion in stock market. It can be understood that the derived features may undergo feature selection and/or extraction and therefore, only a few approaches focus on the fusion techniques applicable to such features, however, the existing work presents the scope of exploring feature fusion based on analytical as well as operational aspects.

One of the potential challenges of feature fusion is the difficulty to choose useful features [61]. Selection of appropriate features can significantly enhance the prediction performance; such features may be chosen based on various factors such as their characteristics, capabilities of exploring as well as exploiting information space, diversity, correlations, to name a few. Therefore, feature fusion requires more attention towards which kind of features would be fused together and how the fusion would be carried out. Such rationale may be adopted at a larger level with justification on the need of fusion for the specific application; also, the possibility of supporting feature fusion with information fusion and/or model fusion should be carefully examined. The feature extraction and selection techniques have an important role for this task; the usefulness of such techniques can be beneficial to derive informative features, however, the limitations of such methods may result into a poor set of features which, in turn, is likely to reduce the prediction performance; for example, nonlinearity of CCA [62]. Therefore, it is necessary to evaluate the technical aspects of the feature constructing approaches. It can also be noticed that stock market data are presented in different forms; the same may be extended at fusion level for retrieving rich set of features that can improve the prediction accuracy.

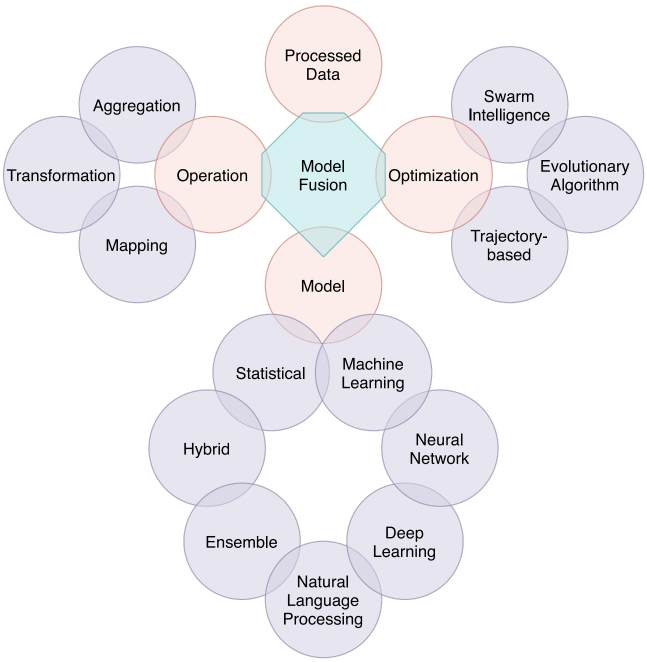

Fig. 4. Graphical representation of model fusion in stock market prediction.

4. Model fusion in stock market

Information fusion and feature fusion are applied to stock market prediction; these approaches respectively focus on the data and derived attributes that can enhance the intrinsic perspectives. Another way to improve prediction is model-based fusion where inherent characteristics of prediction models are exemplified to contribute towards improved forecasting. Based on the architecture design where fusion would be integrated, a model fusion can be classified into centralized, decentralized, distributed, and hierarchical architectures [7]. The considered models may be homogeneous or heterogeneous and they are aimed to be fused to explore as well as exploit the prediction capabilities of a fused model. Some of the model fusion techniques [63] include linear weighted fusion, cross-fusion, waterfall fusion, additive fusion, predictive fusion, stack fusion, etc. Among various applications of such model fusion techniques, financial markets are also studied to various extents.

To elaborate the wide spread of model fusion in stock-related forecasting applications, we demonstrate a graphical representation as shown in Fig. 4. For improving the prediction performance of a model, one of the important tasks is parameter optimization; as compared to the randomly selected parameters, different metaheuristics as well as trajectory-based techniques can be employed for tuning the model parameters. On the other hand, the model fusion can be carried out by performing various operations in order to fuse specific features of two or more models; it must be noted that these operations may differ from the possible operations carried out for feature fusion given by Fig. 3. Such operations as well as optimization can be applied to a wide range of models; also, the hybrid or ensemble models can be integrated with others for a model fusion. While we provide a generalized representation of the model fusion, the same can be modified depending on the type of input data as well as the target application; the given representation can be referred to identify which of the potential techniques, i.e., operation and/or optimization, is considered for model fusion and further applied to stock-related forecasting applications as shown in Fig. 1. This section describes various stock market applications addressed using model fusion.

4.1. Stock price/trend prediction

Direct or fusion-based techniques can be applied on the data processing as well as feature-related steps of a prediction approach and they are further given as inputs to the prediction model; one of the expected outcomes of a prediction model is reliable forecasting. Model fusion can be considered as an application to the stock price/trend prediction; this fusion may combine two or more architectures or it may apply modules of different prediction models in order to fuse the capabilities of stock market prediction.

A fusion model was proposed to predict the trend of stock close price in Ref. [64]; authors considered ANFIS model to advance the capabilities of ANN and Sugeno fuzzy system where the membership function parameters were adjusted using clonal selection-based immune algorithm. Authors provided historical stock market data along with technical indicators to the ANFIS model where the data served as antigen and the prepared fuzzy rules were treated as the antibody; comparison indicated performance improvement using the proposed fusion model. On the other hand, considering wavelet analysis as an effective approach to process data information at various scales, it was fused with two models namely, ANN and ANFIS for predicting next-day close price in Ref. [65]. Authors adopted one dimensional discrete wavelet with level 3 and applied denoising on the normalized historical stock data; ANN and ANFIS models, with and without wavelet analysis, were evaluated individually and the prediction results were compared. These approaches indicated the significance of fusion over neural networks as well as fuzzy systems.

