Contents lists available at www.journal.unipdu.ac.id

Register

Journal Page is available to www.journal.unipdu.ac.id/index.php/register

Research article

An empirical study on the various stock market prediction methods

Jaymit Bharatbhai Pandya [a,*], *Udesang K. Jaliya* [b]

[a,b] *Gujarat Technological University, Gujarat, India*
[b] *Department of Computer Engineering, Birla Vishvakarma Mahavidyalaya Engineering College, Gujarat, India*
email: [a,*] erpandyajaymit@gmail.com
* *Correspondence*

ARTICLE INFO

Article history:
Received 26 August 2021
Revised 10 October 2021
Accepted 18 November 2021
Available online 23 March 2022

Keywords:
Clustering and Data Mining
CNN
Sentiment Analysis
Stock Market Prediction
Time Series Analysis

Please cite this article in IEEE style as:
J. B. Pandya and U. K. Jaliya, "An empirical study on the various stock market prediction methods," *Register: Jurnal Ilmiah Teknologi Sistem Informasi*, vol. 8, no. 1, pp. 58-80, 2022.

ABSTRACT

Investment in the stock market is one of the much-admired investment actions. However, prediction of the stock market has remained a hard task because of the non-linearity exhibited. The non-linearity is due to multiple affecting factors such as global economy, political situations, sector performance, economic numbers, foreign institution investment, domestic institution investment, and so on. A proper set of such representative factors must be analyzed to make an efficient prediction model. Marginal improvement of prediction accuracy can be gainful for investors. This review provides a detailed analysis of research papers presenting stock market prediction techniques. These techniques are assessed in the time series analysis and sentiment analysis section. A detailed discussion on research gaps and issues is presented. The reviewed articles are analyzed based on the use of prediction techniques, optimization algorithms, feature selection methods, datasets, toolset, evaluation matrices, and input parameters. The techniques are further investigated to analyze relations of prediction methods with feature selection algorithm, datasets, feature selection methods, and input parameters. In addition, major problems raised in the present techniques are also discussed. This survey will provide researchers with deeper insight into various aspects of current stock market prediction methods.

1. Introduction

Various common behaviors, like economic activities, climate indices, or energy data whose temporal characteristics are an important source for making the prediction [1]. Among them, stock market prediction is the most important performance of economical firms and the personal depositors for making the investment resolutions. The stock market is an evolutionary, non-linear dynamic system, and its prediction is considered a difficult job [2]. The main intention of the stock market identification is to identify the potential rate of company stock deals on exchange. Dependable identification of future stock prices has able to acquire important profits [3]. Generally, the prediction movement is divided into three terms, namely short, medium, and long. The predicting period may be within a few minutes, hours, or days in a week. The predicting period within one week to month denotes the medium term, and the period of one to several years represents the long term [4].

Several varieties of prediction methods using soft computing approaches were developed to enhance the identification accuracy in stock market prediction [4]. For the time series modeling, there are various techniques developed. The traditional statistical methods, Auto Regressive Integrated Moving Average (ARIMA) [5], Generalized Auto Regressive Conditional Heteroskedasticity (GARCH) [58], and the moving averages are linear within their forecasting of future values. Hence, in time series prediction, the applicability of various machine learning techniques [6] and deep learning techniques

[7], along with their ability to extract features [7], are tested extensively by various researchers. Among them, the most researched techniques are the Support Vector Machine (SVM), and other familiar approaches are Neural Networks [2, 8], Decision Trees [2, 9] and Random Forests [10], Logistic Regression [2, 10, 11] Discriminant Analysis [2, 12] and K-Nearest Neighbor [2, 13]. Different approaches apart from prediction have been used. Kim [14] tested various data mining tools to identify the group of parameters providing the highest accuracy. Leung., et al. [15] optimized level estimation methods, minor forecasting error does not translate into gain or loss. Many methods of different areas have been adapted and have been suggested to be used in time series forecasting. Several such methods are short term electricity price prediction using hybrid evolutionary approach [16], predicting tumor using threshold segmentation in brain images [17], intelligent and ensemble learning for option prediction [18], prediction of option price for uncertain stock models [19], option price prediction considering market sentiment and information [20] and calculating returns from options trading based on discontinuity in spot prices [21]. Several forecasting methods used optimization algorithms such as Genetic Algorithm (GA) [22], Cuckoo Search [4], Firefly Optimization [23], Auditory Algorithm [24]. There are many modified algorithms such as CMS-PSO [25], Self-regulating PSO [26], static and adaptive mutation for Genetic Algorithm [22, 27], and nature-inspired algorithms such as Rider Based Optimization [28], Elephant Herd Optimization [29], Moth Algorithm [30]. The applicability of these algorithms for optimizing neural networks of time series analysis is yet to be tested.

Several surveys on stock market prediction methods have been conducted. Hu., et al. [31] surveyed several methods to predict stocks and forex based on historical time series data. The authors have noted, stocks and forex have similar characteristics and hence, analyzed the effects of deep learning methods such as CNN, LSTM, DNN, and RNN. Authors have surveyed papers only from DBLP database. Shah., et al. [32] have presented a taxonomy of stock market prediction methods published between 2008 to 2016. The survey was aimed to find methods that are commonly applied to stock market prediction and challenges in the prediction process. Gandhmal and Kumar [33] surveyed different machine learning approaches such as ANN, SVM, Naïve Bayes, RNN, Fuzzy method, K-means, filtering methods, for stock market prediction using historical data. These methods have been categorized as prediction methods and classification methods. The authors also discussed the limitations and significance of the methods. Thakkar and Chaudhari [34] evaluated the applicability of DNN on temporal stock market data. Authors focused on RNN, CNN, DNN, and their variations. The authors concluded that DNN has better prediction capability. However, DNN can be used effectively only when hyperparameters of the network are properly controlled otherwise, it may affect results negatively. Rao., et al. [35] surveyed methods such as ARIMA, time series linear model, RNN, Hidden Markov Model, SVM, and ANN. These techniques were categorized based on their usefulness for prediction and optimization purposes.

This paper investigates different existing stock market prediction techniques. We mainly focus on modern techniques, accentuating their limitations and significance. Methods having historical time series data and sentimental data as input are considered. The techniques are categorized into time series analysis-based techniques and sentiment analysis-based techniques. We discuss research gaps and issues found in prediction techniques. This survey analyses techniques based on toolset used, datasets used, categorization of methods, performance matrices used, and input attributes used. Therefore, this review is useful for the development of effective stock market prediction techniques.

This survey paper is arranged as following: Section 2 elaborates reviews of stock market prediction, and Section 3 illustrates the research gaps with the issues. Section 4 describes the study of the approaches using the utilized datasets, performance metrics, along with finally concludes the paper in Section 5.

2. Literature Survey

The research works adopting different approaches developed for stock market predictions are deliberated in this section. These are categorized into time series analysis-based techniques and sentiment analysis-based techniques. The time series analysis-based techniques are further divided into classical machine learning algorithm-based techniques, CNN and RNN based techniques, evolutionary learning-based techniques, and other stock market prediction techniques. The challenges correlated with these approaches are evaluated to inspire the researchers to develop novel stock prediction techniques.

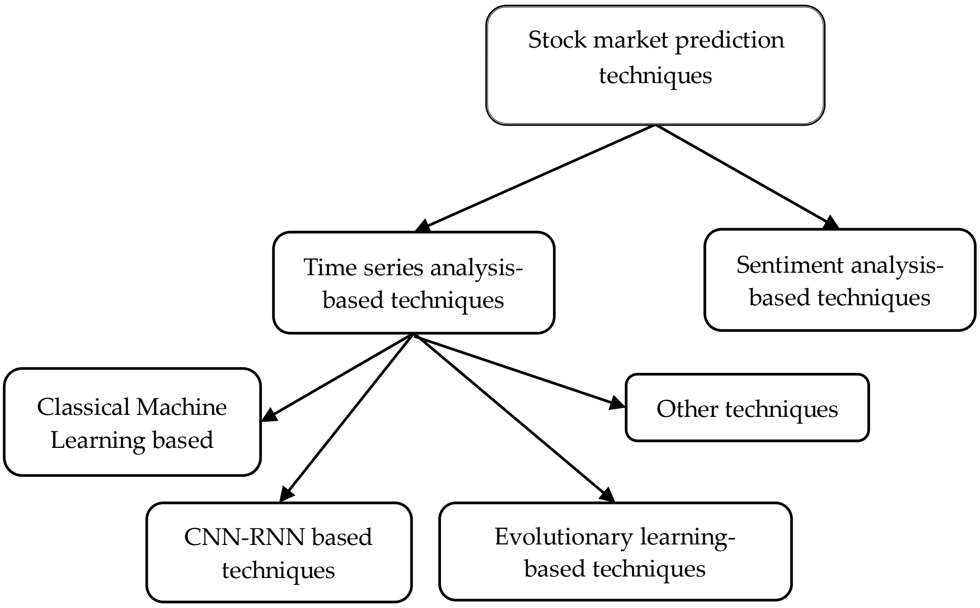
Stock market prediction techniques
Time series analysis-based techniques
Sentiment analysis-based techniques
Classical Machine Learning based
Other techniques
CNN-RNN based techniques
Evolutionary learning-based techniques

Fig. 1. Categorization of the stock market prediction

The research works considering several approaches used for the stock market predictions. Major works take both or one of the following data as input. 1. Qualitative data. 2. Quantitative data. Qualitative data generally consist of news, sentiments from social networking sites, reports of companies, articles, etc [36]. Quantitative data consist of historical time series numbers, indicators extracted from time series data, etc [36]. Hence, we have categorized our surveyed methods into time series analysis based techniques and sentiment analysis based techniques [36]. The time series analysis-based techniques are further categorized based on the used methods. The initial approach to assess and predict the stock market was through classical techniques like Linear Regression, Support Vector Machine, Nearest Neighbour, Naïve Bayes classifier, K-means clustering, etc. These approaches are covered in classical machine learning algorithms-based techniques. We identified that recent models combining CNN, RNN, or LSTM with other models are widely researched. Hence, we have categorized them separately. The values of hyperparameters must be optimized to achieve good prediction [34], which leads to the need of evolutionary algorithms for controlling hyperparameters. These techniques are categorized into evolutionary learning-based techniques. Various researchers have adopted various unconventional approaches for stock market predictions. These approaches are covered in other stock market prediction techniques. Fig. 1 shows this categorization.

Another perspective to observe recent researched models is ensemble learning. Ensemble techniques combine several base models in order to produce one optimal predictive model. Such methods are robust in nature. The stock market prediction mainly requires feature selection from raw data, training the neural network, and controlling hyperparameters. All three tasks require the use of different specialized algorithms and their integration. The majority of the techniques are analyzed in the next section and termed to follow ensemble learning methodology as they integrate multiple techniques to increase prediction accuracy.

2.1. Time series analysis-based techniques

a) Classical machine learning algorithms-based techniques

The researches that utilized classical machine learning-based approaches are discussed in this section:

Babu., et al. [37] presented an effective clustering algorithm named Recursive K-means clustering (HRK) and Hierarchical agglomerative for the stock market prediction. This method has three stages for effective stock market prediction. Initially, every financial statement was transformed into feature vectors. The hierarchical agglomerative clustering approach was utilized, segmenting the transformed feature vectors to the clusters. Then, for every cluster, K-means clustering approach was applied for dividing every cluster into the sub-clusters. Therefore most feature vector in every sub-clusters belongs to the similar class. After that, for every sub-clusters, its centroid was selected as the representative feature vector. At last, to predict the stock price movements, representative feature selection was used.

