IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received March 11, 2021, accepted April 1, 2021, date of publication April 15, 2021, date of current version May 5, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3073507

Ensemble Technique With Optimal Feature Selection for Saudi Stock Market Prediction: A Novel Hybrid Red Deer-Grey Algorithm

SAUD S. ALOTAIBI

Department of Information Systems, College of Computer and Information Systems, Umm Al-Qura University, Mecca 21421, Saudi Arabia

e-mail: ssotaibi@uqu.edu.sa

ABSTRACT The forecast of the stock price attempts to assess the potential movement of the financial exchange's stock value. The exact estimation of the movement of share price would contribute more to investors' profit. This paper introduces a new stock market prediction model that includes three major phases: feature extraction, optimal feature selection, and prediction. Initially, statistical features like mean, standard deviation, variance, skewness, and kurtosis is extracted from the collected stock market data. Further, the indexed data collected are also computed concerning standard indicators like Average True Range (ATR), Exponential Moving Average (EMA), Relative Strength Index (RSI), and Rate of Change (ROC). To acquire best-predicted results, it is more crucial to select the most relevant features. Such that, the optimal features are selected from the extracted features (technical indicators based features, statistical features) by a new hybrid model referred to Red Deer Adopted Wolf Algorithm (RDAWA). Further, the selected features are subjected to the ensemble technique for predicting the stock movement. The ensemble technique involves the classifiers like Support Vector Machine (SVM), Random Forest1 (RF1), Random Forest2 (RF2), and optimized Neural Network (NN), respectively. The final predicted results are acquired from the Optimized Neural Network (NN). To make the precise prediction, the training of NN is carried out by the proposed RDAWA via fine-tuning the optimal weight. Finally, the performance of the proposed work is compared over other conventional models with respect to certain measures.

INDEX TERMS Saudi stock market prediction, close price, second order technical indicators, pre classifier.

NOMENCLATURE

Abbreviation	Description
ROC	Rate of Change
MS-VAR	Markov-switching vector autoregressive
RSI	Relative Strength Index
SMPPF	Stock Market Prices Prediction Framework
GA	Genetic Algorithm
RNN	recurring neural network
MM-HPA	Multi-Model based Hybrid Prediction Algorithm
EMA	Exponential Moving Average
SVM	Support Vector Machine
ATR	Average True Range
RF2	Random Forest2
NN	Neural Network
GWO	Grey Wolf Optimizer
RF1	Random Forest1
RDA	Red deer algorithm
HMM	Hidden Markov Model
KNN	K-Nearest Neighbor

The associate editor coordinating the review of this manuscript and approving it for publication was Christian Pilato.

I. INTRODUCTION

Saudi Arabia is a well-established country in the oil markets, and being the core member of the Organization of ''Petroleum Exporting Countries (OPEC), Saudi Aramco, the national oil and gas company'', by producing and maintaining billions of gallons of Saudi oil, including some 260 billion tonnes in inventory [1], [9]–[13]. ''Saudi Stock Exchange (SSE)'' is the largest exchange in the Middle East, identified locally by its Arabic name Tadawul. This is twice the size of the second-ranked Kuwait Stock Exchange by market capitalization. [14], [15]. Market fluctuations in Tadawul typically cause related movements [16]–[18] in share value or 'Market Capitalization or Market Cap'.

Nowadays, uncertainty characterizes the financial and economic environs, and thereby, understanding the predictability of future asset returns is the trending research topic for professionals and academics. The forecasting of the stock price attempts to assess the potential movement of a financial exchange's stock value. The accurate price movement would contribute more benefit to investors in production. However, the most challenging aspect is the forecasting of stock movement as it involves the factors like interest rates, politics, economic growth, etc. The hybrid optimization algorithms have been reported to be promising for certain search problems [19]. This makes the use of machine learning models for stock forecasting with the training of price fluctuations of days and even the minute. Models like RNN, HMM [2], [6] are more common in forecasting stock prices. To be more accurate, the Metaheuristic tactics are incorporated with the learning algorithm [6]. The major contribution of this research work is:

- Introduces the ensemble technique by leveraging the traditional classifiers like SVM, RF1, RF2, and optimized NN.
- A new RDAGW algorithm is introduced to train the NN by tuning the optimal weight.

The rest of the paper is organized as: Section II describes the contemporary works undergone in stock market prediction. Section III tells about the feature extraction process: SOTI and statistical features. Section IV represents the constructed ensemble classifier with RDAGW based optimized NN. The final results acquired are discussed in Section V. This paper is concluded in Section VI.

II. LITERATURE REVIEW

A. RELATED WORKS

In 2019, Tissaoui *et al.* [1] have conducted an empirical study on the predictability and complex contingent association among foreign volatility risk indices and Saudi stock returns. To test the long-run and short-run persistence of shocks on the complex conditional correlation, the authors have introduced a mixed regression framework. The "DCC-GARCH (1.1) and CCF-Approaches" were the basis of the proposed model. The Saudi market return cross-correlation analyses have indicated a greater presence of spreading shocks than the commodity markets.

In 2020, Trichilli *et al.* [2] have introduced an HMM-based on transition matrix to investigate the relationship between the Islamic index returns and the investor's sentiment in MENA countries. The transition matrix and the steady-state probabilities were estimated using the HMM. In Islamic market indexes, the uncertainties were captured using the proposed model. In addition, the possible effects of the dynamics of the Islamic market on the persistence of these regimes or States were also determined.

In 2019, Oueslati *et al.* [3] have examined the performance of a variety of RFV return forecasting variables and techniques in "Saudi Arabia and Malaysia". In Saudi Arabia, the market excess returns were documented for predicting the changes in oil prices, the dividend yield, and inflation. The diffusion index was utilized for forecasting the stock return predictability. The empirical results had exhibited the rationally time-varying expected returns of Saudi Arabia and Malaysia's irrational pricing.

In 2018, Chowdhury *et al.* [4] have examined the Saudi stock market's autocorrelation pattern corresponding to the stock returns from January 2004 to December 2015. The authors have also investigated the stock's autocorrelation structure and the portfolio returns of Saudi Arabia. Specifically, the authors have examined the relation of the Saudi Arabian stock market's autocorrelation return to factors like day of the week, stock trading, performance on a preceding day, and volatility.

In 2020, Hung [5] have utilized the Markov-switching autoregression to detect the RSB regime-shift behavior in the stock returns of the Gulf Arab countries. In addition, they have captured the dynamic interrelatedness existing among the stock returns and the exchange by MS-VAR model. Further, from the study, it was evident that both the low-volatility regime and the high-volatility regime had supported the perseverance for all markets.

In 2020, Polamur *et al.* [6] have proposed SMPPF for the accurate forecasting of the stock market price. The underlying algorithm behind the SMPPF was the MM-HPA, which was constructed by blending the linear (autoregressive moving average model) and non-linear models (RNN). Further, the optimal parameters were explored using the GA. The empirical results of the proposed framework had exhibited its supremacy in capturing the non-linear data.

In 2018, Zhang *et al.* [7] have proposed a stock prediction framework on the basis of the Hidden Markov model, where heterogeneous knowledge such as web news and historical quantitative data was combined. By implementing the market correlation information into this system, the event sparsity problem was alleviated. The proposed model was tested on the China A-share market dataset, and the results showed the best predictive performance.

In 2018, Chen *et al.* [8] have developed a basic hybridized framework with the amalgamation of the feature weighted SVM as well as feature weighted KNN for predicting the indices in the stock market effectively. Initially, the collected data were classified using the feature-weighted SVM. Then, by measuring the k-weighted nearest neighbors, the future stock market indexes were expected. The experiments have used Chinese stock market indices such as the Shanghai and Shenzhen stock exchange indices for assessing the system's effectiveness.