A large number of historical stock market data are analysed and various technical indicators can be derived; such data can be appended to prepare a set of features that can be further utilized for stock market prediction. However, operations with such large amount of data having non-linear complex relationships may be difficult. An OWA approach was fused with ANFIS model in study [66] for stock price prediction; here, OWA was adopted to reduce the higher dimensional stock dataset into an aggregate value and fuzzy rules were created using ANFIS. The predicted stock price indicated higher profits as compared to previous work. On the other hand, a Bayesian-regularized feed-forward ANN model was proposed in Ref. [67] to predict the one-day-ahead stock market trend; the prediction results indicated comparable mean absolute percentage error (MAPE) with one of the existing fusion models [68].

A two-stage fusion approach was proposed in Ref. [69] to predict future close price; authors proposed to integrate SVR in the first stage and fused SVR, ANN, and random forest (RF) in the second stage that combined SVR-ANN, SVR-RF, and SVR-SVR models for prediction. Results indicated performance improvement with fused models as compared to single prediction model. Subsequently, for predicting the stock price, a blended model fusion approach was proposed in study [70] where the regression and backpropagation using ANN algorithms were integrated to enhance the prediction.

Another approach for the stock trend prediction integrated recurrent neural network (RNN) and representative pattern discovery (RPD) in study [71]; the advantages of deriving short-term and long-term trends of time-series data using RNN and RPD, respectively, were fused to build an improved model. The representative patterns (RPs) were derived such that the subsequences contained higher similarity with others whereas lower similarity with each others [71]; the dynamic time warping (DTW) distance between such RPs and historical stock data was calculated and matched pattern joints were used to adjust the prediction results of RNN; an increased trend prediction performance was achieved using this fusion approach.

Considering the wisdom of crowds, a fusion approach was proposed to use social media sentiments and technical indicators for stock trend prediction [72]; authors developed deep random subspace ensembles (DRSE) technique for efficient mining of market fluctuations and hence, flagging the stock trend.

4.2. Portfolio management

The returns of investment are crucial concerns for the investors; while the dynamic stock market offers fluctuations, investors may choose to be assured about the expected return along with the calculated risk. To ensure that amongst the time-varying return distributions, the desired returns are obtained, investors are likely to choose diverse assets and combine them under a portfolio. Such diversification of portfolio can be addressed in several ways; individuals may choose to select assets from different sectors [73]; individuals may aim for cross-country investments [74] and hence, diversify portfolios at domestic and international levels based on the geopolitical risks and financial stresses [75] to handle turbulences of economic conditions; other potential approach may diversify the investment styles using fundamental analysis and/or technical analysis; also, the concept of financial cognitive dissonance, i.e., changing the investment beliefs, may be adapted by the investors [76].

Stock market offers a variety of assets; selecting beneficial assets that can gain higher returns and hence, creating an optimal portfolio is a challenging task. Apart from the econometric models [77], computational models can also be applied, as well as fused, to address portfolio management. One of such fusion approaches was proposed in Ref. [78] with SVM and mean–variance (MV) for portfolio selection. Authors evaluated cardinality of the selected portfolios as well as associated risks and returns; the proposed approach (SVM + MV) indicated better performance as well as higher net profitability.

4.3. Risk/return forecasting

The desire of earning higher profits of the investment can enforce to move towards risky assets. Hence, it is necessary to estimate the expected returns and corresponding level of risk prior to an investment. Also, the duration of investment such as short-term, medium-term, or long-term, should be carefully chosen.

While the investors worry about the returns of their investment in stocks, the lenders may consider the potential risk of not receiving the owed amount as well as its interest; the credit risk concept can be related to this kind of a scenario of having possible loss in case the borrower fails to repay the borrowed amount. This becomes a critical factor that demands systematic study of the investment and potential credit risk. A fusion-based credit risk assessment approach was proposed in Ref. [79]; authors proposed to apply an enhanced decision support model (EDSM) using relevance vector machine and DT to interpret the forecasting results. Also, the effectiveness of various criteria was evaluated and the future measurement designs were examined with the help of corporate transparency and information disclosure feasibility. Such assessments can be helpful in forecasting credit risks for the investors and hence, taking a well-examined decision.

It can be noticed that diversity is one of the major concerns while forecasting the stock market; such diversities are explored using various fusion techniques. For stock risk and return predictions, multiple diverse base classifiers were proposed to be fused for a common input data and a Meta classifier was developed to learn from outputs of such individual classifiers [80]. Authors included Bagging, Boosting, and AdaBoost methods with various base classifiers and analysed the prediction performance.

4.4. Other applications

The financial market offers a wide range of interesting concepts; an investor's point of interest can decide various factors such as how he/she would select the instruments, which assets would get higher weightage, what would be the trading strategy, when should the invested shares be sold out, etc. Such applications can be targeted for prediction using model fusion.

Investment in a company can be dependent on various factors; the profitability of a company is found to have considerable influence on the company's value [81]. The profitability of each share of the given company stocks can be calculated using earnings per share (EPS); an early work on the EPS forecasting was conducted in Ref. [82] using model fusion of autoregressive (AR) model and ANFIS. This approach was extended with further comparisons and analysis to predict EPS using the model fusion [83].