Gupta and Sharma [38] developed a K-means clustering technique for stock market identification. In their research, to predict the stock market effectively, a hybrid combinatorial approach was used. Initially, the K-means clustering technique was utilized for clustering datasets, a horizontal partition-based decision tree was applied for classifying the clustered data. This developed method gives an accurate identification of the real value, so results obtained from this method were more effective and precise.

Wang., et al. [39] modeled a sparse representation-based technique for the stock market prediction. The sparse representation method was applied for smoothing the time series. After that, fuzzy tools were utilized for converting the time series to the fuzzy series, and the weights were calculated using the number of occurrences of the fuzzy sets. At last, values of time series were predicted using the inverse representation of sparse representation and the weights.

Chen and Hao [40] introduced feature weighted SVM approach for the stock market identification. Initially, for estimating the virtual significance of every feature and for locating weights of every feature, the information gain was determined. Then, the weights were used for computing the inner product during the kernel functions in the SVM to forecast stock indices price movement. After that, the computed weights were used in calculating the Euclidean distance in the neural network for predicting the stock prices. Some features, such as closing price, entire volume on the stock market catalogs, historical highest cost, opening cost, and lowest cost was selected for input features.

Barak and Modarres [41] developed a data mining technique for effective stock market prediction. The advantage and disadvantages were considered by analyzing the size and the leaves of the tree approaches. Then, for the selected features, the risk and the return were predicted. Lastly, the hybrid technique based on a filter with the function-based clustering was employed that characteristics chosen were the best indicators for the return and risk process. The outcome shows the proposed hybrid approach was the effective feature selection method and the effective prediction approach.

b) CNN and RNN-based techniques

The researches that utilized convolution neural network and recurrent neural network-based approaches are discussed in this section:

The motif-based sequence reconstruction along with CNN was introduced by Wen., et al. [42] for the stock market prediction. Here, the motif-based sequence reconstruction was adopted for decreasing the noisy-filled financial temporal series, and the neural network was used for capturing the spatial arrangement of the time series. Then, the vocabulary-based method and the modified dynamic time warping were also utilized for obtaining a group of motifs for each and every time series. The basic patterns present in the reconstructed sequence were learned with the help of CNN, which offers needed information for the downs and ups forecast.

The stock price trend prediction method using the encoder structure was developed by Chen., et al. [43] for the stock market prediction. Here, the dual feature extraction model based on various time duration was adopted, which has the ability to extract the underlying market data effectively. The CNN and the piecewise linear regression model were developed for extracting the long-term temporal features and also the short-term market features of financial time series in the various time durations. The performance of the prediction was improved by portraying the stock market information with dual features. The performance outcome shows that the developed trend predictive model achieves good generalization capability and the stock market prediction capability.

Long., et al. [44] developed a deep neural network method for the prediction of the stock market. The node to vector method was developed for attaining the vector representation of company nodes. Then, cluster the investors based on the trading behaviors and also attains the investor clusters for the additional examination of their trading models. After that, the buying volume matrix, selling volume matrix, and the transaction number matrix were generated using trading data, and obtained data were subjected to the CNN for deep feature extraction. A deep stock-trend prediction neural network was developed for predicting the stock market effectively.

Hoseinzade and Haratizadeh [45] introduced a CNN method for the stock market identification. The convolutional operation was utilized for calculating the alterations caused by applying the filter on input data. The size of the filter displays coverage of the specific filter, and every filter used the shared set of weights for executing the convolutional process. Then, the pooling layer was dependable for sub-sampling data, and it helped to decrease the computational cost and manage the overfitting issues. The two differences of the deep CNN were developed as well as applied for extracting the higher-level features from the rich set of original variables.

Baek and Kim [46] developed a data augmentation approach for the prediction of the stock market. The developed ModAugNet structure contains two parts, namely the prediction LSTM part and the overfitting-prevention LSTM part. The prediction LSTM part was named a prediction module, whereas the overfitting-prevention LSTM part is termed as prevention segment. In this case, the prediction part considered only the target index as an input, whereas the prevention segment considered the various combinations of related stocks every time. After that, various occurrences were generated and the method was trained by enlarged quantity of real data points. Finally, the stock market prediction was composed by using a unified network of equally trained prevention part and prediction part.

Hiransha., et al. [47] modeled a deep learning approach for stock market identification. The four methods of deep learning approaches, like Recurrent Neural Network (RNN), CNN, MLP, and LSTM were utilized in this developed method. In this approach, the various two leading stock market was considered for identification. The stock market was trained along with four utilized deep learning approaches. Conversely, the outcome of the developed method shows that CNN was the best stock market prediction approach.

Zhou., et al. [48] presented Generative Adversarial Network-Frequency Data (GAN-FD) to identify the stock market. This method utilized the Long Short-Term Memory as well as the Convolutional Neural Network for identifying high-frequency stock market. The input was taken from the freely obtainable index offered by the trading software for avoiding the complex technical and financial theory analysis. This method imitates the trading form of authenticated trader also utilizes the technique of rolling partition training set. The training set was used for analyzing the prediction performance of the developed model. Using deep learning architecture, the method attained the prediction capability better than the other standard methods. The direction prediction loss was reduced by adversarial training method.

Nelson., et al. [49] presented a LSTM neural network to predict stock market prediction. The data was collected in time series for identifying the stock market. A log return transformation was utilized by means of normalization for stabilizing the variance and mean. The network performance was validated using the performance metrics and the financial outcomes were gathered and compared with the baseline approaches. Finally, the statistical test was conducted to determine the important improvements.

c) Evolutionary learning-based techniques

The researches that utilized evolutionary learning-based approaches are discussed in this section:

Tsai and Hsiao [50] developed a technique for stock market identification by combining numerous feature selection methods. The multiple feature selection approaches are joined for the identification of more delegate variables for good prediction. The three familiar feature selection technique was joined, decision trees (CART), Principal Component Analysis (PCA), and GA. The combination methods were used to sort out the unreliable variables based on the multi-intersection, union, and intersection schemes. The Back Propagation Neural Network (BPNN) method was utilized for the prediction model. This combined feature selection method has the ability to provide high accuracy and low errors than the single feature selection method. For analyzing the level of important variance of prediction accuracy based on the various feature selection approaches, a t-test was used.

Asadi., et al. [51] modeled a hybrid intelligent method, termed Pre-processed Evolutionary LM Neural Networks (PELMNN) for the stock market identification. This proposed approach was the integration of genetic approaches, data pre-processing techniques, and the Levenberg Marquardt (LM) approach for the learning feed forward neural network. Data transformation approach was used for scaling the input data and the stepwise regression was also used for selecting the input data to remove the dissimilar variables in the pre-processing level. Then, for developing the neural networks, its initial weights were changed along with the LM approach by using the GA. These attained weights were utilized as the initial weights for LM backpropagation approach to the local search. At last, the output information was returned to the original value and identified value was created.

Anish and Majhi [52] developed a hybrid non-linear adaptive method for predicting the stock market. In this study, the feature selection approach and feedback functional link ANN approach was used. In the factor analysis, the input was selected from the raw data, which was a dominant statistical attributes reduction method. The feedback functional link ANN was used along with the recursive least square training method was developed for the prediction process. The recursive least square training method decreases the training time because it needs a less number of iterations. In order to overcome the local minima problem, the GA and the Particle Swarm Optimization (PSO) were developed. The performance of this method was compared with the other approaches like PCA and Discrete Wavelet transform.

The cuckoo optimized SVM was introduced by Devi., et al. [4] for the stock market prediction. However, the SVM was also used for managing the non-linear classification effectively. It classifies the data by plotting the samples from the low dimensional input space into the high dimensional characteristic space. The cuckoo search method based on the PSO approach was used for tuning the parameters of the SVM. The outcomes show that the cuckoo search SVM was attained more accuracy than the SVM.

Das., et al. [23] presented a Firefly and Online Sequential Extreme Learning Machine for the stock market identification. The firefly framework is the combination of the GA and the optimized feature reduction firefly optimization. Then, the feature reduction approaches were applied to the prediction model, such as Recurrent Back Propagation Neural Network (RBPNN), Extreme Learning Machine, and Online Sequential Extreme Learning Machine. On the other hand, the PCA method was used in the multivariate data analysis to decrease the dimension. Finally, the performance of the proposed approach was calculated using the four several stock market databases.

The Translation Invariant Morphological Time-lag Added Evolutionary Forecasting (TIMTAEF) was introduced by Araújo [53] for the stock market prediction. The developed TIMTAEF approach contains a GA and modular morphological neural network. The MGA population was trained by Backpropagation for enhancing the parameters. After that, the developed method selected the most tuned forecasting method for the time series depiction. It executed the behavioral statistical test and developed a phase fix procedure for regulating the time phase alternations, which was observed from the stock market. In the end, the performance metrics were computed, and it was utilized with the fitness function for enhancing the time series occurrence description.

Jabin [54] developed a feed forward ANN for the stock market prediction. The BPNN used the gradient descent method to tune the network parameters. The backpropagation approach joins both gradient descent and hill-climbing, in which gradient descent was used to improve the performance. This developed method learns the weight for a fully associated feed-forward multilayer network. Moreover, the gradient descent was utilized for reducing squared error among output values as well as the target values.

Araújo and Ferreira [55] modeled Evolutionary Morphological-Rank-Linear Forecasting (EMRLF) approach to identify the stock market. The developed EMRLF approach was an integration of Morphological-Rank- Linear (MRL) filter also the MGA. In this case, the smallest amount of the time lags was sufficient to signify a time series for the purpose of prediction. Then, MGA was improved by using the Least Mean Square (LMS) approach for altering the parameters of the MRL filter. MGA method was used for identifying the parameters, such as the primary parameters of MRL filter and least amount of time lags and their equivalent precise location to represent the time series. Furthermore, LMS approach was utilized for training every individual of the MGA population. After that, the construction of the prediction model executes the behavioral statistical analysis along with the phase fix procedure for regulating time phase distortions.

Oyewala., et al. [24] developed Auditory Algorithm (AA), which follows the pathway of the auditory system-like that of the human ear. The algorithm has shown an ability to detect exponential decay of the stock market. Identification of exponential decay helps in the detection of upward/downward movement or stability of the stock market. The algorithm outperformed four machine learning methods, such as Linear Regression, Support Vector Machine, Feed Forward Neural

Network, and Recurrent Neural Network and two continuous-time models, namely, stochastic differential equation and geometric Brownian motion. The algorithm was tested on five different stocks from Nigerian Stock Exchange.

d) Other stock market prediction techniques

This subsection describes techniques inspired by various engineering concepts or techniques which are not commonly used for stock market prediction:

Xi., et al. [56]] developed a neural network method for predicting the stock market. Initially, several noise data, like singular values of the continuous function jump discontinuity point was considered. A class of constructive decompose RBFNN was used for restoring the singular value of the continuous function along with the limited number of jumping discontinuity points. The continuous element was approximated by using the basic neural networks, which have better performance and the finest network architecture. Subsequently, the RBFNN was created for fitting the singular value and for optimizing the neural network structure.