B. REVIEW

The stock market prediction seems to be an important activity and is of particular significance as correctly forecasting stock markets will lead to lucrative profits by establishing appropriate choices. Because of non-stationary, blaring, and chaotic data, stock market prediction is a big problem, and hence the prediction becomes impossible for shareholders to spend capital to make profits. It had long been one of the most challenging activities performed by humans to forecast the stock price. Thousands and thousands of hours were spent attempting to beat the competition reliably. Up to this point, nobody had succeeded, not even experienced investors who are right just about half the time. In the latest approaches, multiple strategies are developed to forecast stock market movements. When developing these hypotheses, several variables are considered, such as business fundamentals, climate, supply and demand, investor psychology, and so on. Yet some people think that with machine learning, optimism will be on the horizon, and its immense capacities will one day shortly defy this phenomenon and unlock the gate for people to riches. Machine learning is a system of data processing that, without relying on a preset equation, learns from experience using numerical data to 'learn' knowledge directly from data. In other words, the more and more data this is served, it gets smarter. Recently, several research studies have used machine learning to successfully forecast stock market changes. The ANN in [28] is utilized for both Classification and Forecasting purposes. It had the ability to handle complex non-linear patterns, noisy as well as missing data. But, here, the parameters are highly sensitive to overfitting problems. The Autoregressive integrated moving average model in [28] is utilized for clustering as well as forecasting purpose. This technique works goods for short-term forecasting and for the linear time series. But, it is incapable of handling the non-linear data, and this process is tedious too. In [28], the fuzzy C-Means is applicable for clustering-based financial data forecasting. This can be more significant for the small as well as medium datasets. Apart from this, it suffers from noise and isn't capable of handling huge datasets. In addition, the Generalized Regression Neural Network in [28] is easy to implement for forecasting the financial time-series data. Moreover, here the training approach is quicker and easier. But, this technique suffers from the drawbacks like huge memory consumption, and it is highly computationally complex in terms of time and cost. In addition, the Hidden Markov Model in [28] is significant in forecasting the higher dimensional data. But, here, the parameters are set manually and hence takes a huge time. The Particle Swarm Optimization utilized for forecasting stock market data is easy to implement and requires fewer parameters. But, this technique seems to be expensive. Moreover, the SVM in [28] provides an optimal global solution in forecasting time-series stock market data. This technique is sensitive to parameter selection as well as outliners. The Support Vector Regression [28] can handle huge data sets and can forecast time-series data more accurately. Apart from this advantage, it suffers from higher sensitivity in terms of user parameter selection.

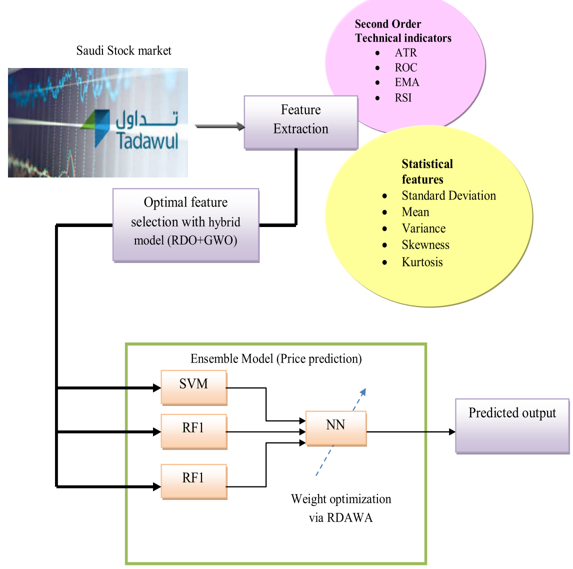


FIGURE 1. The architecture of the proposed work.

III. PROPOSED SAUDI STOCK MARKET PREDICTION MODEL: CONCEPTUAL DESCRIPTION

A. AN OVERVIEW

This paper introduces a new Saudi stock market prediction model, including three major phases: feature extraction, optimal feature selection, and prediction. The architecture of the proposed work is shown in Fig.1. Initially, the statistical and higher-order statistical features are extracted from the stock market data. Along with this, the standard indicator-based features like ATR, EMA, RSI, and ROC are also extracted, $F = f^{staticl} + f^{SOTIl}$. The feature dimension seems to be the greatest issue in any of the prediction models, which degrades the prediction accuracy. Hence, in this work, optimal features are selected using a new RDAGW algorithm. The selected features $F^{optimal}$ are subjected to ensemble technique for prediction purposes, which combines the SVM, RF1, RF2, and optimized NN, respectively. The final predicted results will be acquired from the Optimized NN, therefore to increase its prediction accuracy; the training will be carried out by the new RDAGW algorithm via tuning the optimal weights. The proposed hybrid optimization model will be the conceptual amalgamation of two standard optimization models: RDA and GWO, respectively.

IV. FEATURE EXTRACTION: STATISTICAL FEATURES AND TECHNICAL INDICATORS BASED FEATURES

This is the initial phase, where the statistical and higher-order statistical features, along with second-order technical indicator-based features (SOTI) get extracted.

A. SOTI BASED FEATURES

The Saudi stock market prices cannot be predicted accurately with the explanatory variables (e.g. transaction volume, google trend as well) when it comes to time series forecasting.

TABLE 1. Review on conventional stock market prediction models: Features and challenges.

Methods	Data	Purpose	Advantages	Disadvantages
ANNs: Artificial Neural network [28]	Non-time series, Time-series, and Financial time series	Classification and Forecasting	✓ High capacity to fix difficult non-linear pattern ✓ The model can help both linear and non-linear Model can help both linear and non-linear processes ✓ High precision for calculating the similarity in data categories ✓ stable and can accommodate noisy and incomplete data	❖ Overfitting Adaptive to the selection of parameters ❖ For any unknown data, ANNs just offer expected target values without any variance information to validate the forecast.
ARIMA: Autoregressive integrated moving average model [28]	Time-series, Financial time-series	Forecasting and Clustering	✓ Functioning well for linear time series ✓ It would be the most excellent economic forecasting tool. ✓ It offers more stable and accurate for short-run predictions than the comparable models of more complex structural frameworks.	❖ For non-linear time series, it does not function well. ❖ The model defined for one series would not be sufficient for another series. ❖ Needs more detail ❖ It takes a long time to process a big dataset for
FCM: Fuzzy c means [28]	Non-time series, Time-series, and Financial time series	Clustering	✓ It fits good when looking for spherical clusters ✓ Working successfully with small to medium datasets	❖ Sensitive to noise ❖ Issues in handling high dimensional datasets
GRNN: Generalized Regression Neural Network [28]	Non-time series, Time-series, and Financial time series	Classification and Forecasting	✓ Simple to introduce owing to a much quicker preparation phase than most ANNs. ✓ Useful for real-time forecast execution ✓ Engage in Fast Training	❖ More memory space is required to store the model. ❖ Because of its large scale, it can be computationally costly.
HMM: Hidden Markov Model [28]	Non-time series, Time-series, and Financial time series	Clustering, Classification, and Forecasting	✓ · Solid mathematical base ✓ Capable of high-level knowledge modeling	❖ Needs to set the parameters and is driven by personal assumptions that may be inaccurate, which resulting in incorrect clusters ❖ It takes some time to process a big dataset
PSO: Particle Swarm Optimization [28]	Non-time series, Time-series, and Financial time series	Forecasting	✓ Easy to implement ✓ Very few parameters to tweak	❖ Lacks a sound statistical framework to analyze the potential growth of related theories
SVM: Support Vector Machine[28]	Non-time series, Time-series, and Financial time series	Classification and Forecasting	✓ Can have the best global approach and have outstanding predictive accuracy potential ✓ Works on several problems in grouping	❖ Sensitive to outliers ❖ Sensitive to parameter selection
SVR: Support Vector Regression [28]	Time-series and Financial Time Series	Forecasting	✓ Especially suitable for multiple-input handling ✓ Provides elevated forecast accuracy ✓ Capacity to resolve the problem of overfitting	❖ Sensitive to users' defined free parameters

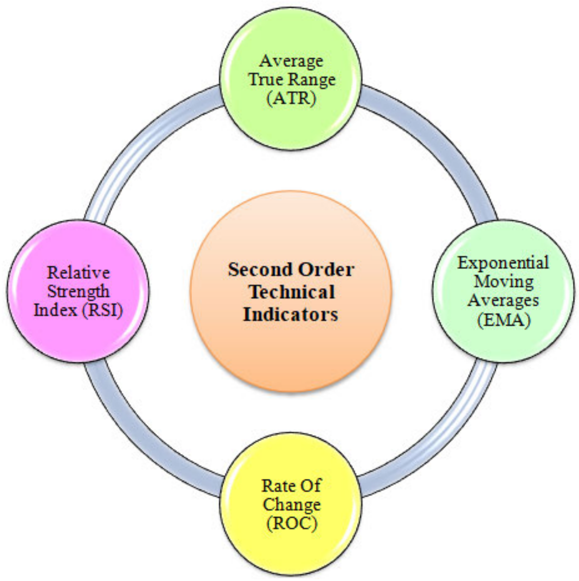


FIGURE 2. SOTI based features.