The sentiments associated with stock-related text are important attributes; an analytical approach was proposed with SVM and bootstrapping techniques to evaluate the stock sentiments [84]. Authors derived the stock review blogs using a web crawler and classified the emotions within the review text using SVM and reconstructed the set using bootstrapping classifier. The results indicated performance improvement using such model fusion.

One of the major concerns while investing in a company is its revenue; it indicates the turnover of a firm and hence, investors may look forward to the companies having a promising sale and therefore, higher returns of the investments. For predicting a company's future revenue, a model fusion was proposed in Ref. [63]; the features including multi-factor and time-series were integrated with the linear regression model and a training model of extreme gradient boosting, namely, XGBoost model. Authors selected the additive fusion method based on the prediction results and additional weights to forecast the company's revenue trend. Thus, a variety of financial applications based on model fusion refer to the potential combinations of homogeneous as well as heterogeneous models. Integration architecture-based and/or approach-based fusion techniques may enhance the forecasting capabilities; the same may be further improved by optimizing the parameters using some of the widely integrated swarm intelligence approaches such as particle swarm optimization (PSO) [85] as well as evolutionary computation-based algorithms such as genetic algorithm (GA) [86,87].

4.5. Summary

Various prediction models are developed and integrated with stock market data to predict future market behaviour; the capabilities of such models can be extended with optimal set of parameters, supportive input features, and other techniques. Model fusion is one of the important aspects to enhance the prediction performance of the given models. This section outlines how various model fusion are applied to stock market prediction.

It can be observed that model fusion enables the learning capability of the associated models; while individual models may not be able to learn different market perspectives, a fusion technique can be employed for this task. Such models may be fused using their architectural combinations along with different operations such as mapping, transformation, etc. However, the disadvantages of individual architectures [7] may propagate while fusing them with others; for example, curse of dimensionality problem with VAR [88] and ANFIS [89]; overfitting because of complex RF trees [90]. Therefore, it becomes essential to reduce the potential limitations of the associated models in order to minimize the expansion of individual's drawback(s). As the model selection is crucial to support the usefulness of pre-processing as well as feature selection/extraction steps, it is also required to choose appropriate operations in order to assist the forecasting model. A considerable amount of work provides the intuition towards fusing specific models. The same may be extended with information as well as feature fusion techniques in order to inherit exploration and exploitation capabilities of the corresponding models. The model fusion may also be dependent on the input data types; hence, a generalized conceptualization may be developed to study how specific kinds of models should be fused for the given parameters. Though different stock market applications are aimed using model fusion, a large number of opportunities can be noticed to apply model fusion for other stock-related forecasting tasks.

5. Discussions

The stock market complexities are widely addressed with a variety of information sources, attributes, methods, and other relevant aspects. It can be observed that a considerable amount of work has been carried out based on fusion at various levels; such fusion techniques are employed for enhancing the market information, extracting useful trading rules, as well as preparing a complementary system where the limitations of one or more approaches can be overcome. In this survey article, we have covered the research work over the last decade to study the genesis of applicabilities of fusion techniques. While we have majorly discussed information, feature, and model fusions for various stock market applications, other aspects associated with financial markets are worth mentioning where fusion techniques are integrated in different ways; such applications can provide different perceptions that may have been explored to a limited extent. To summarize the reviewed fusion-based approaches, we provide a graphical representation of our survey as shown in Fig. 5. The primary three levels of fusion, i.e., information, feature, and model, and respective work for stock market prediction is demonstrated with reference to the referred articles in this survey. Such abstract overview can be helpful to gain clear insights about the existing work; also, it can build a motivational aspect for potential future extensions.

5.1. Other fusion approaches

In the previous sections, different stock-related forecasting applications were covered under specific fusion categories, however, there are other extensions and enhancements on fusion approaches. It has been observed that fusion of prediction results could provide improved predictions [91,92]. Hence, predictions and prediction models were treated as information and sources, respectively, whereas the combinational approach was considered as fusion in Ref. [93]; authors derived the sentiment scores based on such information fusion and analysed the short-term performance of IPOs.

The financial markets are association of a large number of inherent correlations; with a growth of the related trading strategies, the traders' actions also get inline with the market correlations which then get reflected in market dynamics as well [94]. Such correlations are important to study because the fragment of an overall component may result into market crash over a period of time. Based on the prospect theory [95] and price fluctuations being considered as a dynamic process, Ref. [96] proposed to develop a multiparametric analysis framework for decision making in financial investments (MIAMI model) on short time-frames; the approaches including knowledge discovery, technical analysis, information fusion, and soft computing were applied along with fuzzification such that the contributions were fused in energy and entropy decision variables to derive useful trend analysis. While fusion–fission approach was applied to study and predict a market crash [94], a potential integration of such statistical approaches can be desired for higher-level of fusion. Subsequently, a combinational fusion analysis (CFA) was integrated for the task of portfolio management in Ref. [97]; as a combination of selecting appropriate attributes and specific approaches, the portfolio management was proposed with multiple algorithms, instead of single one, based on the diversity or performance strength and the results were given scores for deriving a combinational approach. On the other hand, as an application of individuals' sentiments on the market research, a fusion approach was proposed [98]; while stock market data could be considered as a potential application, authors studied the impact of sentiment analysis-based fusion.