Kannan., et al. [57] modeled data mining techniques for the stock market prediction. The five methods namely, Bolliener Bands (BB), Relative Strength Index (RSI), Bolliener Signal, Moving Average (MA), Typical Price (TP), Stochastic Momentum Index (SMI), and Chaikin Money Flow indicator (CMI) were utilized for identifying, whether the stock market is increase or decrease. Furthermore, this method considered several global events on the prediction of the stock market. The effectiveness of developed stock market prediction was discovered by comparing the moving average crossover.

Ballings., et al. [2] introduced a standard collection approach with a single classifier for the stock market prediction. This method contains the profitability indicators, liquidity indicators, and solvency indicators. In this study, the positive classes were over-sampled since the former made sure that there was no valuable data was discarded. For the logistic regression, a penalized technique was used to avoid the overfitting problem. This developed approach gives up the small bias in order to decrease the variance of the predicted values, and therefore, enhances the performance of the prediction. The feed forward neural network was utilized and one layer of hidden layer was used for the efficient prediction.

Aithal., et al. [58] introduced a data mining approach for identifying the stock market effectively. Initially, the knowledge, like data collection, data transformation, and data cleaning, was obtained. Then, the Bartlett tests, correlation matrix, and the Kaiser Meyer Olkin test were applied to reduce the dimensionality. Consequently, the dimensionality was decreased by using the PCA method and it was linked to the microelectronic systems, which were grouped together into the factors. Subsequently, the K1-kaiser and also scree test method was also used for preserving the factors. Then, the PCA along with varimax approach was used for finding the factors with the maximum variation. The feed forward neural network along with the sigmoid activation and backpropagation function shows the best accuracy and performance.

e) The deep learning-based techniques

The deep learning-based techniques introduced for stock market prediction are elaborated in this section.

Verma., et al. [59] modeled an ANN technique to predict stock market indices. A number of activation functions were executed with the decisions for a cross validation sets. Initially, the data was normalized and it was subjected to the ANN. The input vectors of the training data were normalized, such that every feature was non-zero. The target values were normalized based on the activation function. Then, the test vector was further scaled by a similar factor that likewise the training data was normalized. The ANN for the particular test vector was scaled back through a similar factor as target values for the training data.

Guresen., et al. [60], introduced an effective ANN model for the stock market predictions. This method contains a dynamic ANN, hybrid neural networks, and the Multi-Layer Perceptron (MLP). This method used the Generalized Autoregressive Conditional Heteroscedasticity (GARCH) for extracting the input variables. The non-linear processing element and the considerable interconnectivity were the significant characteristics in the MLP. The MLP was trained with the backpropagation algorithm, and the gradient descent method was used.

de Faria., et al. [61] developed a neural network method and the adaptive exponential smoothing approach to predict the stock market. For capturing non-linear behaviors, various methods were utilized. The backpropagation approach was used for the training in which the weights of the neurons were changed and it was back propagated with the network. The network architecture includes one input layer along with a number of neurons equal to the number of days in an input window, whereas the output layer had only one neuron subsequent to the prediction outcome. The adaptive estimation approach was executed and the two-performance metrics were analyzed for computing the prediction efficiency.

Moghaddam., et al. [62] modeled an ANN for predicting the stock market. MLP was utilized for solving the regression type issues. Neurons take the input parameters and inserts based on assigned weights and add a bias. The transfer function was applied and output values were calculated. The feed forward ANN was trained by the backpropagation approach. The performance metrics were analyzed and it was used for estimating the efficiency of prediction.

Patel., et al. [63] presented two-stage fusion methods to predict the stock market effectively. In the first stage of a fusion, the super vector regression model was used. In the second level of a fusion, the random forest and ANN were utilized. The super vector regression predicts the future values of statistical parameters and it was subjected to prediction models in the second level. Ten technical indicators were picked as inputs in every prediction model for better efficiency. The proposed method was also used in other areas, such as energy consumption prediction and weather prediction.

Yuan., et al. [64] presented a feature selection and a machine learning-based method for the stock market identification. Initially, the feature selection method was applied; it selected the features, which decreased the complexity, and avoided dimensional failure. The time sliding cross-validation approach was used for creating the model as more appropriate for the real situation. For both the stock price prediction and feature selection, Random Forest approach was used, which performs better than other methods. In the Random Forest approach, the long-short portfolio was developed for validating the efficiency of the method.

The instantaneous frequency approach was introduced by Zhang., et al. [65] for the stock market prediction. This method mainly focused on the class of signals, termed as simple wave, and the instaneous frequency is called counting instaneous frequency. Then, the intrinsic mode function was listed with the help of their frequencies, in the order of higher frequencies to lower frequencies. The instantaneous frequency was computed and it was utilized in the wave cycle projection for every simple wave. Additionally, the Radial Basis Function Neural Network (RBFNN) was joined along with the proposed method to identify the stock market. Moreover, Empirical Mode Decomposition (EMD) was adopted for attaining the simple wave.

Yang., et al. [66] presented a hybrid stock selection approach, which includes the stock prediction for efficiently predicting the stock market. The proposed approach mainly consists of two levels stock identification and stock scoring. Initially, the stock returns for the next period were forecasted by the Computational Intelligence (CI) approach of Extreme Learning Machine, which has quick computing speed and powerful learning capacity. Then, the stock scoring approach was introduced as the linear combination of a predicted factor and basic factors using a CI optimization for the top-ranked stocks and weights were picked for an evenly weighted portfolio.

Karhunen [67] introduced a contemporary statistical and machine learning method for identifying the stock market. This developed approach contains Logistic Regressions, Ordinary Last Squares, similarity-based classification, and regularized regressions. Then, the similarity-based classification approach was used, which rapidly computed the parameters and was robust for non-linearity. Moreover, the outcomes were both statistically and economically important for statistical analysis and increasing the trading charge. Additionally, the outcomes were created when the objective of the analysis was modified.

Zhou., et al. [68] presented an enhanced neural network approach, termed EMD and Factorization Machine-based Neural Network (EMD2FNN) for stock market prediction. Then, the EMD and the factorization machine were utilized in this method to identify the stock market trend. In this research, the EMD was used to defeat the non-stationary of the stock market and also for decomposing the original financial time series into various elements. Every extracted intrinsic mode function has the oscillatory models through the scales in a narrow series and analyzed as a quasi-stationary element. The factorization machine was utilized to grasp the non-linear connections between the inputs in which also effective in the calculation because of its linear complexity. Finally, using the functional neural network, the stock market was predicted and evaluated by various other techniques.

Ramezanian., et al. [69] presented a method with the integration of genetic network programming and the MLP to identify the stock market. Genetic network programming was applied for the rule extraction and forecasting the stock returns. After that, the MLP neural network was utilized for the classification of data and generating a set of related rules. After the data classification and rules extraction, adopting an econometric model Auto Regressive Moving Average-Generalized Auto Regressive Conditional Heteroskedasticity (ARMA)-GARCH was used for predicting the stock returns. The connection for ruling pools was attained by using the produced weights. The outcomes of this proposed approach increase the error handling enhancement also provide good prediction accuracy.

Wang [70] developed a stock pricing and stochastic volatility using the Taylor series for predicting the stock market. Initially, the financial market time series decomposition method was used for enhancing the predictive performance of the neural network. The stock market index information was decomposed into the empirical mode data series from the higher frequency to lower frequency. The time sequence was grasped from the various scales and directions, and the stock index was analyzed. Then, PCA was used for decreasing the data noise and compressing the redundant data, which helps for reducing the training time and enhance the performance of prediction. After that, the interval time series decomposition approach was applied for improving the prediction performance of the valley and the peak. Furthermore, the deep network training time and the data scale were used for enhancing the prediction performance.

Liu., et al. [71] presented a Heterogeneous Autoregressive Model (HAR) with the time-varying parameters (TVP) for the prediction of the stock market. Various usual constant coefficient HAR type approach includes the jump, variance, volatility leverage effect, which was enlarged to the TVP approaches. The HAP method was developed with the time-varying parameters for computing the errors. The QLIKE loss function was selected as the measure for assessing the prediction accuracy. Furthermore, the combined HAR method includes the jump factors and the continuous components for the effective outcome. Additionally, the TVP HAR method was included with leverage effect, continuous volatility element, and the signed jump for predicting the stock market.

2.2. Sentiment analysis-based techniques

This section describes the sentiment analysis-based techniques in stock market identification using various approaches.

Nguyen., et al. [72] modeled a sentiment analysis method to predict the stock market prediction. In this study, two techniques were developed for extracting the sentiments, such as JST-based and Aspect-based method. The two datasets, such as mood information and the historical price dataset was utilized for predicting the stock market. The SVM was used to effectively manage the high dimensional data's and for the classification. Moreover, the developed method automatically extracts the related sentiments and the topics from the texts, which was present in the message board.

Wang., et al. [73] developed the model independent structure for predicting the stock market. This approach was efficient during the numerical simulations. Also, for the identification process, the method included compositional data, scalar data, and functional data. Furthermore, the transformation approaches were developed for managing the functional information and the compositional information. After that, the model-independent structure was used to identify the stock market using preliminaries. Lastly, a logistic regression method was utilized and the equivalent computation process was employed.

Dang and Duong [74] presented a time series analysis with the enhanced text mining approaches for predicting the stock market. At first, articles were taken from online websites and extracted file content presented in the format of plain text. Then, every article from the collection was pre-processed for getting an optimized dataset. After that, every article was labeled into the exact class of negative, positive, or neutral with the help of stock price. Then, the dataset was processed by the natural language phase, which consists of the term weighting and the term selection. Finally, the SVM was utilized for training as well as testing the data.

Wang., et al. [1] modeled a hybrid time-series predictive neural network for predicting the stock market. The distributed model was utilized for mapping in which the news was mapped to the vector space. Then, for decreasing the dimension of the word vector matrix and for obtaining the precise and useful text information, the sparse autoencoder method was introduced. Then, a hybrid neural network model was established for predicting stock volatility, which was combined with a deep convolutional layer for capturing the text features. In order to decrease feature prediction fault, this developed method was joined the price features and the news.

Ding., et al. [75] presented a deep learning approach for the stock market identification. Initially, events were extracted from news text and represented as a dense vector. Then, for modeling the long-term influences and the short-term influences of events, deep CNN was utilized. Furthermore, the bag of words was developed for representing the news documents, and also the SVM was developed for creating the prediction method. Moreover, the randomization test was implemented in this model to validate the statistical importance.

Liu and Wang [3] developed Numerical-based Attention (NBA) approach for the prediction of the stock market. The attention-based method was introduced in this method to efficiently utilize the complementarity among the news and numerical data. The stock trend information in the news was converted into a significant distribution of the numerical information. Therefore, the news was encoded to guide the selection of numerical data. This approach has removed the noise and produced the entire use of trend information in the news. Furthermore, for assessing NBA approach, the numerical data and the news corpus were collected for making three datasets from the two sources.

Zhang., et al. [76] introduced the multiple source multiple instance method to identify the stock market prediction efficiently. Initially, a sentiment analyzer was used for obtaining collective sentiments from specific databases. The extracted sentiments were subjected to the multiple source multiple instance methods or efficiently merged the events. The multi-source super group-level labels, multi-source group level labels, and instance-level labels were differentiated for making the predictions interpretable. However, the main intention of this method was for predicting the multi-source super group-level labels, which showed the increase or decrease of the stock market index. Furthermore, the event representation learning process was also developed to capture the event information efficiently.