Therefore, the SOTI of the indexed data is to be computed for the stock prices, considering as the important feature for prediction. Fig.2 shows the SOTI based features. A brief description of these technical indicators is described below:

1) ATR [35]

This evaluation shows the sensitivity of the range of the period and tends to consider the certain disparity from the closure of the prior period. ATR may be used as a gauge for historical trail uncertainty in technical research. Though it is known as a lagging measure, it offers some visibility into 'where uncertainty is now? and where the last cycle was (day, week, month, etc.)?. Eq. (1) represents the statistical formula for ATR. Here, $True_{range}$ (true range) is the biggest of the three price differences:

$$ATR = Avg(True_{range}, time) \tag{1}$$

The price gap between yesterdays close to today's high, or yesterdays close to today's low, or the distance between today's low and HP is the largest of the three price differences (true range). The mathematical formula is shown in Eq. (2). The $True_{range}$ is defined in Eq. (3).

$$ATR = Avg(True_{range}, time) \tag{2}$$

$$True_{range} = \max(high - low), (high - P_{past}), (P_{past} - low) \tag{3}$$

2) EMA [36]

It is more receptive than simple moving averages to the current market price, so with EMAs, the lag time is less. The weighting assigned to the most recent price depends on the chosen moving average period. The shorter the EMA span, the more weight will be added to the latest price. For the present day $EMA_{p_{day}^{value}}$, the EMA can be computed as per Eq. (4).

$$EMA_{p_{day}^{value}} = \left[P_{day}^{value} * \left(\frac{S}{1 + Days}\right)\right] + EMA_{p_{yesterday}^{value}} * \left(1 - \left(\frac{S}{1 + Days}\right)\right) \tag{4}$$

In addition, $EMA_{p_{yesterday}^{value}}$ are the EMA value of the previous day, and the smoothing factor is symbolized as S.

3) RSI [37]

The ratio of the smoothed average of *time*-period gains to the absolute value of the smoothed average of *time*-period losses is the relative strength RS. It measures the strength of the trend, and the mathematical formula for RSI is shown in Eq. (5).

$$RSI = 100 - \left(\frac{100}{1 + RS}\right) \tag{5}$$

4) ROC OR MOMENTUM INDICATOR [38]

The Rate of Change (ROC) measure compares the current price to the previous price over the number of cycles chosen. The ROC measures the percentage change in price value between (C_{price}) and the price recorded a certain number of periods ago $(\Pr ice)$. This measure is often regarded as an indicator of momentum. For the specified period, the current price is (C_{price}) contrasted from the previous price. The ROC is defined as per Eq. (6).

$$ROC = ((C_{price}/\Pr ice) - 1.0) * 100 \tag{6}$$

The extracted SOTI based features are denoted as $f^{SOTI I} = \{ATR, EMA, RSI, ROC\}$.

B. STATISTICAL FEATURES

In addition to the SOTI based features, the statistical features are also extracted. The extracted statistical features are shown in Fig. 3.

1) STANDARD DEVIATION

By evaluating each data point's variation relative to the mean, the standard deviation is measured as the square root of variance. There has a larger variance within the data set if the data points are beyond the mean. Thus, the more the data is spaced out, the higher the standard deviation. The mathematical formula for standard deviation SD is shown in Eq. (7).

$$SD = \sqrt{\frac{\sum_{i=1}^{N} (Y_i - \overline{Y_i})^2}{N - 1}} \tag{7}$$

In which, Y_i is the i^{th} point in the collected input data, $\overline{Y_i}$ is the mean of the collected input data, and N is the count of data points in the collected input data.

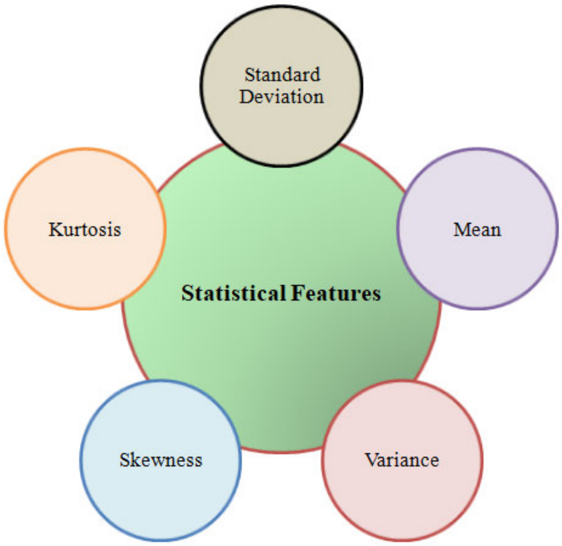


FIGURE 3. Statistical features.

2) MEAN

The mean is the first raw moment, and it is defined as $\overline{Y} = \frac{1}{N}\sum_{i=1}^{N} Y_i$

3) VARIANCE

In probability theory and mathematics, variance assumes a random variable's square deviation from its mean. The mathematical formula for variance is shown in Eq. (8).

$$Var(Y) = E\left[(Y-\mu)^2\right] \tag{8}$$

4) SKEWNESS

It is a measure of the asymmetry of the probability distribution of a real-valued random variable over its mean in probability theory and statistics. The value of skew can be positive, zero, negative, or unknown.

5) KURTOSIS

It is the mathematical calculation that indicates how often a distribution's tails vary from the normal distribution's tails. In other words, kurtosis determines when the extreme values are found in the tails of a given distribution. The computed statistical features are denoted as $f^{statistical}$. The final feature set is denoted as $F = f^{statistical} + f^{SOTII}$, from which the optimal features are extracted via the RDAWA model.

V. RDAWA FOR OPTICAL FEATURE SELECTION AND NN TRAINING

A. PROPOSED RDAWA

Feature dimensionality reduction is a significant step for better prediction results. In this work, the features of statistical and higher-order statistical features along with second-order technical indicator-based features (SOTI) get extracted. The optimal features F^{opt} are selected from the extracted features by the new RDWA model. Moreover, the training processing NN is also carried out using the same algorithm. The objective (fitness) of the algorithm is given in Eq. (33).

RDA was developed based on the Red Deers' mating behaviour. It is a high convergent model with less probability of getting trapped into the local optima. Similarly, the GWO was developed based on the hierarchical and hunting behaviour of the grey wolves. The GWO is good at solving complex optimization problems. Therefore, the RDA and GWO are hybridized in this research work. The hybrid optimization algorithms are more significant in solving different optimization issues in complex scenario [29]–[34]. The procedure of the proposed Red Deer Adopted Wolf Algorithm (RDAWA) is as follows: In the red deer algorithm, the solution gets updated using GWO, where α, β and δ are newly computed to determine the percentage of mating.

Step 1: Initialize the population of search agents as $pop = X_1, X_2, \ldots, X_M$. The current iteration is denoted as itr and the maximal count of iterations is denoted as Max^{itr}.

Step 2: Compute the fitness using Eq.(33).

Step 3: Sort the solutions based on fitness. The best deer is set as *Male* and the rest is set as *hind*. The best solution is denoted as X^*.

Step 4: While $itr < Max^{itr}$ do

Step 5: Move to roar male stage, where the male is roared as per Eq. (9). If the current solution is better than the existing one, then update the new position of the search agent.

$$New_{male} = \begin{cases} Old_{male} + a_1 * \left(\left(\begin{matrix} UB- \\ -LB \end{matrix}\right) * a_2\right) \\ \quad +LB \quad if\ a_3 \geq 0.5 \\ Old_{male} - a_1 * \left(\left(\begin{matrix} UB- \\ -LB \end{matrix}\right) * a_2\right) \\ \quad + +LB \quad if\ a_3 < 0.5 \end{cases} \tag{9}$$

Here, Old_{male} and New_{male} denotes the current and the new position of the red deer. In addition, a_1, a_2, a_3 is the randomly distributed integer between 0 to 1. In addition, UB and LB denotes the upper and lower bounds of solutions.

Step 6: Using Eq. (10) and Eq. (11), sort the males and then form the commanders and stags. Using Eq. (10), compute the male commanders count, and the stags count is computed as per Eq. (11).

$$Num_{Com} = round\ \{\gamma.Num_{male}\} \tag{10}$$

$$Num_{stag} = Num_{male} - Num_{Com} \tag{11}$$

Here, Num_{Com} and γ is the count of the male and initial value.

Step 7: The fight between male commanders and stags stage is undergone using Eq. (12) and Eq. (13). Then, update the position of the stags and male commanders.

$$New1 = \frac{(stag + Com)}{2} + b_1 * \left(\left(\begin{array}{c} Upper- \\ -Lower \end{array} \right) * b_2 \right) + Lower \quad (12)$$

$$New2 = \frac{(stag + Com)}{2} - b_1 * \left(\left(\begin{array}{c} Upper- \\ -Lower \end{array} \right) * b_2 \right) + Lower \quad (13)$$

Here, $New1$ and $New2$ denotes the 2 newly generated solutions by the fighting process. In addition, *Upper* and *Lower* denotes the upper and lower bounds in the search space.