Apart from the traditional fusion, an approach to inject the knowledge to derive the actual information was proposed in study [99]; it was developed as a defusing technique to infer the relevant base data from the aggregated historical data. Another application of fusion was given as a one-vs-one scheme with optimizing decision directed acyclic graph (ODDAG) to predict the listing status of companies [100]. One of the recent applications included the uncertain possibility–probability information fusion for stock selection based on the relative closeness of the alternative stocks [101]. The hyperlinks on Twitter that referred to specific companies in stock market, i.e., cashtags, were integrated with natural language processing (NLP) and data fusion approaches in Ref. [102]; authors prepared the company-specific corpora based on such collected tweets and classified whether the corresponding tweet was related to a particular stock exchange.

The wide range of applications of the extended fusion techniques have a clear indication that the restrictions on exploiting fusion methods should be eliminated and their diversity should be employed for the economic models.

5.2. Potential future directions

The primary focus of our survey has been on the applications of stock markets that are addressed using fusion techniques. These techniques include information, feature, and model fusion; we have included the major applications of stock market. There are certain extensions suggested in respective articles whereas the other potential inferences from the studied approaches can be made for future research directions.

The basic understanding of how fusion can be applied to various domains is an important task. It has been noticed that specification of the specific category of fusion and/or type of fusion technique is given in a limited number of articles; also, the words such as "fusion", "ensemble", and "hybrid" are interchangeably used in some articles and hence, maintaining the principle conceptual meaning of the appropriate terminology becomes a crucial responsibility for future articles. While we have primarily considered the articles having closest explanation about fusion at some point in the prediction phase, other articles that have not been exclusively said to have used fusion, however, a part of it represents fusion approach can be studied and explored.

Various information fusion-based methods have considered quantitative, derived, and appended data. The data sources in a stock market play a critical role; while selection of such sources may differ based on the target application, a combination of static and active data sources [103] can also be carried out on a dynamic level for information fusion. Also, the rationale behind choosing a specific group of data can be advisable for more insights to the approach. One of the important concerns is perspective analysis; while perspective information fusion is applied to various domains [104], the same can be adopted for personalized recommendations on stock market applications. The data fusion is expanded to information fusion, however, an exhaustive extraction can be carried out to prepare a knowledge base and integrate the same for knowledge fusion [105] for stock market.

It can be observed that comparatively a restricted amount of work has been obtained based on feature fusion; though these features come along with the processing step with various attributes, an appropriate fusion of such data can enhance the prediction performance. One of the possibilities of adopting feature fusion can be the usage of image fusion techniques that can derive useful patterns from the stock charts [106]. Subsequently, appropriate combination of information fusion and feature fusion may be helpful in achieving higher-order knowledge base as well as description for the given approach. The model fusion is presented in a broad term; other ways of applying related fusion include classifier fusion, architecture fusion, as well as decision fusion. Such fusion techniques are further classified as per the proposed outline of the article, for example, hierarchy of techniques that combine classifiers [107]. While the classification categories may differ, unification of such wide spread classes and hence, showing a contextual presentation may be a potential future work. Subsequently, the integration of fusion at multiple levels, i.e., information, feature, model, etc. may widen the abilities of an approach to study complex market patterns from diverse perspectives. Also, the balanced approaches integrating economical aspects with computational intelligence can increase the opportunities

Fig. 5. Graphical summary of fusion in stock market prediction.

of reliable predictions; the dependency of such predictions for stock market data must be evaluated for an economic collapse such as the COVID-19 outbreak.

Our survey primarily focuses on various stock market applications addressed using different fusion techniques. It can be observed that the amount of work carried out has potential expansion in terms of fusion of the considered techniques as well as algorithms. The recent advances of neural networks and computing paradigms [108,109] can be consolidated for fusion; on the other hand, the computational intelligence [21,85,110] can also be supplemented with such models to increase the learning capabilities of fusion techniques. The consideration of various financial aspects can redirect the existing work towards a remarkable enhancement; one of such approaches is the financial behaviour which may be further adapted for cognitive analysis and potential stock market predictions. Hence, it is worth mentioning that data fusion, in general, has leveraged stock market predictions and has a prospective future scope.

6. Concluding remarks

The non-linear stock market predictions have been attractive for investors as well as researchers; enormous amount of work has been carried out at various levels where different aspects of market dynamics are explored to make reliable predictions. Amongst such techniques, fusion plays an important role in determining the usefulness of combinational information. To study the significance of fusion in stock market, we conduct an exhaustive survey over a decade, i.e., 2011 – 2020, and review the fusion-based approaches for various stock market applications.

Fusion can take place at diverse components within a prediction process; we elaborate the necessity of applying fusion in stock market and we consider the three kinds, namely, information fusion, feature fusion, and model fusion to cover its major aspects. The target area where fusion is integrated can decide on its impact on the overall prediction performance as well. We also study fusion-based stock market applications such as stock price/trend prediction, index forecasting, portfolio management, risk/return forecasting, and other applications such as company's revenue, market dynamics, EPS prediction, etc. To provide a general overview of fusion categories and potential stock market applications, we build an infographic presentation along with different fusion classification techniques; it has been observed that a formal classification of the proposed fusion-based approach is provided in limited articles. Therefore, classification-based categorization and characteristics of the articles may be challenging. For the completeness of this survey, we discuss the other potential fusion techniques that have been applied to the financial market, followed by the potential future directions based on fusion in stock market. To the best of our knowledge, our survey is the primary work to review the recent advances and implications of fusion in stock market prediction.

CRediT authorship contribution statement

Ankit Thakkar: Conceptualization of the idea, Flow of the paper, Writing - original draft, Writing and editing of the revised manuscript. **Kinjal Chaudhari:** Conceptualization of the idea, Flow of the paper, Writing - original draft, Writing and editing of the revised manuscript.