Chen., et al. [77] introduced a Factorization Machine (FM) for forecasting the stock market identification. Initially, the relationship of FM to a generalized linear model and the SVM was analyzed and also the reasons for using FM in high dimensional data. Then, for predicting the stock market on precise textual features, social media was utilized. After that, the better performance of FM was described by comparing along with the various non-trivial baselines and exploring the sensitivity of FM for representing the textual.

Oliveira., et al. [78] introduced a robust methodology for predicting the stock market variables, such as portfolios, trading volume of diverse indices, returns, and volatility. This approach utilized the attention and sentiment indicators, which were extracted from the survey indices and microblogs. The Kalman filter was used to combine the survey sources and the microblogs. The Diebold-Mariano test and the various machine learning methods were used for predicting the accuracy, which was based on the major two schemes, termed microblog-based and the baseline model. The less noisy Kalman filter sentiment indicator was developed that joined the measures of various periodicities. In addition, Twitter sentiment analysis was used for effective forecasting.

Zhang., et al. [79] introduced tensor factorization and coupled matrix technique for the stock market identification. This method combines the sentiments, events, and quantitative features were extracted from several data sources, like social media, stock quantitative data providers, and web news. The tensor was used for combining the heterogeneous data and confines the intrinsic relations between investors and events. Because of the sparsity of the tensor, the stock correlation matrix, the stock quantitative feature matrix, and two auxiliary matrices were generated and integrated to help the tensor decomposition because of the sparsity of the tensor. The multiple related stocks were concurrently predicted throughout their commonalities, which were permitted by sharing the collaboratively factorized minimum rank matrices among the tensor and matrices.

3. Research Gaps Identified

This section illustrates the research gaps and the problems faced by previous stock market prediction approaches. The analysis is depicted in Table 1. The research gaps and issues are arranged in the same sequence of categorization followed in the previous section.

Table 1. Research gaps and issues

Authors	Research gaps and issues
Wang., et al. [39]	A clustering approach was introduced for calculating weights using a number of occurrences. However, the method was not able to manage the multi-factor prediction.
Chen and Hao [40]	They introduced feature weighted SVM approach. The application of correlation-weighted approaches is to be tested.
Barak and Modarres [41]	Relation of risk and return also feature selection were established, but an optimized metaheuristics algorithm was not developed in the method for obtaining better prediction results.
Wen., et al. [42]	Other datasets with high stability were not considered for the stock market prediction
Chen., et al. [43]	The encoder was not effectively recognizing the entire relative information in the time series.
Long., et al. [44]	The sentiment and text information were not combined to improve the performance.
Baek and Kim [46]	The augmentation method was developed to identify the stock market. However, the method failed to implement an optimized multi-modular trading system for efficiently dealing with multimodal data.
Zhou., et al. [48]	The adversarial neural network was used to identify the stock market. However, the method does not combine the predictive models under the multi-scale limitations for obtaining better accuracy.
Tsai and Hsiao [50]	The rate of prediction accuracy by combination technique was not increased.
Devi., et al. [4]	The method failed to include the social media sentiments and analyze the impact of several factors, like dollar price, crude oil price, and gold price, towards the stock market index for improving the system performance.
Devi., et al. [4]	The integrated framework was developed for predicting stock market prediction, but the method did not include several filters for recognizing more efficient rules to the classification and selection process by the classifiers.
Das., et al. [23]	The method was not extended to the other financial features along with analyzing their controlling factors.
Araújo [53]	The method failed to adapt the classic approaches for evaluating the performance.
Araújo [55]	The robust approach was used for predicting the stock, but the method was not recognized the influential microblog users and did not consider their involvement in predicting explicit stocks. The evolutionary morphological rank linear method was developed for stock market prediction. However, the developed method was failed to consider the financial return and risk to discover the extra cost-effective benefits.
Oyewala., et al. [24]	Applicability of auditory algorithm shall be extended to predict cryptocurrency, gold, other stock markets, and other commodities like silver, copper, oil, and gas.
Aithal., et al. [58]	The data mining approach was developed for identifying the stock market. However, this technique was not extended for recognizing the qualitative data manipulating stock markets.
Guresen., et al. [60]	ANN was used to identify the stock market. Although, this system failed to use the time series for understanding the inner dynamic of hybrid model performance.
Patel., et al. [63]	The machine learning approach was introduced for stock market prediction, although the method failed to utilize more statistical parameters as input for finding better correlations.
Yuan., et al. [64]	An integrated feature selection and machine learning approach was developed for stock market prediction, but this method failed to optimize the feature selection algorithm for obtaining better performance.
Yang., et al. [66]	The method failed to extend the other tasks in quantitative asset management, like market-timing determination and portfolio formulation.
Karhunen [67]	The statistical and machine learning approach was developed for forecasting the stock market, but this technique failed to consider the high-quality data sets for better prediction.
Wang [70]	Stocking pricing model-based Taylor series was developed to identify the stock market, although the approach was failed to develop multi-scale financial data fusion approaches for effective prediction.
Nguyen., et al. [72]	The sentiment analysis model was introduced, but the method was still failed to automatically extract the sentiments and topics concurrently for stock market identification. The model independent framework was developed for identifying the stock market; however, the method was not added the other informative variables in the framework for improving the prediction power.
Dang and Duong [74]	The stock prices prediction and the technical analysis were not combined for enhancing the performance of the system.
Wang., et al. [1]	The hybrid time series predictive neural network was developed for stock market prediction even though the method failed to segment the sequence window for good accuracy.
Liu and Wang [3]	NBA approach was developed for stock market prediction, but the method has not explored the efficiency of the developed method in industrial or index level data.

4. Analysis and Discussion

This division describes the discussion of stock market identification techniques used, use of optimizational gorithms and feature selection methods, tools used, dataset utilized, performance evaluation metrics, and input parameters. The aim of this discussion is to identify mostly used and suitable algorithms, the significance of optimization algorithm and feature selection method, to find the applicability of other optimization algorithms and feature selection methods, to identify commonly used input parameters and performance evaluation matrices, to identify potential input parameters, to identify tools used mostly, to identify the most and the least used datasets, and to identify less used stock exchange datasets. Secondly, this discussion will also help to find directions for future research.

4.1. Analysis of techniques used

This subsection depicts the review on the basis of techniques and methodology used in research papers, analyzed for stock market identification. Fig. 2 shows CNN, RNN and LSTM, SVM, and ANN are widely researched and accepted algorithms for stock market prediction. It has been observed that recent researches revolve around complex and deep structures based on LSTM and CNN. Apart from this, different unconventional methodologies originating from different fields of engineering are also tested.

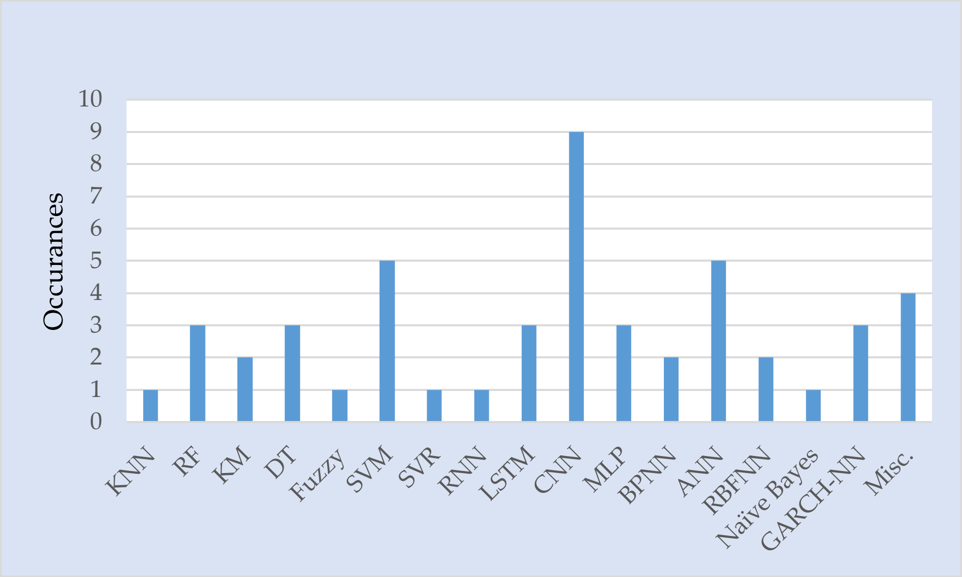

Fig. 2. Evaluation based on the techniques used

4.2. Analysis of optimization algorithm used

This subsection illustrates the usefulness of the optimization algorithms utilized by the existing stock market identification methods for identifying correct weights neural network structures. Fig. 3 shows that out of 23 neural network-based techniques covered, 11 techniques are using optimization methods. It has been observed by various researchers that controlling and optimizing hyperparameters affects prediction accuracy [34, 80]. However, many of the algorithms are using genetic algorithm for optimization. Fig. 4 shows only a few other algorithms, such as Firefly, Cuckoo, PSO, Auditory, have been tried. The applicability of various optimization algorithms is yet to be tested.

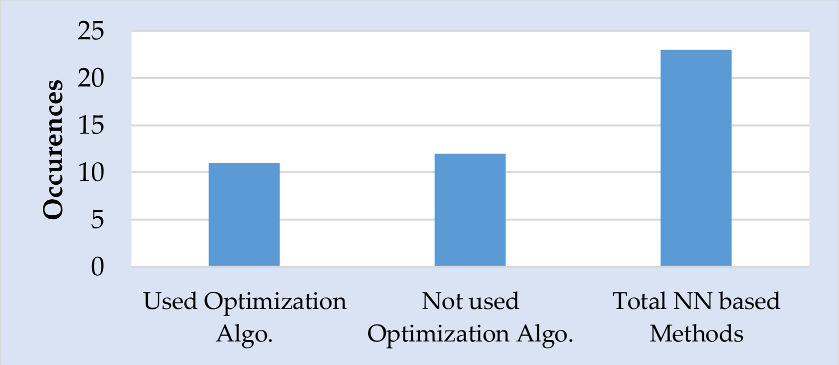

Fig. 3. Optimization algorithm used for controlling hyperparameters

4.3. Analysis of feature selection techniques used

http://doi.org/10.26594/register.v8i1.2533

This subsection analyzes the importance of feature selection in stock market prediction. It is known that the stock market generates abundant data, including stock prices, ratios, news, sentiments, indicators, etc. It is important to select features that weigh more in the prediction of the stock market. Fig. 5 shows techniques used for sentiment analysis do more use of feature selection as compared to time series analysis. A limited number of works for identifying proper input parameters in the time series analysis part are observed.