Step 8: Form harems stage via Eq. (14), Eq. (16), and Eq. (17), respectively. Using Eq. (14) for the male commanders.

$$Norm_n = power_n - \max_i \{power_i\} \quad (14)$$

Here, $power_n$ denotes the n^{th} commander's power, and the normalized value is $Norm_n$. The normalized power is computed using Eq. (15). The hinds count of the harem is computed as per Eq. (16). Moreover, the harems commander is mated with ε percent of hinds as per Eq. (17). Here, $Num^{mate}_{harem_n}$ is the count of hinds.

$$power_n = \left| \frac{Norm_n}{\sum_{i=1}^{N_{Com}} Norm_i} \right| \quad (15)$$

$$Num_{harem} = round.\{power_n.Num_{hind}\} \quad (16)$$

$$Num^{mate}_{harem_n} = round.\{\varepsilon.Num_{hind}\} \quad (17)$$

Step 9: Randomly mate the selected commander male with the chosen harem using Eq. (18).

$$Offs = \frac{(hind + Com)}{2} + (Upper - Lower) * c \quad (18)$$

where, *Offs* denotes the new solution; and c is a random number.

Step 10: The harem is randomly selected, and it is denoted as κ

Step 11: In the harem, the count of hinds is computed as per Eq. (19).

$$Num_{Com} = round\left\{\delta.Num_{harem_k}\right\} \quad (19)$$

where Num_{harem_k} denotes the k^{th} hinds counts

Step 12: Randomly mate the selected commander male with the chosen harem using Eq. (19).

Step 13: Select the closest hind by computing the distance between the hinds and stag.

Step 14: With the selected hind, mate the stag using Eq. (19).

Step 15: Subsequently, update the position of the search agent using GWO using Eq. (20).

Step 16:

$$\vec{X}(tr+1) = \frac{\vec{X}_1 + \vec{X}_2 + \vec{X}_3}{3} \quad (20)$$

Here,

$$\begin{aligned} \vec{A}_\alpha &= |\vec{B}_1 \cdot \vec{X}_\alpha - \vec{X}|, \\ \vec{A}_\beta &= |\vec{B}_2 \cdot \vec{X}_\beta - \vec{X}|, \\ \vec{A}_\delta &= |\vec{B}_3 \cdot \vec{X}_\delta - \vec{X}| \end{aligned} \quad (21)$$

$$\begin{aligned} \vec{X}_1 &= \vec{X}_\alpha - \vec{C}_1 \cdot \left(\vec{A}_\alpha\right), \\ \vec{X}_2 &= \vec{X}_\beta - \vec{C}_2 \cdot \left(\vec{A}_\beta\right), \\ \vec{X}_3 &= \vec{X}_\delta - \vec{C}_3 \cdot \left(\vec{A}_\delta\right) \end{aligned} \quad (22)$$

Here, the assessment of $\vec{C}$ and $\vec{B}$ is defined in Eq. (23) and Eq. (24), in which, mi is linearly minimized from 2 to 0, $\vec{v}_1$ and $\vec{v}_2$ are the random vectors in the range [0, 1]. Moreover, the mathematical model of encircling behavior is in Eq.(25) and Eq. (26).

$$\vec{C} = 2mi \cdot \vec{v}_1 - mi \quad (23)$$

$$\vec{B} = 2 \cdot \vec{v}_2 \quad (24)$$

$$\vec{A} = |\vec{B} \cdot \vec{X}^P(tr) - \vec{X}(tr) \quad (25)$$

$$\vec{X}(tr+1) = \vec{X}^P(tr) - \vec{C} \cdot \vec{A} \quad (26)$$

As per the proposed logic, α is said to be the percentage of mating in a harem and β is the percentage of the mating of a commander with another harem. In addition, δ is the percentage of commanders. The values of α, β and δ are computed using Eq. (27), Eq. (28), and Eq. (29), respectively.

$$\alpha = 0.1 + 0.9 * \frac{iter}{\max(iter)} \quad (27)$$

$$\delta = 0.5 + 0.5 * \frac{iter}{\max(iter)} \quad (28)$$

$$\beta = 1 - \alpha \quad (29)$$

The extracted optimal feature is denoted as F^{opt}, which is fed as input to RF1, RF2, and SVM, respectively.

VI. PROPOSED ENSEMBLE TECHNIQUE FOR STOCK MARKET PREDICTION

Ensembles are also much more descriptive than the actual classifiers. In order to solve the same problem, the ensemble methods, also known as committee-based learning or learning multiple classifier systems, train several hypotheses. In general, the ensembles are good at enhancing the precision of the classifiers. It is a powerful "Meta characterization strategy". This is an effective "meta characterization strateg" which blends fragile beginners with vigorous newcomers to stimulate the efficacy of fragile beginners. The "ensemble strategy" is used in this research to improve the accuracy of the Saudi stock market prediction. The classifier is built in this research work by mixing the NN, RF1, RF2, and SVM, respectively. The extracted F is fed as input to RF1, RF2, and SVM, the output from SVM, RF1, and RF2 is fed as input to optimized NN.

A. RF1 AND RF2

The selected features F^{opt} are given as the input to the RF classifier. RF is an algorithm of labelling that combines the weak classifiers such as DT to develop a solid classifier. The community of trees is constructed by RF, and the classification is ultimately accomplished by having these trees to choose the most renowned class. In comparison, the normal CART decision, the tree utilizes the total features, and the randomness of features is guaranteed by the RF method. The RF model provides high precision of classification than the DT; however, the interpretability is not evident as the characteristics with an essential function are unrecognized. The resultant from RF1 and RF2 is Out^{RF1}, Out^{RF2}, respectively.

B. SVM

Meanwhile, the selected features F^{opt} are fed as the input to SVM. A non-linear mapping is used in the SVM to transform the normal training data into a high dimension. In its novel dimension, SVM investigated the linear optimum hyperplane. In addition, the decision boundary is a hyperplane that separates the tuples from one group to another. Moreover, through the margins and support vectors, the SVM has found the hyperplane. The hyperplane of2-class linear separable problem is calculated as per Eq. (30),

$$HP = VT^T Zi + dis = 0 \tag{30}$$

where dis indicates the distance among the origins to the hyperplane and VT denotes the normal vector. The output of SVM is denoted as Out^{SVM}.

C. OPTIMIZED NEURAL NETWORK

Optimized NN is responsible for providing the final classified result by getting the outputs of SVM and RFs (Out^{SVM}, Out^{RF1}, Out^{RF2}) as input denoted as Out. The network model of NN is given in Eq. (31), Eq. (32), and Eq. (3), respectively. As the main contribution, the training is carried out by the proposed RDAWA via optimizing the weights wg.

$$HIDDEN = AF\left(wg^N_{bias,hid} + \sum_{inp=1}^{N_{inp}} wg^N_{inp,hidz}.Out\right) \tag{31}$$

$$OUT_{pre} = AF\left(wg^r_{bias.out} + \sum_{hid=1}^{N_{hid}} wg^r_{hid.out}.HIDDEN\right) \tag{32}$$

$$F(er) = \underset{\{wg^N_{bias.hid}, wg^N_{inp,hid}, wg^r_{bias.out}, wg^r_{hid.out}\}}{\arg\min} \times \sum_{out=1}^{N_{out}} |Actual - \Pr edicted| \tag{33}$$

TABLE 2. Parameters of optimized NN.

Notation	Description
M	count of features
$inp = 1,2,\cdots N_{inp}$	input neuron
$hid = 1,2,\cdots N_{hid}$	hidden neurons
$out = 1,2,\cdots N_{out}$	output neuron
N_{inp}	count of input neuron
N_{hid}	count of hidden neuron
N_{out}	count of the output neuron
$wg^N_{inp,hid}$	weight from inp^{th} to hid^{th} is
$wg^N_{bias.hid}$	bias weight of hid
$wg^r_{bias.out}$	bias weight of out
wg^r_{zo}	weight from hid^{th} to out^{th}
$F(er)$	Error Function
$\Pr edicted$	the predicted output from NN
$Actual$	Actual output
AF	activation function