Declaration of competing interest

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

References

[1] M. Van Rooij, A. Lusardi, R. Alessie, Financial literacy and stock market participation, J. Financ. Econ. 101 (2) (2011) 449–472.
[2] G. Bonanno, F. Lillo, R.N. Mantegna, Levels of complexity in financial markets, Physica A 299 (1–2) (2001) 16–27.
[3] A. Thakkar, K. Chaudhari, CREST: Cross-reference to exchange-based stock trend prediction using long short-term memory, Procedia Comput. Sci. 167 (2020) 616–625.
[4] C. Keller, M. Siegrist, Investing in stocks: The influence of financial risk attitude and values-related money and stock market attitudes, J. Econ. Psychol. 27 (2) (2006) 285–303.
[5] R.C. Cavalcante, R.C. Brasileiro, V.L. Souza, J.P. Nobrega, A.L. Oliveira, Computational intelligence and financial markets: A survey and future directions, Expert Syst. Appl. 55 (2016) 194–211.
[6] H.-Y. Chen, C.-F. Lee, W.K. Shih, Technical, fundamental, and combined information for separating winners from losers, Pac.-Basin Finance J. 39 (2016) 224–242.
[7] F. Castanedo, A review of data fusion techniques, Sci. World J. 2013 (2013).
[8] P. Kazienko, E. Lughofer, B. Trawinski, Editorial on the Special Issue "Hybrid and Ensemble Techniques in Soft Computing: Recent Advances and Emerging Trends", Springer, 2015.
[9] A. Gabriela ğiĠan, The efficient market hypothesis: Review of specialized literature and empirical research, Procedia Econ. Finance 32 (2015) 442–449.
[10] A. Shleifer, Inefficient Markets: An Introduction to Behavioural Finance, OUP Oxford, 2000.
[11] Z. Copur, Handbook of Research on Behavioral Finance and Investment Strategies: Decision Making in the Financial Industry: Decision Making in the Financial Industry, IGI Global, 2015.
[12] S. Khuntia, J. Pattanayak, Adaptive market hypothesis and evolving predictability of bitcoin, Econom. Lett. 167 (2018) 26–28.
[13] A.W. Lo, Reconciling efficient markets with behavioral finance: the adaptive markets hypothesis, J. Invest. Consult. 7 (2) (2005) 21–44.
[14] J.H. Kim, A. Shamsuddin, K.-P. Lim, Stock return predictability and the adaptive markets hypothesis: evidence from century-long US data, J. Empir. Financ. 18 (5) (2011) 868–879.
[15] A. Urquhart, F. McGroarty, Are stock markets really efficient? Evidence of the adaptive market hypothesis, Int. Rev. Financ. Anal. 47 (2016) 39–49.
[16] A. Charles, O. Darné, J.H. Kim, Exchange-rate return predictability and the adaptive markets hypothesis: Evidence from major foreign exchange rates, J. Int. Money Finance 31 (6) (2012) 1607–1626.
[17] J. Chu, Y. Zhang, S. Chan, The adaptive market hypothesis in the high frequency cryptocurrency market, Int. Rev. Financ. Anal. 64 (2019) 221–231.
[18] S. Kute, S. Tamhankar, A survey on stock market prediction techniques, Int. J. Sci. Res. (2013).
[19] M. Woźniak, M. Graña, E. Corchado, A survey of multiple classifier systems as hybrid systems, Inf. Fusion 16 (2014) 3–17.
[20] J.A. Balazs, J.D. Velásquez, Opinion mining and information fusion: a survey, Inf. Fusion 27 (2016) 95–110.
[21] I. Ibidapo, A. Adebiyi, O. Okesola, Soft computing techniques for stock market prediction: A literature survey, Covenant J. Inform. Commun. Technol. 5 (2) (2017).
[22] B.P.L. Lau, S.H. Marakkalage, Y. Zhou, N.U. Hassan, C. Yuen, M. Zhang, U.-X. Tan, A survey of data fusion in smart city applications, Inf. Fusion 52 (2019) 357–374.
[23] X. Zhao, Y. Jia, A. Li, R. Jiang, H. Xie, Y. Song, W. Han, Multi-source knowledge fusion: A survey, in: 2019 IEEE Fourth International Conference on Data Science in Cyberspace (DSC), IEEE, 2019, pp. 119–127.
[24] F. Rundo, F. Trenta, A.L. di Stallo, S. Battiato, Machine learning for quantitative finance applications: A survey, Appl. Sci. 9 (24) (2019) 5574.
[25] A.M. Ozbayoglu, M.U. Gudelek, O.B. Sezer, Deep learning for financial applications: A survey, 2020, arXiv preprint arXiv:2002.05786.
[26] H.F. Durrant-Whyte, Sensor models and multisensor integration, in: Autonomous Robot Vehicles, Springer, 1990, pp. 73–89.
[27] H. Haleh, B.A. Moghaddam, S. Ebrahimijam, A new approach to forecasting stock price with EKF data fusion, Int. J. Trade Econ. Finance 2 (2) (2011) 109.
[28] W.A. Rosenkrantz, Why Stock Prices Have a Lognormal Distribution, Department of Mathematics and Statistics, University of Massachusetts at Amhers, 2003.
[29] M.J. Sharpe, Lognormal Model for Stock Prices, University of California, San Diego, San Diego, California, 2004.