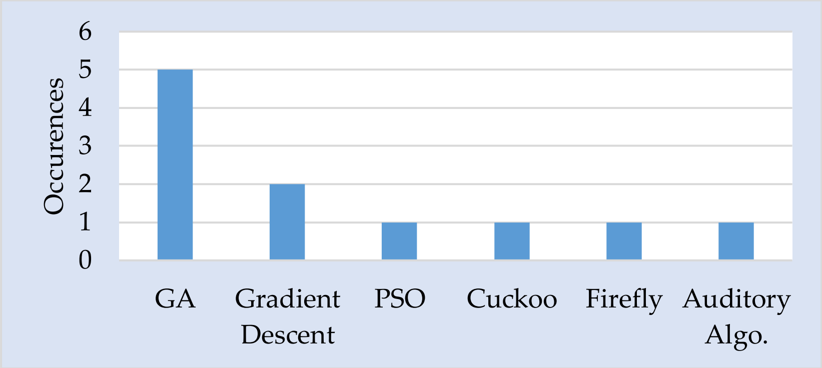

Fig. 4. Usage frequency of different optimization algorithms

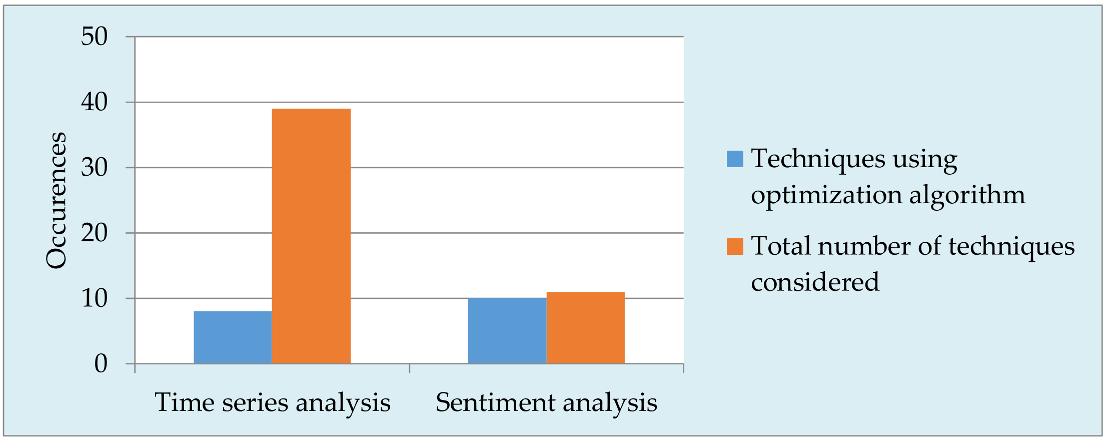

Fig. 5. Evaluation based on feature selection technique used

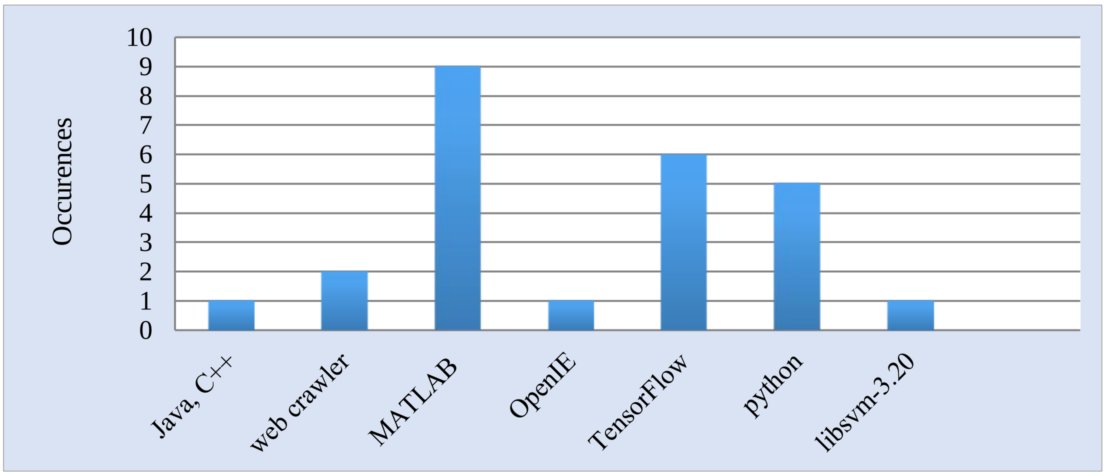

Fig. 6. Evaluation based on the toolset

4.4. Analysis of employed tools

This subsection illustrates the toolset utilized by the existing stock market identification methods. Fig. 6 shows the analysis based on the toolset used for stock market prediction. The software toolsets utilized in the research papers are MATLAB, Tensorflow, Python, JAVA, Web crawler, OpenIE, libsvm. In Fig.

6, it is clearly interpreted that MATLAB software is often utilized software tool for the stock market identification method.

4.5. Evaluation of employed datasets

This subsection detailed the investigation carried out based on datasets utilized by existing research works.

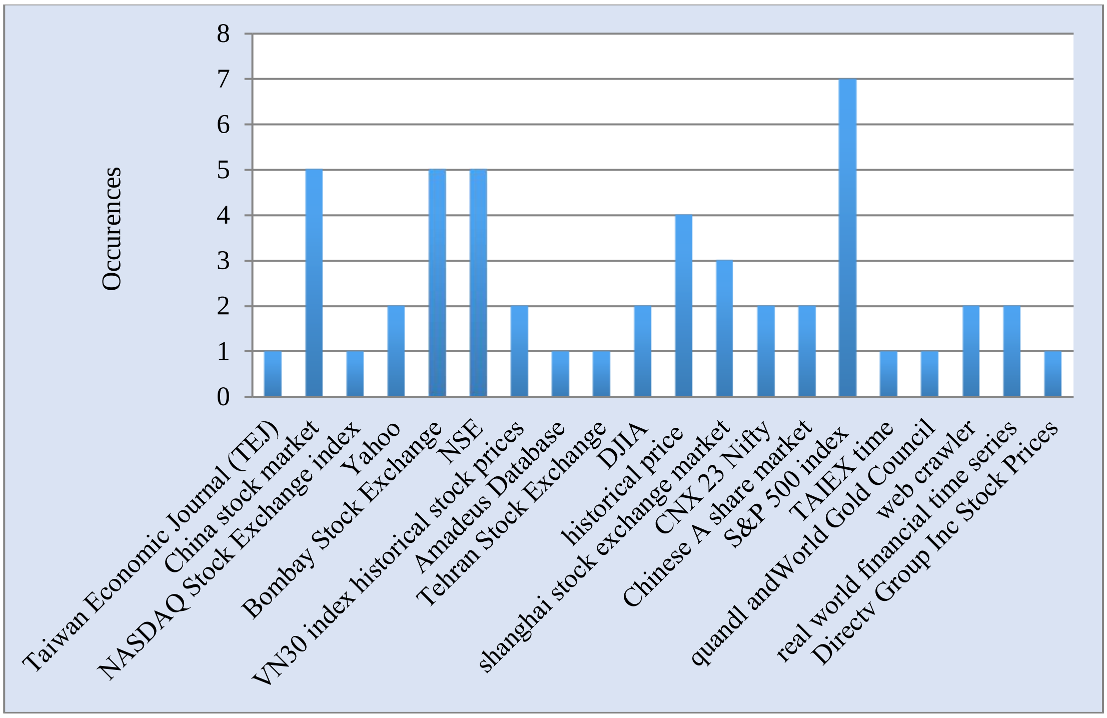

Fig. 7. Evaluation based on datasets

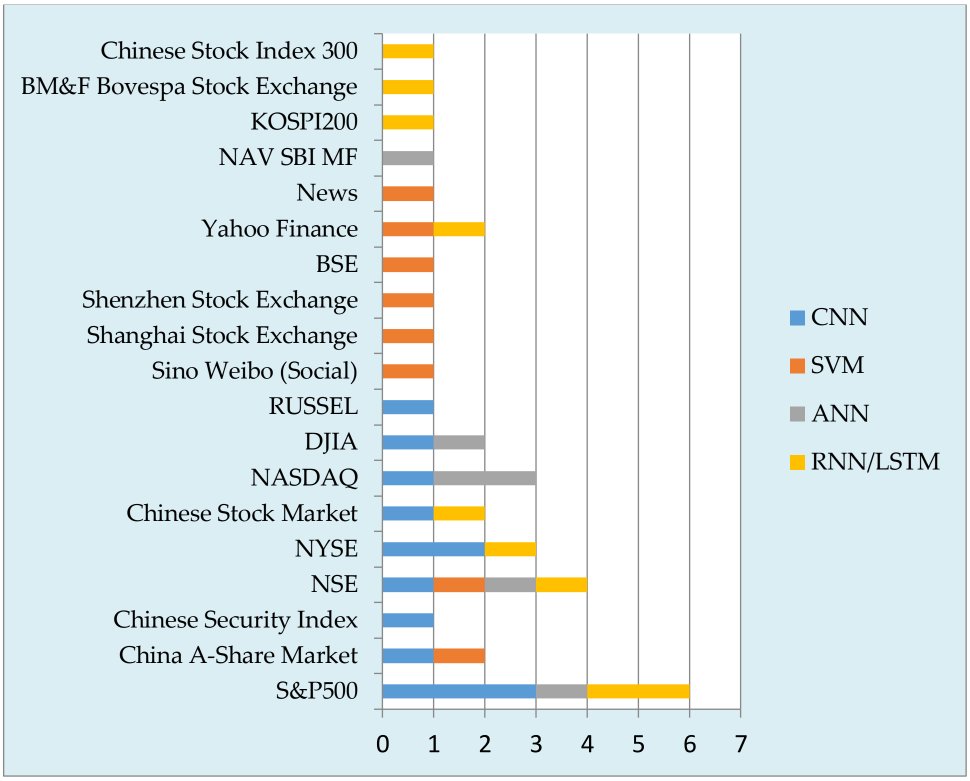

Fig. 8. Comparative analysis of techniques using different datasets

Fig. 7 shows the several datasets utilized for stock market identification. The commonly used datasets in the stock market prediction are Taiwan Economic Journal (TEJ) database, China stock market dataset, NASDAQ Stock Exchange index, National Stock Exchange (NSE) financial databases, yahoo, Bombay Stock Exchange (BSE), VN30 index historical stock prices, Amadeus Database, Tehran Stock

Exchange dataset, Standard & Poor's 500 stock (S&P 500) index, RMRF, DJIA, historical price dataset, mood information dataset, China Security Index 300 (CSI300), shanghai stock exchange market dataset, RUSSELL 2000, CNX 23 Nifty, Directv Group Inc Stock Prices, Chinese A share market database, web crawler, Technical Assistance and Information Exchange (TAIEX) time dataset, Quandl and World Gold Council dataset, and the real-world financial time series dataset. In Fig. 7, it is comprehensible that the most repeatedly used dataset is S&P 500 index.

Fig. 8 shows comparative analyzes of datasets of different stock exchanges used by different techniques. It is evidently in Fig. 8 that S&P500 is widely researched by various techniques. Similarly, in Asia, Chinese Stock Exchanges are widely researched. However, very few techniques have been tested on stock exchanges' data of other countries and unconventional data sources like news and social sentiments. It is observed that different exchanges have different kinds of financial flows, which build exchanges [81]. The techniques that were successful on S&P500, are yet to be tested on data of other exchanges.

Table 2 shows links of majorly researched stock exchange datasets. These links lead to stock exchange historical data sections or websites providing such datasets. The required dataset can be downloaded by just selecting the date range and script.