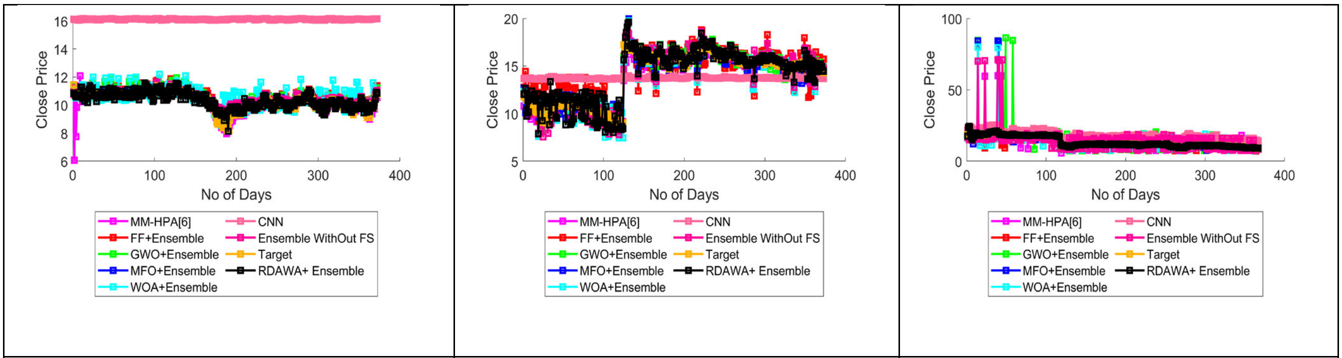
VII. RESULTS AND DISCUSSIONS

A. SIMULATION SETUP

The proposed Saudi stock market prediction model was implemented in MATLAB. The data for the evaluation was collected for three companies, namely: ALINMA company, DAR ALARKAN company, and SABIC company from [39]. The analysis was done by varying the window size from 7, 15, 20, respectively. The outcome determines the precise close price vales of the Saudi stock market on 8th day, 16th day, and 21st day, respectively. Here, the convergence analysis is made for RDAWA on ALINMA Company, DAR ALARKAN Company, and SABIC company datasets by varying the count of iterations. These evaluations will be undergone to verify the supremacy of RDAWA in fitness function achievement over existing approaches like FF+ Ensemble classifier, GWO + Ensemble classifier, MFO + Ensemble classifier, WOA + Ensemble classifier, and MM- HPA [6], respectively. Moreover, the evaluation is carried out in terms of error measures like RMSE, MSE, MAE, and MAPE, respectively.

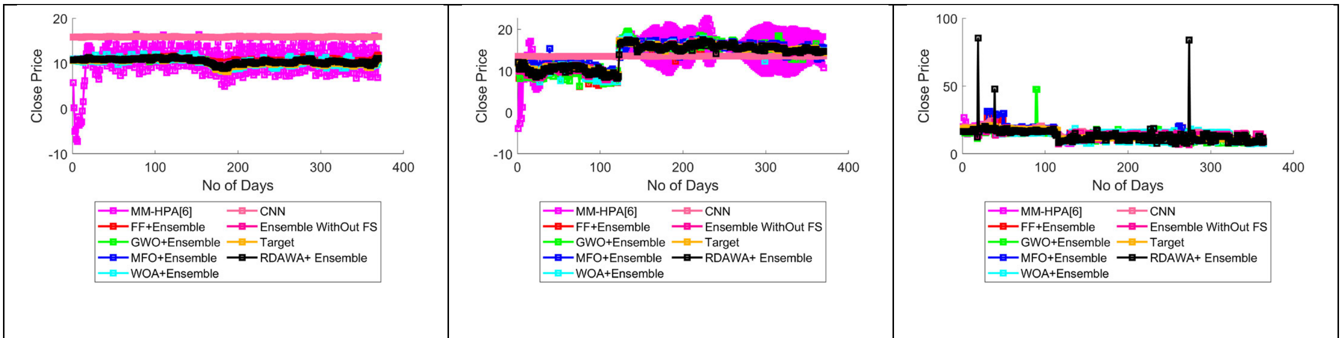
B. PREDICTION ANALYSIS

The predicted results using ALINMA company, DAR ALARKAN Company, and SABIC company are given for three varying window sizes. On observing the acquired results, it is clear that the predicted closed prices of the



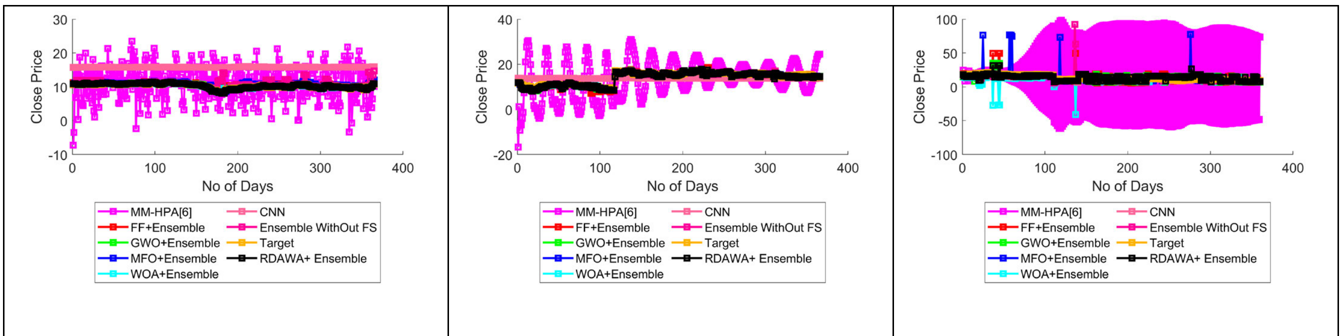
(a) (b) (c)

FIGURE 4. **Predictive analysis of the proposed as well as the existing works in case of ALINMA company datasets for (a) window size = 7, (b) window size = 15, and (c) window size = 30.**



(a) (b) (c)

FIGURE 5. **Predictive analysis of the proposed as well as the existing works in case of DAR ALARKAN company datasets for (a) window size = 7, (b) window size = 15, and (c) window size = 30.**



(a) (b) (c)

FIGURE 6. **Predictive analysis of the proposed as well as the existing works in case of SABIC company datasets for (a) window size = 7, (b) window size = 15, and (c) window size = 30.**

proposed work are closer to the target value, while compared to the existing models. The results acquired with ALINMA company, DAR ALARKAN Company, and SABIC company dataset are shown in Fig. 4, Fig.5, and Fig.6, respectively. This has been proved for all the window sizes.

C. CONVERGENCE ANALYSIS

The convergence analysis is undergone in this research work to exhibit the supremacy of the proposed RDAWA in achieving the desired objective as in Eq. (33). this result proves that 'how the proposed optimization algorithm converged for solving the given issue. Moreover, the convergence analysis is undergone by varying the count of the iteration from 10, 20, 30, 40, and 50, respectively. Since the RDAGW models were tested with the data collected from three different companies; the cost function is evaluated for each of them using the RDAWA. The cost function of the fitness function of RDAWA for ALINMA Company is shown in Fig. 7(a). Here, the cost function of all the techniques seems to be higher at the least count of iterations. As the count of iterations seems to increase after the 20^{th} iterations, a steep

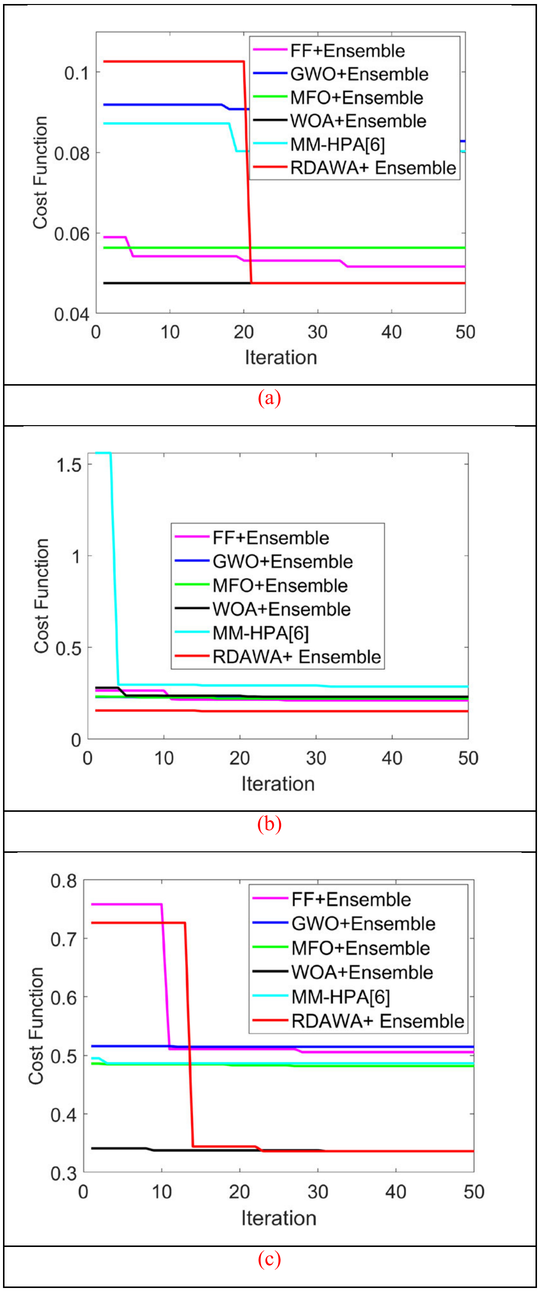


FIGURE 7. Convergence analysis of RDAGW+ Ensemble Classifier model over existing models for (a) ALINMA company, (b) DAR ALARKAN company, and (c) SABIC company.