[30] B.A. Moghaddam, H. Haleh, S. Ebrahimijam, Forecasting trend and stock price with adaptive extended kalman filter data fusion, in: Proceedings of IEEE International Conference on Economics and Finance Research, 2011, pp. 119–123.
[31] B. Weng, M.A. Ahmed, F.M. Megahed, Stock market one-day ahead movement prediction using disparate data sources, Expert Syst. Appl. 79 (2017) 153–163.
[32] X. Zhang, Y. Zhang, S. Wang, Y. Yao, B. Fang, S.Y. Philip, Improving stock market prediction via heterogeneous information fusion, Knowl.-Based Syst. 143 (2018) 236–247.
[33] J. Huang, Y. Zhang, J. Zhang, X. Zhang, A tensor-based sub-mode coordinate algorithm for stock prediction, in: 2018 IEEE Third International Conference on Data Science in Cyberspace (DSC), IEEE, 2018, pp. 716–721.
[34] J. Park, H. Leung, K. Ma, Information fusion of stock prices and sentiment in social media using granger causality, in: 2017 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI), IEEE, 2017, pp. 614–619.
[35] J. Park, K. Ma, H. Leung, Prediction of stock prices with sentiment fusion and SVM granger causality, in: 2019 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech), IEEE, 2019, pp. 207–214.
[36] J. Long, Z. Chen, W. He, T. Wu, J. Ren, 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. (2020) 106205.
[37] C. Lundblad, The risk return tradeoff in the long run: 1836–2003, J. Financ. Econ. 85 (1) (2007) 123–150.
[38] S.T. Rachev, S. Mittnik, Stable Paretian Models in Finance, Vol. 7, Wiley, 2000.
[39] S.T. Rachev, C. Menn, F.J. Fabozzi, Fat-Tailed and Skewed Asset Return Distributions: Implications for Risk Management, Portfolio Selection, and Option Pricing, Vol. 139, John Wiley & Sons, 2005.
[40] L. Zhang, L. Zhang, W. Teng, Y. Chen, Based on information fusion technique with data mining in the application of finance early-warning, Procedia Comput. Sci. 17 (2013) 695–703.
[41] X. Li, X. Huang, X. Deng, S. Zhu, Enhancing quantitative intra-day stock return prediction by integrating both market news and stock prices information, Neurocomputing 142 (2014) 228–238.
[42] H. He, A. Subramanian, S. Choi, P.K. Varshney, T. Damarla, Social media data assisted inference with application to stock prediction, in: 2015 49th Asilomar Conference on Signals, Systems and Computers, IEEE, 2015, pp. 1801–1805.
[43] G. Yi, H. Lei, L. Ziqiang, Port customer credit risk prediction based on internal and external information fusion, Open Cybern. Syst. J. 9 (1) (2015).
[44] A. Oztekin, R. Kizilaslan, S. Freund, A. Iseri, A data analytic approach to forecasting daily stock returns in an emerging market, European J. Oper. Res. 253 (3) (2016) 697–710.
[45] S. Feuerriegel, J. Gordon, Long-term stock index forecasting based on text mining of regulatory disclosures, Decis. Support Syst. 112 (2018) 88–97.
[46] K. Teymourian, M. Rohde, A. Paschke, Fusion of background knowledge and streams of events, in: Proceedings of the 6th ACM International Conference on Distributed Event-Based Systems, 2012, pp. 302–313.
[47] K. Teymourian, M. Rohde, A. Paschke, Knowledge-based processing of complex stock market events, in: Proceedings of the 15th International Conference on Extending Database Technology, 2012, pp. 594–597.
[48] Y. Xu, L. Wu, X. Wu, Z. Xu, Belief fusion of predictions of industries in China's stock market, in: International Conference on Belief Functions, Springer, 2014, pp. 348–355.
[49] K. Chang, Z. Tian, Market analysis and trading strategies with Bayesian networks, in: 2015 18th International Conference on Information Fusion (Fusion), IEEE, 2015, pp. 1922–1929.
[50] X. Shao, B. He, J. Guo, T. Yan, The application of AUV navigation based on adaptive extended Kalman filter, in: Oceans 2016-Shanghai, IEEE, 2016, pp. 1–4.
[51] L. Rokach, Decision forest: Twenty years of research, Inf. Fusion 27 (2016) 111–125.
[52] S. Karamizadeh, S.M. Abdullah, M. Halimi, J. Shayan, M. javad Rajabi, Advantage and drawback of support vector machine functionality, in: 2014 International Conference on Computer, Communications, and Control Technology (I4CT), IEEE, 2014, pp. 63–65.
[53] M. Qiu, Y. Song, Predicting the direction of stock market index movement using an optimized artificial neural network model, PLoS One 11 (5) (2016) e0155133.
[54] P. Du, S. Liu, J. Xia, Y. Zhao, Information fusion techniques for change detection from multi-temporal remote sensing images, Inf. Fusion 14 (1) (2013) 19–27.
[55] Z. Guo, H. Wang, Q. Liu, J. Yang, A feature fusion based forecasting model for financial time series, PLoS One 9 (6) (2014).