Table 2. Links for accessing different datasets

Datasets	Links
NSE	https://www1.nseindia.com/products/content/equities/indices/historical_index_data.htm
BSE	https://www.bseindia.com/markets/equity/EQReports/StockPrcHistori.aspx?ffla=1
S&P500	https://www.wsj.com/market-data/quotes/index/SPX/historical-prices
Yahoo Finance	https://finance.yahoo.com/quote/%5EGSPC/history/
NASDAQ	https://data.nasdaq.com/databases/DY4/data https://www.nasdaq.com/market-activity/quotes/historical
NYSE	https://www.nyse.com/market-data/historical
Shanghai Composite Index	https://in.investing.com/indices/shanghai-composite-historical-data
MF NAVs	https://www.sbimf.com/en-us/navs
News	https://www.vsestock.vn https://www.hsx.vn https://www.hsn.vn
Microblogging site	https://open.weibo.com/wiki/API%E6%96%87%E6%A1%A3/en

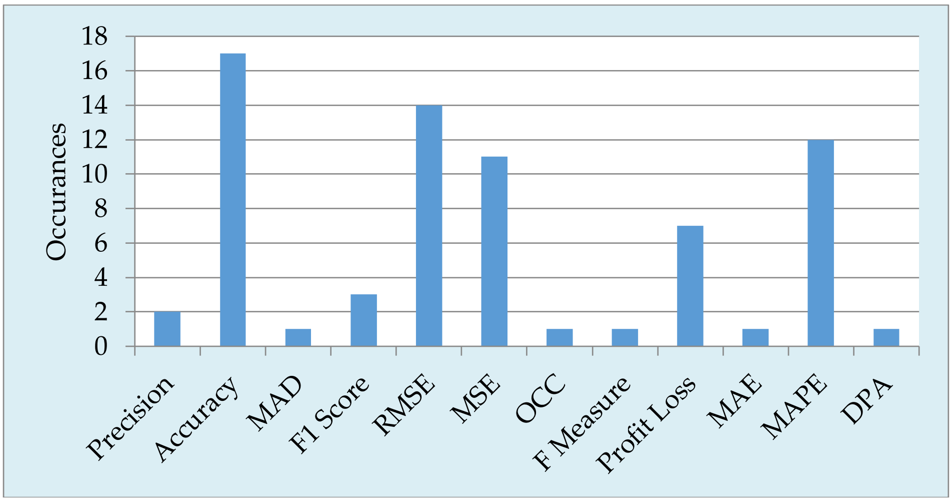

Fig. 9. Evaluation based on evolution matrices used

4.6. Analysis of evaluation metrics

The evaluation based on performance metrics is examined in this subdivision. The performance metrics considered are Accuracy, Root Mean Squared Error (RMSE), Mean Square Error (MSE), Direction

Prediction Accuracy, Mean Absolute Deviate, Recall, Average Profit, Operating Characteristic Curve, Mean Absolute Error (MAE), Relative Variance, Precision, F1-Score, and Matthews Correlation Coefficient (MCC).

In Table 3 and Fig. 9, it can be evaluated that Accuracy, MSE, RMSE, MAPE are commonly preferred. However, considering the relationship of RMSE and MSE, it can be inferred that they are the most used and suitable evolution parameters for stock market prediction. Profit Loss is an unconventional evaluation parameter, which has been widely used and justifies the actual goal of the prediction process. Fig. 10 shows a comparative analysis using different evolution matrices. It can be observed that SVM, LSTM, and CNN based methods majorly used accuracy and RMSE, while ANN mostly relies on RMSE and MAPE. It can also be stated that Accuracy and RMSE are widely accepted among the majority of the techniques, as compared to other evolution matrices.

Table 3. Evaluation based on the performance metrics

Performance Metrics	A Number of Research Papers
Accuracy	[1, 3, 4, 37, 38, 41, 42, 44, 45, 50] [58, 71, 72, 73, 74, 76, 79, 82]
Mean Square Error	[3, 23, 46, 53, 54, 55, 60, 63, 71, 73] [78]
Root Mean Squared Error	[4, 39, 40, 43, 48, 56, 61, 63, 65, 66] [68, 69, 70, 71]
Direction Prediction Accuracy	[48]
Mean Absolute Deviate	[51]
Recall	[42, 74]
Average profit	[37]
Operating Characteristic Curve	[2]
Mean Absolute Percentage Error	[4, 23, 40, 46, 47, 52, 53, 63, 66, 68] [69, 70]
Profit Loss	[37, 41, 45, 57, 64, 65, 69]
Mean Absolute Error	[46]
Relative Variance	[52]
Precision	[42, 64]
F1-Score	[42, 45, 76]
Matthews Correlation Coefficient	[75, 79]

Table 4. Analysis in terms of accuracy

Accuracy Range	Number of Research Papers
51% - 60%	[3, 44, 49, 82]
61% - 70%	[1, 38, 41, 71, 73, 75, 79]
71% - 80%	[50, 74]
81% - 90%	[4, 42, 45, 72, 77]
91% - 99%	[37, 58, 59, 76]

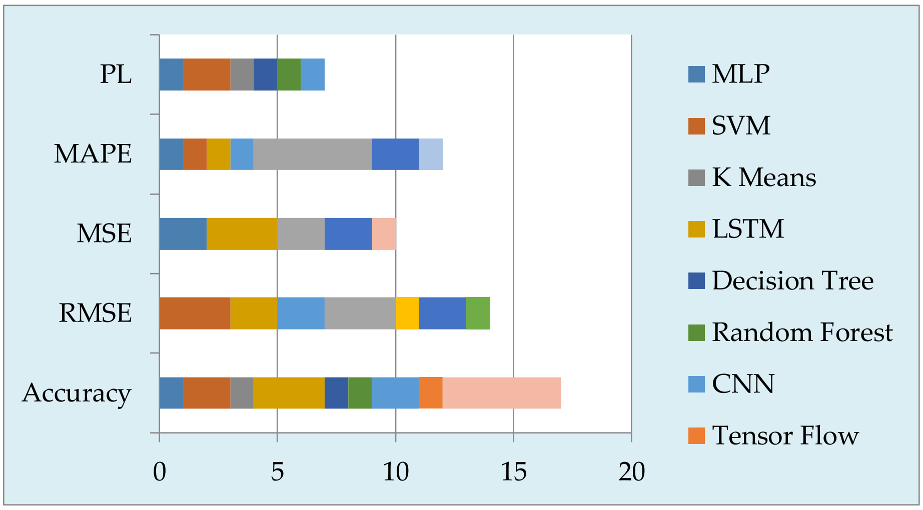

Fig. 10. Comparative analysis of techniques using different evolution matrices

http://doi.org/10.26594/register.v8i1.2533

The evaluation made using the performance metrics is discussed in this subsection using Table 4. Table 4 shows the review based on the Accuracy is specified by five ranges: 51% - 60%, 61% - 70%, 71% - 80%, 81% - 90%, and 91% - 99%. As in Table 4, it well-known that the research papers [37, 58, 59, 76] attained high accuracy with range of 91% - 99% and [3, 44, 49, 82] research papers had low accuracy within the range of 51% - 60%.

4.7. Analysis of input parameters

The evaluation of the input parameters is depicted in this subsection:

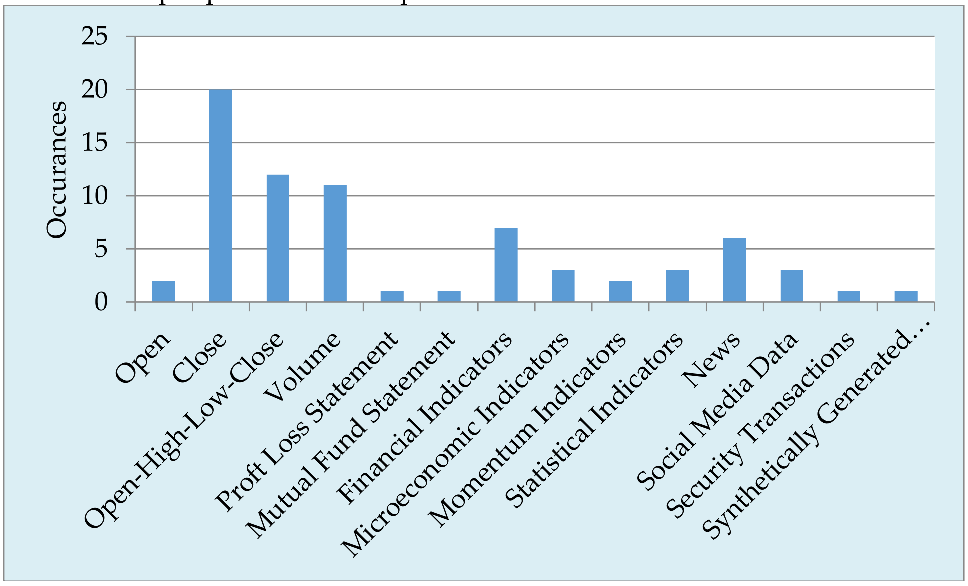

Fig. 11. Evaluation based on input parameters used

Fig. 11 depicts closing price, open-high-low-close prices, volume are highly used input parameters in time series-based analysis, whereas the news remained to be used more compared to social media data in sentiment analysis. Time series-based input parameters are more researched compared to textual data using news and microblogging sites. Financial indicators and statistical indicators derived from stock prices are the second majorly researched combination.

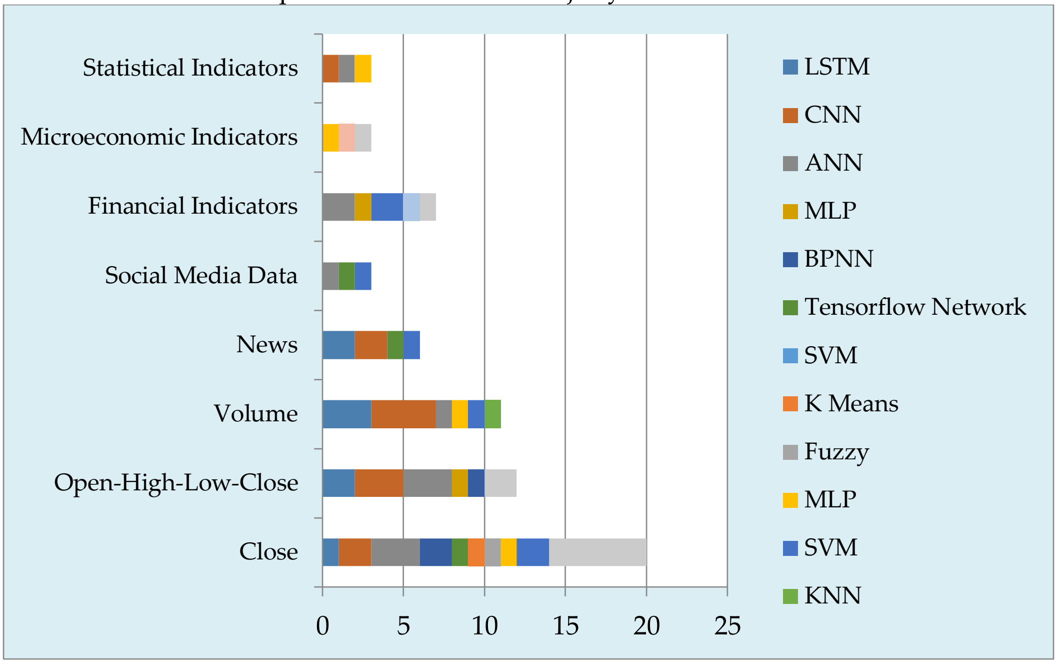

Fig. 12. Comparative analysis of techniques using different input parameters

Fig. 12 shows that LSTM and CNN are equally used time series as well as sentiment analysis. Other neural network techniques are majorly used for the time series part. Closing price, open-high-low-close combination, volume is used along with the majority of the techniques. However, the usefulness of derived time series parameters such as financial indicators, statistical indicators, momentum indicators for stock market prediction is yet to be evaluated.