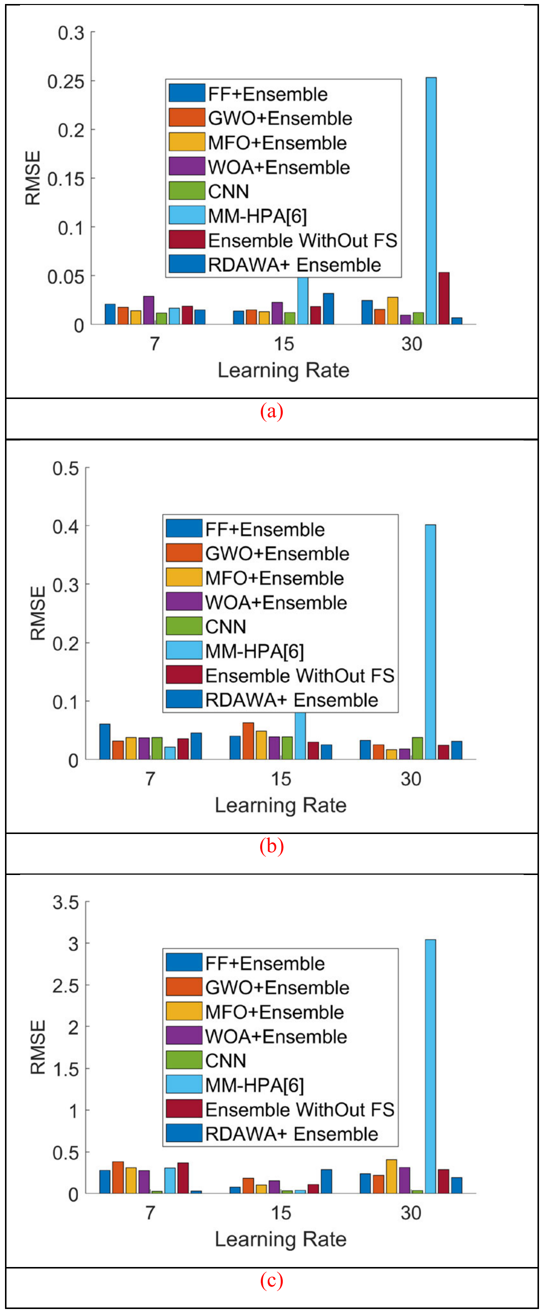


FIGURE 8. RMSE analysis of RDAGW+ Ensemble Classifier model over existing models for (a) ALINMA company, (b) DAR ALARKAN company, and (c) SABIC company.

fall in the cost function is recorded in the proposed work. Beyond the 20^{th} iteration, the proposed work had achieved the least cost function of 0.05. At 30^{th} iteration, the RDAWA is 23.22%, 13.5%, 14.2%, 12.5%, and 1.5% better than existing approaches like FF+ Ensemble Classifier, GWO+ Ensemble Classifier, MFO+ Ensemble Classifier, WOA+ Ensemble Classifier and MM-HPA. In addition, the convergence analysis of RDAWA was evaluated for DAR ALARKAN Company, and the corresponding results acquired are shown in Fig. 7(b). In this case, the RDAWA seems to maintain a fixed cost function overall variation in the count of iteration. Moreover, compared to all the acquired results, the least function is recorded by the proposed work. The convergence of RDAWA is at 0.3, which is the least value when compared to FF+ Ensemble classifier, GWO+ Ensemble Classifier, MFO+ Ensemble Classifier, WOA+ Ensemble Classifier, and MM-HPA. Further, on analyzing the convergence performance of SABIC Company (Fig. 7(c)), the proposed work had achieved the lowest cost function at the higher count of iterations. At 50^{th} iteration, the RDAWA is 23.2%, 23.5%, 24.2%, 12.5%, and 1.5% better than existing approaches like FF+ Ensemble Classifier, GWO+ Ensemble Classifier, MFO + Ensemble Classifier, WOA+ Ensemble Classifier and MM-HPA. Therefore, from the overall evaluation, a clear conclusion can be derived that the RDAWA perfectly suits for weight optimization in NN, while compared to existing approaches like FF+ Ensemble Classifier, GWO+ Ensemble Classifier, MFO+ Ensemble Classifier, WOA+ Ensemble Classifier, and MM-HPA [6], respectively.

D. ANALYSIS ON RMSE

The mathematical formula for RMSE is shown in Eq. (34)

$$RMSE = \sqrt{\frac{1}{M-N}\sum_{o=1}^{M-N}(Actual - \Pr edicted)^2} \quad (34)$$

Here, $Actual$ and $\Pr edicted$ is the actual and the predicted close price values, respectively.

The error minimization during the close price prediction is the major objective behind the current research work. Here, the RMSE of the proposed as well as existing works is evaluated for all the algorithms. This evaluation is undergone by varying the window size from 7, 15, and 30, respectively. The results acquired with ALINMA Company, DAR ALARKAN Company, and SABIC Company datasets are shown in Fig.8. In the case of ALINMA Company, the RSME of the proposed work at learning rate = 30 is ~0.01, which is the least value and it is 6.6%, 5.4%, 6.2%, 4.7%, 3.25%, 98% and 80% better than existing approaches like FF+ Ensemble Classifier, GWO+ Ensemble Classifier, MFO+ Ensemble Classifier, WOA+ Ensemble Classifier, MM-HPA, and Ensemble without FS. Then, in the case of DAR ALARKAN Company, the proposed work had recorded the least RMSE at learning rate = 15 and the corresponding value is ~0.015. In case of SABIC Company datasets at learning percentage = 7, the proposed work is 5%, 6.2%, 5.1%, 4.78%, 4.59%, 3.1%, 4% and 75% better than existing approaches like FF+ Ensemble Classifier, GWO+ Ensemble Classifier, MFO+ Ensemble Classifier, WOA+ Ensemble Classifier, MM-HPA [6], and Ensemble without FS. From the acquired results, it is clear that the RDAGW model had achieved the least RMSE value over every variation in the window size. This best performance is recorded by our RDAGW+ Ensemble Classifier model in ALINMA Company, DAR ALARKAN Company, and SABIC company datasets. Therefore, the RDAGW model is said to be good for predicting the close price of the Saudi stock market precisely.

E. ANALYSIS ON MAPE

The mathematical formula for MAPE is shown in Eq. (35)

$$MAPE = \frac{1}{M-N}\sum_{o=1}^{M-N}\left|\frac{Actual - \Pr edicted}{Actual}\right| \quad (35)$$

The MAPE of both the existing works is evaluated for ALINMA Company, DAR ALARKAN Company, and SABIC Company datasets. This evaluation is done by varying the window size from 7, 15, and 30, respectively. The results acquired with ALINMA company, DAR ALARKAN company, and SABIC company datasets are shown in Fig.9. As per the recorded results, the RDAGW model has recorded the least value over every variation in the window size. In case of ALINMA Company, the least MAPE is recorded by the RDAGW models at 30th window size. Among all the companies MAPE value, the least value is recorded in the SABIC Company datasets at window size = 7. At window size = 7, the proposed work for SABIC Company datasets is 80%, 85%, 62%, 60%, 70%, 55%, and 80% better than existing approaches like FF+ Ensemble Classifier, GWO+ Ensemble Classifier, MFO+ Ensemble Classifier, WOA+ Ensemble Classifier, MM-HPA and Ensemble without FS. Therefore, from the evaluation, it is clear that the RDAGW+ Ensemble

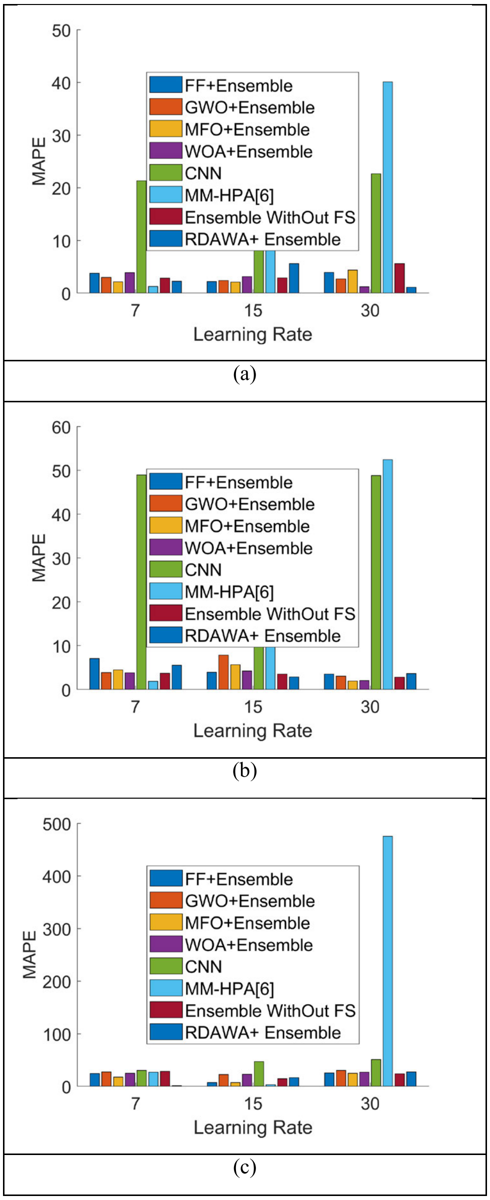


FIGURE 9. MAPE analysis of RDAGW+ Ensemble Classifier model over existing models for (a) ALINMA company, (b) DAR ALARKAN company, and (c) SABIC company.