[56] G. Kaur, J. Dhar, R.K. Guha, Minimal variability OWA operator combining ANFIS and fuzzy c-means for forecasting BSE index, Math. Comput. Simulation 122 (2016) 69–80.
[57] T. Kim, H.Y. Kim, Forecasting stock prices with a feature fusion LSTM-CNN model using different representations of the same data, PLoS One 14 (2) (2019).
[58] A. Picasso, S. Merello, Y. Ma, L. Oneto, E. Cambria, Technical analysis and sentiment embeddings for market trend prediction, Expert Syst. Appl. 135 (2019) 60–70.
[59] A. Ang, D. Papanikolaou, M.M. Westerfield, Portfolio choice with illiquid assets, Manage. Sci. 60 (11) (2014) 2737–2761.
[60] J. Irukulapati, D.F. Hsu, C. Schweikert, Long-term portfolio management using attribute selection and combinatorial fusion, in: 2018 IEEE 17th International Conference on Cognitive Informatics & Cognitive Computing (ICCI* CC), IEEE, 2018, pp. 593–599.
[61] M. Bedworth, J. O'Brien, The omnibus model: a new model of data fusion?, IEEE Aerosp. Electron. Syst. Mag. 15 (4) (2000) 30–36.
[62] O. Arandjelović, Discriminative extended canonical correlation analysis for pattern set matching, Mach. Learn. 94 (3) (2014) 353–370.
[63] Q. He, Z. Cai, W. Liu, Prediction of listed companies' revenue based on model-fused, in: 2019 International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI), IEEE, 2019, pp. 257–260.
[64] M. Gunasekaran, K. Ramaswami, A fusion model integrating ANFIS and artificial immune algorithm for forecasting Indian stock market, J. Appl. Sci. 11 (16) (2011) 3028–3033.
[65] N. Homayouni, A. Amiri, Stock price prediction using a fusion model of wavelet, fuzzy logic and ANN, in: International Conference on E-Business, Management and Economics, Vol. 25, 2011, pp. 277–281.
[66] C.-H. Cheng, L.-Y. Wei, J.-W. Liu, T.-L. Chen, OWA-based ANFIS model for TAIEX forecasting, Econ. Model. 30 (2013) 442–448.
[67] J.L. Ticknor, A Bayesian regularized artificial neural network for stock market forecasting, Expert Syst. Appl. 40 (14) (2013) 5501–5506.
[68] M.R. Hassan, B. Nath, M. Kirley, A fusion model of HMM, ANN and GA for stock market forecasting, Expert Syst. Appl. 33 (1) (2007) 171–180.
[69] J. Patel, S. Shah, P. Thakkar, K. Kotecha, Predicting stock market index using fusion of machine learning techniques, Expert Syst. Appl. 42 (4) (2015) 2162–2172.
[70] K. Jhaveri, D. Shah, S. Bhanushali, E. Johri, Financial market prediction using hybridized neural approach, in: 2016 International Conference on Computation of Power, Energy Information and Commuincation (ICCPEIC), IEEE, 2016, pp. 009–014.
[71] L. Zhang, X. Fan, C. Xu, A fusion financial prediction strategy based on RNN and representative pattern discovery, in: 2017 18th International Conference on Parallel and Distributed Computing, Applications and Technologies (PDCAT), IEEE, 2017, pp. 92–97.
[72] Q. Wang, W. Xu, H. Zheng, Combining the wisdom of crowds and technical analysis for financial market prediction using deep random subspace ensembles, Neurocomputing 299 (2018) 51–61.
[73] S. Mohamad, T. Hassan, Z.M. Sori, Diversification across economic sectors and implication on portfolio investments in Malaysia, Int. J. Econ. Manage. 1 (1) (2006) 155–172.
[74] J. Driessen, L. Laeven, International portfolio diversification benefits: Cross-country evidence from a local perspective, J. Bank. Financ. 31 (6) (2007) 1693–1712.
[75] Y.S. Kim, K.J. Park, O.B. Kwon, Geopolitical risk and trading patterns of foreign and domestic investors: Evidence from Korea, Asia-Pac. J. Financ. Stud. 48 (2) (2019) 269–298.
[76] A.K. Chaudhary, Impact of behavioral finance in investment decisions and strategies–a fresh approach, Int. J. Manage. Res. Bus. Strategy 2 (2) (2013) 85–92.
[77] M.W. Brandt, Portfolio choice problems, in: Handbook of Financial Econometrics: Tools and Techniques, Elsevier, 2010, pp. 269–336.
[78] F.D. Paiva, R.T.N. Cardoso, G.P. Hanaoka, W.M. Duarte, Decision-making for financial trading: A fusion approach of machine learning and portfolio selection, Expert Syst. Appl. 115 (2019) 635–655.
[79] T.-C. Wu, M.-F. Hsu, Credit risk assessment and decision making by a fusion approach, Knowl.-Based Syst. 35 (2012) 102–110.
[80] S. Barak, A. Arjmand, S. Ortobelli, Fusion of multiple diverse predictors in stock market, Inf. Fusion 36 (2017) 90–102.
[81] H. Harnovinsah, S. Alamsyah, The mediation influence of value relevance of accounting information, investment decision and dividend policy on the relationship between profitability and the company's value, J. Akunt. 21 (2) (2017) 170–183.
[82] L.-Y. Wei, A fusion ANFIS model for forecasting EPS of leading industries in Taiwan, in: 2011 International Conference on Machine Learning and Cybernetics, Vol. 1, IEEE, 2011, pp. 1–4.
[83] L.-Y. Wei, C.-H. Cheng, H.-H. Wu, Fusion anfis model based on ar for forecasting eps of leading industries, Int. J. Innovative Comput. Inf. Control 7 (9) (2011) 5445–5458.