5. Conclusion

A survey on several stock market identification approaches is explicated in this study. The research papers used in this survey are gathered from different sources, like Springer, Elsevier, MDPI, Google scholar, IEEE, and so on. The gathered research papers are divided into two categories, time series analysis and sentiment analysis, considering two types of input data, historical data and textual data in the news and microblogging sites, respectively. The time series analysis is a widely researched area, and hence it is further divided into subsections based on techniques used. Classical techniques cover early approaches for stock market prediction. It shows learning methods are applicable in stock market prediction. Furthermore, complex techniques can provide better results. It is identified that methods combining LSTM and CNN with other models are widely researched methods in recent times. Hence, it is discussed in a separate section to assess the current research direction. The evaluation techniques section identifies the need for optimization algorithms for controlling hyper-parameters of learning networks. The methods inspired by heterogeneous engineering concepts or sparsely used learning approaches are covered in the other techniques section. The sentiment-based techniques analyze the effects of events happening in the physical world and thereby creating sentiments.

The gathered research papers are reviewed to find research gaps and problems faced. Major research gaps found are as follows: 1) Time series analysis majorly uses daily stock prices and tries to predict prices for the next few days. Prediction for the long term is not frequently researched; 2) CNN, LSTM, SVM, and ANN are widely used in stock market prediction. These techniques have been tested in combination with other methods on a wide range of datasets from various stock exchanges. However, their applicability in sentiment-based analyses needs to be verified. Besides, the performance of other deep learning methods in combination with different algorithms shall be evaluated; 3) The usefulness of optimization algorithms has been proved. GA remains the most used technique. The work on optimization techniques catering to the needs of the stock market is lacking. Apart from that, other optimization techniques such as cuckoo search, firefly, dragonfly, auditory technique, shall be evaluated in combination with widely used prediction techniques; 4) Closing price, Open-High-Low-Close combination, and volume are commonly used input parameters. Less number of feature selection methods has been applied in time series analysis to identify optimal feature set and their effects on prediction; 5) Various indicators such as momentum indicators, financial indicators, statistical indicators, which cover indications considering different aspects, are also proved to be suitable for prediction. They are yet to be subjected to feature selection methods for identifying the optimum set of indicators. These indicators are evaluated using only a few prediction techniques; 6) Sentiment analysis relies on feature selection for prediction. However, a combination of sentiment and time series data is rarely used as input; 7) S&P500 and Chinese Stock Exchange are widely surveyed and analyzed data sources. Widely used prediction techniques such as CNN, LSTM, SVM shall be subjected for evaluation using datasets from sparsely used exchanges like NSE, BSE, and SSE; 8) The analysis of evaluation matrices shows that accuracy, RMSE, and MAPE are widely used approaches. Profit loss is an unconventional evaluation strategy used for stock market prediction techniques. Algorithmic trading is another area that is getting popular but less researched. Combining profit loss strategy with an algorithmic trading idea has not been evaluated yet.

These are major gaps and limitations that need to be addressed in the future by adapting advanced stock market prediction techniques. Through this review, various dimensions of research have been identified. These directions will play an important role for researchers to develop prediction algorithms in the future.

Author Contributions

J. B. Pandya: Conceptualization, data curation, formal analysis, investigation, resources, software, validation, visualization, writing-original draft and writing-review & editing. U. K. Jaliya: Conceptulization, investigation, resouces, software, supervision, validation, visualization and wrting-review & editing.

Declaration of Competing Interest

We declare that we have no conflict of interest.

References

[1] Y. Wang, H. Liu, Q. Guo, S. Xie and X. Zhang, "Stock Volatility Prediction by Hybrid Neural Network," *IEEE Access,* vol. 7, pp. 154524-154534, 2019.

[2] M. Ballings, D. V. d. Poel, N. Hespeels and R. Gryp, "Evaluating multiple classifiers for stock price direction prediction," *Expert Systems with Applications,* vol. 42, no. 20, pp. 7046-7056, 2015.

[3] G. Liu and X. Wang, "A Numerical-Based Attention Method for Stock Market Prediction With Dual Information," *IEEE Access,* vol. 7, pp. 7357-7367, 2018.

[4] K. N. Devi, V. M. Bhaskaran and G. P. Kumar, "Cuckoo optimized SVM for stock market prediction," in *2015 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS)*, Coimbatore, India, 2015.

[5] H.-C. Liu, Y.-H. Lee and M.-C. Lee, "Forecasting China Stock Markets Volatility via GARCH Models Under Skewed-GED Distribution," *Journal of Money, Investment and Banking,* vol. 7, 2009.

[6] G. E. P. Box, G. M. Jenkins, G. C. Reinsel and G. M. Ljung, Time Series Analysis: Forecasting and Control, New Jersey: Wiley, 2016.

[7] E. Chong, C. Han and F. C. Park, "Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies," *Expert Systems with Applications,* vol. 83, pp. 187-205, 2017.

[8] S. H. Kim and S. H. Chun, "Graded forecasting using an array of bipolar predictions: application of probabilistic neural networks to a stock market index," *International Journal of Forecasting,* vol. 14, no. 3, pp. 323-337, 1998.

[9] M.-C. Wu, S.-Y. Lin and C.-H. Lin, "An effective application of decision tree to stock trading," *Expert Systems with Applications,* vol. 31, no. 2, pp. 270-274, 2006.

[10] A. Booth, E. Gerding and F. McGroarty, "Automated trading with performance weighted random forests and seasonality," *Expert Systems with Applications,* vol. 41, no. 8, pp. 3651-3661, 2014.

[11] D. Brownstone, "Using percentage accuracy to measure neural network predictions in Stock Market movements," *Neurocomputing,* vol. 10, no. 3, pp. 237-250, 1996.

[12] P. Ou and H. Wang, "Prediction of Stock Market Index Movement by Ten Data Mining Techniques," *Modern Applied Science,* vol. 3, no. 12, 2009.

[13] M. V. Subha and S. T. Nambi, "Classification of Stock Index movement using k-Nearest Neighbours (k-NN) algorithm," *WSEAS Transactions on Information Science & Applications,* vol. 9, no. 9, pp. 261-270, 2012.

[14] S.-D. Kim, "Data Mining Tool for Stock Investors' Decision Support," *The Journal of the Korea Contents Association,* vol. 12, no. 2, pp. 472-482, 2012.

[15] M. T. Leung, H. Daouk and A.-S. Chen, "Forecasting stock indices: a comparison of classification and level estimation models," *International Journal of Forecasting,* vol. 16, no. 2, pp. 173-190, 2000.

[16] G. J. O. Osório, J. C. O. Matías and J. P. S. Catalão, "Hybrid evolutionary-adaptive approach to predict electricity prices and wind power in the short-term," in *2014 Power Systems Computation Conference*, Wroclaw, Poland, 2014.

[17] M. M. Beno, V. I. R, S. S. M and B. R. Rajakumar, "Threshold prediction for segmenting tumour from brain MRI scans," *International Journal of Imaging Systems and Technology,* vol. 24, no. 2, pp. 129-137, 2014.

[18] X. Wei, Z. Xie, R. Cheng, D. Zhang and Q. Li, "An Intelligent Learning and Ensembling Framework for Predicting Option Prices," *Emerging Markets Finance and Trade,* vol. 57, no. 15, pp. 4237-4260, 2021.

[19] R. Gao, W. Wu, C. Lang and L. Lang, "Geometric Asian barrier option pricing formulas of uncertain stock model," *Chaos, Solitons & Fractals,* vol. 140, p. 110178, 2020.

[20] I. Zghal, S. B. Hamad, H. Eleuch and H. Nobanee, "The effect of market sentiment and information asymmetry on option pricing," *The North American Journal of Economics and Finance,* vol. 54, p. 101235, 2020.

[21] B. Chen and M. Kankanhalli, "Pricing Average Price Advertising Options When Underlying Spot Market Prices Are Discontinuous," *IEEE Transactions on Knowledge and Data Engineering,* vol. 31, no. 9, pp. 1765-1778, 2019.

[22] B. R. Rajakumar, "Impact of static and adaptive mutation techniques on the performance of Genetic Algorithm," *International Journal of Hybrid Intelligent Systems,* vol. 10, p. 11–22, 2013.

[23] S. R. Das, D. Mishra and M. Rout, "Stock market prediction using Firefly algorithm with evolutionary framework optimized feature reduction for OSELM method," *Expert Systems with Applications: X,* vol. 4, p. 100016, 2019.

[24] D. O. Oyewola, A. Ibrahim, J. A. Kwanamu and E. G. Dada, "A new auditory algorithm in stock market prediction on oil and gas sector in Nigerian stock exchange," *Soft Computing Letters,* vol. 3, p. 100013, 2021.

[25] S. Mukhopadhyay and S. Banerjee, "Global optimization of an optical chaotic system by Chaotic Multi Swarm Particle Swarm Optimization," *Expert Systems with Applications,* vol. 39, no. 1, pp. 917-924, 2012.

[26] M. R. Tanweer, S. Suresh and N. Sundararajan, "Self regulating particle swarm optimization algorithm," *Information Sciences,* vol. 294, pp. 182-202, 2015.

[27] B. R. Rajakumar, "Static and adaptive mutation techniques for genetic algorithm: a systematic comparative analysis," *International Journal of Computational Science and Engineering,* vol. 8, no. 2, pp. 180-193, 2013.

[28] D. Binu and B. S. Kariyappa, "RideNN: A New Rider Optimization Algorithm-Based Neural Network for Fault Diagnosis in Analog Circuits," *IEEE Transactions on Instrumentation and Measurement,* vol. 68, no. 1, pp. 2-26, 2019.

[29] G.-G. Wang, S. Deb and L. d. S. Coelho, "Elephant Herding Optimization," in *2015 3rd International Symposium on Computational and Business Intelligence (ISCBI)*, Bali, Indonesia, 2015.

[30] G.-G. Wang, "Moth search algorithm: a bio-inspired metaheuristic algorithm for global optimization problems," *Memetic Computing,* vol. 10, p. 151–164, 2018.

[31] Z. Hu, Y. Zhao and M. Khushi, "A Survey of Forex and Stock Price Prediction Using Deep Learning," *Applied System Innovation,* vol. 4, no. 1, p. 9, 2021.

[32] D. Shah, H. Isah and F. Zulkernine, "Stock Market Analysis: A Review and Taxonomy of Prediction Techniques," *International Journal of Financial Studies,* vol. 7, no. 2, p. 26, 2019.

[33] D. P. Gandhmal and K. Kumar, "Systematic analysis and review of stock market prediction techniques," *Computer Science Review,* vol. 34, p. 100190, 2019.