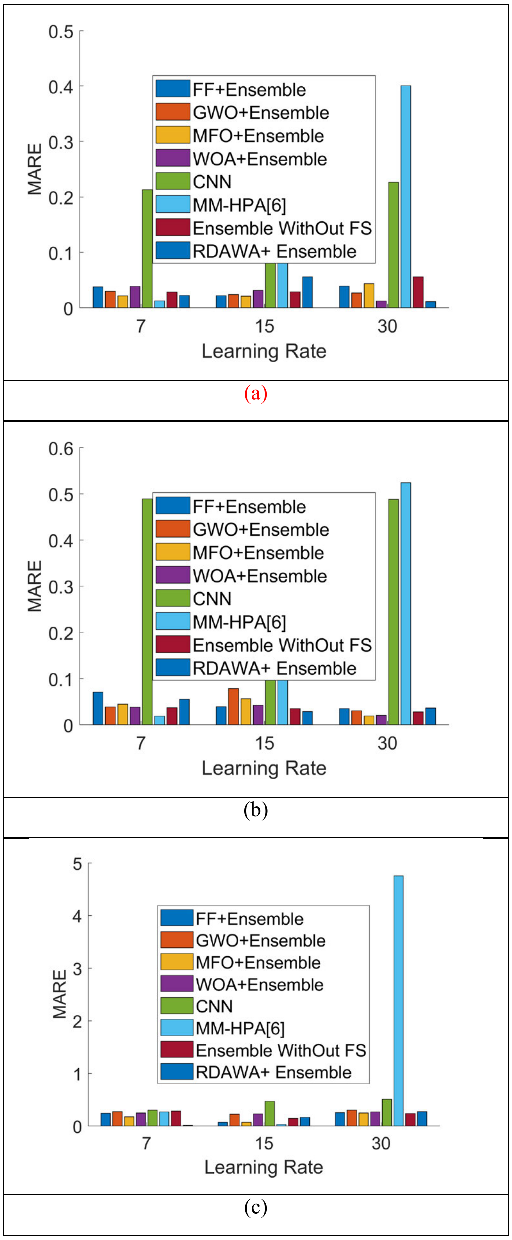


FIGURE 10. MARE analysis of+ Ensemble Classifier model over existing models for (a) ALINMA company, (b) DAR ALARKAN company, and (c) SABIC company.

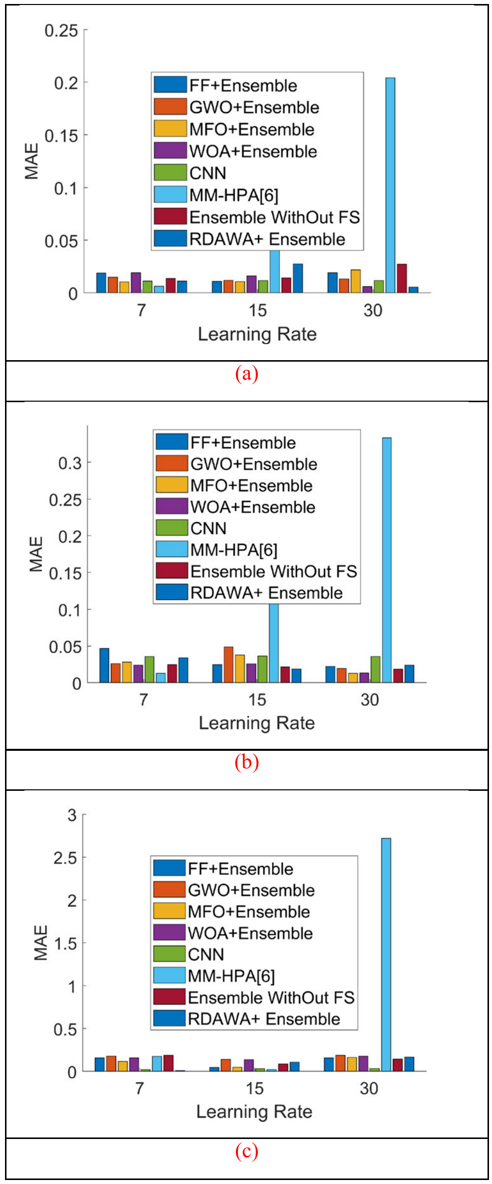


FIGURE 11. MAE analysis of RDAGW model over existing models for (a) ALINMA company, (b) DAR ALARKAN company, and (c) SABIC company.

Classifier model had achieved the least error value, and hence it becomes significant in predicting the closed price values with negligible errors.

F. ANALYSIS ON MARE

The MAPE value recorded for the proposed and the existing values on ALINMA Company, DAR ALARKAN Company, and SABIC Company datasets are shown in Fig. 10. On observing the ALINMA Company dataset, the least RDAGW model had recorded by the proposed algorithm almost for every variation in the window size. However, the least value is recorded by the proposed work at window size = 30. On observing the 30^{th} window size for ALINMA Company dataset, the RDAGW model is 90%, 87.5%, 92.5%, 60%, 97.5%, 98.75%, and 66.6% better than existing approaches like FF+ Ensemble Classifier, GWO+ Ensemble Classifier, MFO+ Ensemble Classifier, WOA+ Ensemble Classifier, MM-HPA [6] and Ensemble without FS. In the case of DAR ALARKAN Company, the MARE of the proposed work at window size = 15 is ~0.05, which is 16.6%, 44.4%, 32.8%, 28.5%, 90%, 65%, and 60% better than existing approaches like FF+ Ensemble Classifier, GWO+ Ensemble Classifier, MFO+ Ensemble Classifier, WOA+ Ensemble Classifier, MM-HPA, and Ensemble without FS. respectively. Therefore, from the evaluation, it is vivid that the proposed work had achieved the least MARE measure while compared to the existing ones.

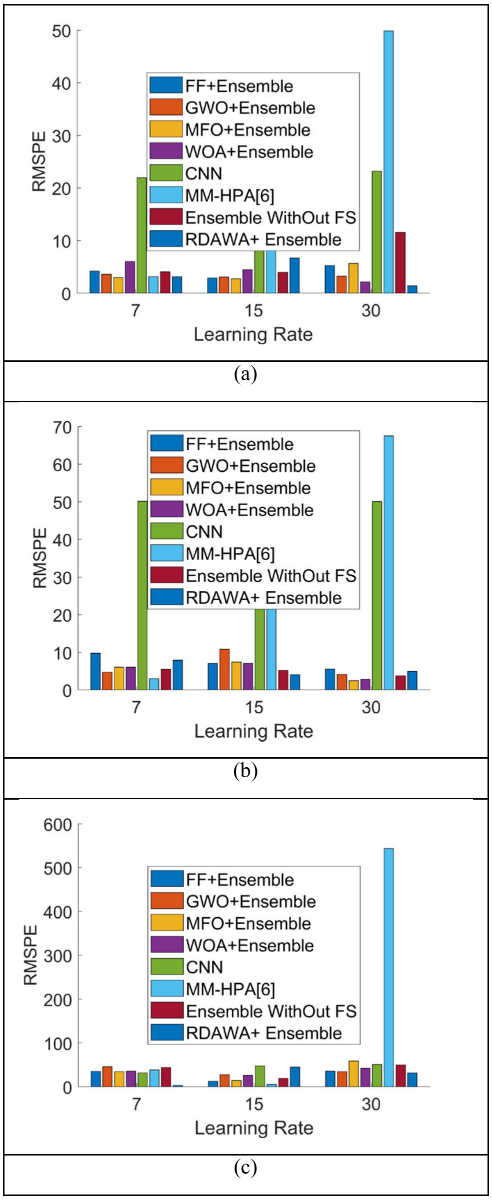


FIGURE 12. RMSPE analysis of+ Ensemble Classifier model over existing models for (a) ALINMA company, (b) DAR ALARKAN company, and (c) SABIC company.