[84] W. Xiaofang, W. Jin, L. Haichuan, L. Shulin, L. Hongjiang, Research and implementation of SVM and bootstrapping fusion algorithm in emotion analysis of stock review texts, DEStech Trans. Comput. Sci. Eng. (iccis) (2019).
[85] A. Thakkar, K. Chaudhari, A comprehensive survey on portfolio optimization, stock price and trend prediction using particle swarm optimization, Arch. Comput. Methods Eng. (2020) http://dx.doi.org/10.1007/s11831-020-09448-8.
[86] G.S. Atsalakis, K.P. Valavanis, Surveying stock market forecasting techniques–Part II: Soft computing methods, Expert Syst. Appl. 36 (3) (2009) 5932–5941.
[87] M.M. Azeem, A. Mohammad, An analysis of applications and possibilities of neural networks (fuzzy, logic and genetic algorithm) in finance and accounting, Donnish J. Bus. Finance Manage. Res. 1 (2) (2015) 9–18.
[88] D. Giannone, M. Lenza, D. Momferatou, L. Onorante, Short-term inflation projections: A Bayesian vector autoregressive approach, Int. J. Forecast. 30 (3) (2014) 635–644.
[89] M.N.M. Salleh, N. Talpur, K. Hussain, Adaptive neuro-fuzzy inference system: Overview, strengths, limitations, and solutions, in: International Conference on Data Mining and Big Data, Springer, 2017, pp. 527–535.
[90] T. Salles, M. Gonçalves, V. Rodrigues, L. Rocha, Improving random forests by neighborhood projection for effective text classification, Inf. Syst. 77 (2018) 1–21.
[91] D. Delen, A comparative analysis of machine learning techniques for student retention management, Decis. Support Syst. 49 (4) (2010) 498–506.
[92] C.M. Fuller, D.P. Biros, D. Delen, An investigation of data and text mining methods for real world deception detection, Expert Syst. Appl. 38 (7) (2011) 8392–8398.
[93] E. Bastı, C. Kuzey, D. Delen, Analyzing initial public offerings' short-term performance using decision trees and SVMs, Decis. Support Syst. 73 (2015) 15–27.
[94] B.K. Teh, S.A. Cheong, The asian correction can be quantitatively forecasted using a statistical model of fusion-fission processes, PLoS One 11 (10) (2016).
[95] D. Kahneman, Prospect theory: An analysis of decisions under risk, Econometrica 47 (1979) 278.
[96] G. Iovane, A. Amorosia, M. Leone, M. Nappi, G. Tortora, Multi indicator approach via mathematical inference for price dynamics in information fusion context, Inform. Sci. 373 (2016) 183–199.
[97] Y. Luo, B.S. Kristal, C. Schweikert, D.F. Hsu, Combining multiple algorithms for portfolio management using combinatorial fusion, in: 2017 IEEE 16th International Conference on Cognitive Informatics & Cognitive Computing (ICCI* CC), IEEE, 2017, pp. 361–366.
[98] S. Das, A. Das, Fusion with sentiment scores for market research, in: 2016 19th International Conference on Information Fusion (FUSION), IEEE, 2016, pp. 1003–1010.
[99] Z. Liu, H.A. Song, V. Zadorozhny, C. Faloutsos, N. Sidiropoulos, H-fuse: Efficient fusion of aggregated historical data, in: Proceedings of the 2017 SIAM International Conference on Data Mining, SIAM, 2017, pp. 786–794.
[100] L. Zhou, Q. Wang, H. Fujita, One versus one multi-class classification fusion using optimizing decision directed acyclic graph for predicting listing status of companies, Inf. Fusion 36 (2017) 80–89.
[101] X. Sang, Y. Zhou, X. Yu, An uncertain possibility-probability information fusion method under interval type-2 fuzzy environment and its application in stock selection, Inform. Sci. 504 (2019) 546–560.
[102] L. Evans, M. Owda, K. Crockett, A.F. Vilas, A methodology for the resolution of cashtag collisions on twitter–a natural language processing & data fusion approach, Expert Syst. Appl. 127 (2019) 353–369.
[103] G. Nachouki, M. Quafafou, Multi-data source fusion, Inf. Fusion 9 (4) (2008) 523–537.
[104] G. Shroff, P. Agarwal, K. Singh, A.H. Kazmi, S. Shah, A. Sardeshmukh, Prescriptive information fusion, in: 17th International Conference on Information Fusion (FUSION), IEEE, 2014, pp. 1–8.
[105] X.L. Dong, E. Gabrilovich, G. Heitz, W. Horn, K. Murphy, S. Sun, W. Zhang, From data fusion to knowledge fusion, 2015, arXiv preprint arXiv:1503.00302.
[106] Q.-S. Sun, S.-G. Zeng, Y. Liu, P.-A. Heng, D.-S. Xia, A new method of feature fusion and its application in image recognition, Pattern Recognit. 38 (12) (2005) 2437–2448.
[107] U.G. Mangai, S. Samanta, S. Das, P.R. Chowdhury, A survey of decision fusion and feature fusion strategies for pattern classification, IETE Tech. Rev. 27 (4) (2010) 293–307.
[108] V. Rajput, S. Bobde, Stock market forecasting techniques: literature survey, Int. J. Comput. Sci. Mobile Comput. 5 (6) (2016) 500–506.
[109] R. Sutkatti, D. Torse, Stock market forecasting techniques: A survey, Int. Res. J. Eng. Technol. 6 (5) (2019).
[110] G. Kumar, S. Jain, U.P. Singh, Stock market forecasting using computational intelligence: A survey, Arch. Comput. Methods Eng. (2020) 1–33.