[34] A. Thakkar and K. Chaudhari, "A comprehensive survey on deep neural networks for stock market: The need, challenges, and future directions," *Expert Systems with Applications,* vol. 177, p. 114800, 2021.

[35] P. S. Rao, K. Srinivas and A. K. Mohan, "A Survey on Stock Market Prediction Using Machine Learning Techniques," in *ICDSMLA 2019*, Singapore, 2019.

[36] S. M. Idrees, M. A. Alam and P. Agarwal, "A Prediction Approach for Stock Market Volatility Based on Time Series Data," *IEEE Access,* vol. 7, pp. 17287-17298, 2019.

[37] M. S. Babu, N. Geethanjali and B. Satyanarayana, "Clustering Approach to Stock Market Prediction," *International Journal of Advanced Networking and Applications,* vol. 3, no. 4, pp. 1281-1291, 2012.

[38] A. Gupta and S. D. Sharma, "Clustering-Classification Based Prediction of Stock Market Future Prediction," *International Journal of Computer Science and Information Technologies,* vol. 5, no. 3, pp. 2806-2809, 2014.

[39] W. Wang, Y. Shi and R. Luo, "Sparse Representation Based Approach to Prediction for Economic Time Series," *IEEE Access,* vol. 7, pp. 20614-20618, 2019.

[40] Y. Chen and Y. Hao, "A feature weighted support vector machine and K-nearest neighbor algorithm for stock market indices prediction," *Expert Systems with Applications,* vol. 80, pp. 340-355, 2017.

[41] S. Barak and M. Modarres, "Developing an approach to evaluate stocks by forecasting effective features with data mining methods," *Expert Systems with Applications,* vol. 42, no. 3, pp. 1325-1339, 2015.

[42] M. Wen, P. Li, L. Zhang and Y. Chen, "Stock Market Trend Prediction Using High-Order Information of Time Series," *IEEE Access,* vol. 7, pp. 28299-28308, 2019.

[43] Y. Chen, W. Lin and J. Z. Wang, "A Dual-Attention-Based Stock Price Trend Prediction Model With Dual Features," *IEEE Access,* vol. 7, pp. 148047-148058, 2019.

[44] J. Long, Z. Chen, W. He, T. Wu and J. Ren, "An integrated framework of deep learning and knowledge graph for prediction of stock price trend: An application in Chinese stock exchange market," *Applied Soft Computing,* vol. 91, p. 106205, 2020.

[45] E. Hoseinzade and S. Haratizadeh, "CNNpred: CNN-based stock market prediction using a diverse set of variables," *Expert Systems with Applications,* vol. 129, pp. 273-285, 2019.

[46] Y. Baek and H. Y. Kim, "ModAugNet: A new forecasting framework for stock market index value with an overfitting prevention LSTM module and a prediction LSTM module," *Expert Systems with Applications,* vol. 113, pp. 457-480, 2018.

[47] M. Hiransha, E. A. Gopalakrishnan, V. K. Menon and K. P. Soman, "NSE Stock Market Prediction Using Deep-Learning Models," *Procedia Computer Science,* vol. 132, pp. 1351-1362, 2018.

[48] X. Zhou, Z. Pan, G. Hu, S. Tang and C. Zhao, "Stock Market Prediction on High-Frequency Data Using Generative Adversarial Nets," *Mathematical Problems in Engineering,* 2018.

[49] D. M. Q. Nelson, A. C. M. Pereira and R. A. d. Oliveira, "Stock market's price movement prediction with LSTM neural networks," in *2017 International Joint Conference on Neural Networks (IJCNN)*, Anchorage, AK, USA, 2017.

[50] C.-F. Tsai and Y.-C. Hsiao, "Combining multiple feature selection methods for stock prediction: Union, intersection, and multi-intersection approaches," *Decision Support Systems,* vol. 50, no. 1, pp. 258-269, 2010.

[51] S. Asadi, E. Hadavandi, F. Mehmanpazir and M. M. Nakhostin, "Hybridization of evolutionary Levenberg–Marquardt neural networks and data pre-processing for stock market prediction," *Knowledge-Based Systems,* vol. 35, pp. 245-258, 2012.

[52] C. M. Anish and B. Majhi, "Hybrid nonlinear adaptive scheme for stock market prediction using feedback FLANN and factor analysis," *Journal of the Korean Statistical Society,* vol. 45, no. 64–76, 2016.

[53] R. d. A. Araújo, "Translation Invariant Morphological Time-lag Added Evolutionary Forecasting method for stock market prediction," *Expert Systems with Applications,* vol. 38, no. 3, pp. 2835-2848, 2011.

[54] S. Jabin, "Stock Market Prediction using Feed-forward Artificial Neural Network," *International Journal of Computer Applications,* vol. 99, no. 9, pp. 4-8, 2014.

[55] R. d. A. Araújo and T. A. E. Ferreira, "A Morphological-Rank-Linear evolutionary method for stock market prediction," *Information Sciences,* vol. 237, pp. 3-17, 2013.

[56] L. Xi, H. Muzhou, M. H. Lee, J. Li, D. Wei, H. Hai and Y. Wu, "A new constructive neural network method for noise processing and its application on stock market prediction," *Applied Soft Computing,* vol. 15, pp. 57-66, 2014.

[57] K. S. Kannan, P. S. Sekar, M. Sathik and P. Arumugam, "Financial stock market forecast using data mining techniques," in *Proceedings of the International Multiconference of Engineers and computer scientists*, 2010.

[58] P. K. Aithal, A. U. Dinesh and M. Geetha, "Identifying Significant Macroeconomic Indicators for Indian Stock Markets," *IEEE Access,* vol. 7, pp. 143829-143840, 2019.

[59] R. Verma, P. Choure and U. Singh, "Neural networks through stock market data prediction," in *2017 International conference of Electronics, Communication and Aerospace Technology (ICECA)*, Coimbatore, India, 2017.

[60] E. Guresen, G. Kayakutlu and T. U. Daim, "Using artificial neural network models in stock market index prediction," *Expert Systems with Applications,* vol. 38, no. 8, pp. 10389-10397, 2011.

[61] E. L. de Faria, M. P. Albuquerque, J. L. Gonzalez, J. T. P. Cavalcante and M. P. Albuquerque, "Predicting the Brazilian stock market through neural networks and adaptive exponential smoothing methods," *Expert Systems with Applications,* vol. 36, no. 10, pp. 12506-12509, 2009.

[62] A. H. Moghaddam, M. H. Moghaddam and M. Esfandyari, "Stock market index prediction using artificial neural network," *Journal of Economics, Finance and Administrative Science,* vol. 21, no. 41, pp. 89-93, 2016.

[63] J. Patel, S. Shah, P. Thakkar and K. Kotecha, "Predicting stock market index using fusion of machine learning techniques," *Expert Systems with Applications,* vol. 42, no. 4, pp. 2162-2172, 2015.

[64] X. Yuan, J. Yuan, T. Jiang and Q. U. Ain, "Integrated Long-Term Stock Selection Models Based on Feature Selection and Machine Learning Algorithms for China Stock Market," *IEEE Access,* vol. 8, pp. 22672-22685, 2020.

[65] L. Zhang, N. Liu and P. Yu, "A Novel Instantaneous Frequency Algorithm and Its Application in Stock Index Movement Prediction," *IEEE Journal of Selected Topics in Signal Processing,* vol. 6, no. 4, pp. 311-318, 2012.

[66] F. Yang, Z. Chen, J. Li and L. Tang, "A novel hybrid stock selection method with stock prediction," *Applied Soft Computing,* vol. 80, pp. 820-831, 2019.

[67] M. Karhunen, "Algorithmic sign prediction and covariate selection across eleven international stock markets," *Expert Systems with Applications,* vol. 115, pp. 256-263, 2019.

[68] F. Zhou, H.-m. Zhou, Z. Yang and L. Yang, "EMD2FNN: A strategy combining empirical mode decomposition and factorization machine based neural network for stock market trend prediction," *Expert Systems with Applications,* vol. 115, pp. 136-151, 2019.

[69] R. Ramezanian, A. Peymanfar and S. B. Ebrahimi, "An integrated framework of genetic network programming and multi-layer perceptron neural network for prediction of daily stock return: An application in Tehran stock exchange market," *Applied Soft Computing,* vol. 82, p. 105551, 2019.

[70] H. Wang, "Research on application of fractional calculus in signal real-time analysis and processing in stock financial market," *Chaos, Solitons & Fractals,* vol. 128, pp. 92-97, 2019.

[71] G. Liu, Y. Wang, X. Chen, Y. Zhang and Y. Shang, "Forecasting volatility of the Chinese stock markets using TVP HAR-type models," *Physica A: Statistical Mechanics and its Applications,* vol. 542, p. 123445, 2020.

[72] T. H. Nguyen, K. Shirai and J. Velcin, "Sentiment analysis on social media for stock movement prediction," *Expert Systems with Applications,* vol. 42, no. 24, pp. 9603-9611, 2015.

[73] H. Wang, S. Lu and J. Zhao, "Aggregating multiple types of complex data in stock market prediction: A model-independent framework," *Knowledge-Based Systems,* vol. 164, pp. 193-204, 2019.

[74] M. Dang and D. Duong, "Improvement methods for stock market prediction using financial news articles," in *2016 3rd National Foundation for Science and Technology Development Conference on Information and Computer Science (NICS)*, Danang, Vietnam, 2016.

[75] X. Ding, Y. Zhang, T. Liu and J. Duan, "Deep Learning for Event-Driven Stock Prediction," in *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence (IJCAI 2015)*, 2015.

[76] X. Zhang, S. Qu, J. Huang, B. Fang and P. Yu, "Stock Market Prediction via Multi-Source Multiple Instance Learning," *IEEE Access,* vol. 6, no. 50720 - 50728. 2018, pp. 50720-50728, 2018.

ISSN 2502-3357 (online) | ISSN 2503-0477 (print)
8 (1) January 2022 58-80

[77] C. Chen, W. Dongxing, H. Chunyan and Y. Xiaojie, "Exploiting Social Media for Stock Market Prediction with Factorization Machine," in *2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)*, Warsaw, Poland, 2014.

[78] N. Oliveira, P. Cortez and N. Areal, "The impact of microblogging data for stock market prediction: Using Twitter to predict returns, volatility, trading volume and survey sentiment indices," *Expert Systems with Applications,* vol. 73, p. 125, 2017.

[79] X. Zhang, Y. Zhang, S. Wang, Y. Yao, B. Fang and P. S. Yu, "Improving stock market prediction via heterogeneous information fusion," *Knowledge-Based Systems,* vol. 143, pp. 236-247, 2018.

[80] N. P. N. Sreedharan, B. Ganesan, R. Raveendran, P. Sarala, B. Dennis and R. B. R, "Grey Wolf optimisation-based feature selection and classification for facial emotion recognition," *IET Biometrics,* vol. 7, no. 5, pp. 490-499, 2018.

[81] C. Hu and L. T. He, "An Application of Interval Methods to Stock Market Forecasting," *Reliable Computing,* vol. 13, p. 423–434, 2007.

[82] R. Chowdhury, M. R. C. Mahdy, T. N. Alam, G. D. A. Quaderi and M. A. Rahman, "Predicting the stock price of frontier markets using machine learning and modified Black–Scholes Option pricing model," *Physica A: Statistical Mechanics and its Applications,* vol. 555, p. 124444, 2020.