G. ANALYSIS ON MAE

The mathematical formula for MAE is shown in Eq. (36)

$$MAE = \frac{1}{M-N}\sum_{o=1}^{M-N}\frac{|Actual - \Pr edicted|}{M-N} \tag{36}$$

The results acquired with ALINMA company, DAR ALARKAN company, and SABIC company datasets is shown in Fig.11. On analyzing the MAE values, the RDAGW model had recorded the best value (i.e. the least value) in ALINMA Company at window size = 30. Here the recorded MAE value is 0.02(∼), while the MAE of the existing techniques is greater than this. Then, in the case of DAR ALARKAN Company, the RDAGW model has achieved the least value at window size = 15. The MAE of the proposed work is 87.5%, 82%, 68.75%, 70%, 63%, 45%, and 37.5% better than existing approaches like FF+ Ensemble classifier, GWO+Ensemble classifier, MFO+Ensemble classifier, WOA+Ensemble classifier, MM- HPA, and Ensemble without FS, respectively at window size = 15. Further, among all the recorded values of the proposed work, the least value is recorded with SABIC Company datasets. Here, at window size = 7, the MAPE value is negligible. Therefore as a whole, the RDAGW model is said to be appropriate for close price prediction of the Saudi stock market.

H. ANALYSIS ON RMSPE

The RMSPE of existing as well as RDAGW model is evaluated with ALINMA Company, DAR ALARKAN Company, and SABIC Company datasets, and the corresponding values acquired in terms of closed price are shown in Fig. 12. On observing the acquired results, the RDAGW models have recorded the least value for every variation in the window size. While analyzing the SABIC company datasets, the RDAGW model had exhibited the least value as 25, which is 50.5%, 58%, 56%, 55.4%, 53%, 60%, and 24.15% better than existing approaches like FF+ Ensemble classifier, GWO+ Ensemble classifier, MFO+ Ensemble classifier, WOA+ Ensemble classifier, MM- HPA [6], and Ensemble without FS respectively. Moreover, while analyzing the other company datasets, the same enhanced performance was recorded in the RDAGW model. Therefore, it is suggested for close price prediction.

I. STATISTICAL ANALYSIS

The statistical evaluation is undergone in terms of Standard Deviation, mean case scenario, median case scenario, best-case scenario, and worst-case scenario, respectively, by executing the optimization algorithms for several times. This evaluation is done by varying the window size from 7, 15, and 30, respectively. The statistical analysis of ALINMA company, DAR ALARKAN company, and SABIC company datasets for varying window size = 7, 15, and 30 is tabulated in Table 3 to Table 9. On observing the mean performance of ALINMA Company at window size = 7, the RDAGW model had achieved the least value as 1.0762, which is 33.3%, 18.45, 3.75%, 46.3%, 87.6% and 22.59% better than existing approaches like FF+ Ensemble Classifier, GWO + Ensemble Classifier, MFO + Ensemble Classifier, WOA + Ensemble Classifier, MM-HPA, and Ensemble without FS, respectively. Also, the ALINMA Company had recorded the least mean performance value at window size = 15 and 30 as 2.47 and 0.50189, respectively. Moreover, while analyzing the mean performance of ALARKAN Company, the proposed work had achieved the least value for all three variations in the window size. For ALARKAN Company, the mean performance recorded by the RDAGW+ Ensemble Classifier model at window size = 15 is 1.4074, which the existing techniques have recorded the mean values as FF+ Ensemble Classifier = 2.2192, GWO + Ensemble Classifier = 3.7685, MFO + Ensemble Classifier = 2.6291,

TABLE 3. Statistical analysis of RDAGW model over existing techniques for ALINMA company at window size = 7.

Methods	Standard Deviation	Mean	Median	Best	Worst
FF+ Ensemble Classifier	2.182	1.6146	0.037623	0.018895	4.2335
GWO+ Ensemble Classifier	1.7889	1.319	0.029829	0.014985	3.5497
MFO+ Ensemble Classifier	1.4293	1.0373	0.021585	0.010543	2.9818
WOA+ Ensemble Classifier	2.8133	2.0058	0.038871	0.019202	6.055
CNN	11.808	8.6999	0.21316	0.011129	21.948
MM-HPA[6]	1.3706	0.88302	0.016591	0.0062958	3.1441
Ensemble Without FS	1.9269	1.3904	0.028246	0.013669	4.0667
RDAGW+ Ensemble Classifier	1.4855	1.0762	0.022199	0.011105	3.1127

TABLE 4. Statistical analysis of RDAGW model over existing techniques for ALINMA company at window size = 15.

Methods	Standard Deviation	Mean	Median	Best	Worst
FF+ Ensemble Classifier	1.398	1.0222	0.02203	0.010862	2.8616
GWO+ Ensemble Classifier	1.4991	1.0978	0.023832	0.01191	3.0553
MFO+ Ensemble Classifier	1.339	0.97974	0.02118	0.010593	2.7357
WOA+ Ensemble Classifier	2.1204	1.5336	0.031361	0.01606	4.4619
CNN	12.423	9.1536	0.22474	0.011716	23.046
MM-HPA[6]	16.731	12.174	0.25554	0.13084	34.75
Ensemble Without FS	1.9091	1.3867	0.028981	0.014093	3.974
RDAGW+ Ensemble Classifier	3.3524	2.47	0.055667	0.027542	6.6683

TABLE 5. Statistical analysis of RDAGW model over existing techniques for ALINMA company at window size = 30.

Methods	Standard Deviation	Mean	Median	Best	Worst
FF+ Ensemble Classifier	2.5301	1.8426	0.038946	0.01915	5.2356
GWO+ Ensemble Classifier	1.6206	1.1933	0.026787	0.013283	3.2324
MFO+ Ensemble Classifier	2.7577	2.0182	0.043642	0.021849	5.6335
WOA+ Ensemble Classifier	0.97037	0.67818	0.012281	0.0060439	2.1348
CNN	12.508	9.2168	0.22633	0.011791	23.2
MM-HPA[6]	24.707	18.152	0.40053	0.20417	49.847
Ensemble Without FS	5.1201	3.4573	0.056118	0.027256	11.538
RDAGW+ Ensemble Classifier	0.68474	0.50189	0.010942	0.0055089	1.392

TABLE 6. Statistical analysis of RDAGW model over existing techniques for DAR ALARKAN company at window size = 7.

Methods	Standard Deviation	Mean	Median	Best	Worst
FF+ Ensemble Classifier	4.6783	3.4054	0.070857	0.046561	9.7629
GWO+ Ensemble Classifier	2.3617	1.7417	0.038629	0.025915	4.749
MFO+ Ensemble Classifier	2.8927	2.1107	0.04442	0.028245	6.0009
WOA+ Ensemble Classifier	2.7939	1.993	0.038334	0.024101	6.032
CNN	27.005	19.907	0.4889	0.035605	50.081
MM-HPA[6]	1.3912	0.98704	0.021295	0.012962	3.03
Ensemble Without FS	2.5695	1.8543	0.03716	0.024715	5.458
RDAGW+ Ensemble Classifier	3.7358	2.7031	0.055018	0.033926	7.8793

TABLE 7. Statistical analysis of RDAGW model over existing techniques for DAR ALARKAN company at window size = 15.

Methods	Standard Deviation	Mean	Median	Best	Worst
FF+ Ensemble Classifier	3.1908	2.2192	0.040056	0.024568	7.0664
GWO+ Ensemble Classifier	5.1789	3.7685	0.078555	0.048658	10.797
MFO+ Ensemble Classifier	3.5901	2.6291	0.056169	0.03771	7.3863
WOA+ Ensemble Classifier	3.2318	2.2828	0.04256	0.025827	7.051
CNN	27.491	20.265	0.49805	0.036259	50.947
MM-HPA[6]	12.006	8.8055	0.18866	0.13592	24.663
Ensemble Without FS	2.4419	1.758	0.035066	0.021687	5.1967
RDAGW+ Ensemble Classifier	1.9387	1.4074	0.028953	0.018654	4.0686

WOA + Ensemble Classifier = 2.2828, CNN = 20.265, MM-HPA [6] = 8.8055, and Ensemble without FS = 1.758. Moreover, in the case of SABIC Company datasets, the mean performance of RDAGW+ Ensemble Classifier model is 11.956, which is 3.1%, 8.6%, 29.45, 14.2%, 42.9%, 94.25%, and 19.23% better than existing approaches like FF+ Ensemble Classifier, GWO+ Ensemble Classifier, MFO+ Ensemble Classifier, WOA+ Ensemble Classifier, CNN, MM-HPA [6], and Ensemble without FS, respectively. Therefore, the proposed work is suggested as the best technique for close price prediction.

TABLE 8. Statistical analysis of RDAGW model over existing techniques for DAR ALARKAN company at window size = 30.

Methods	Standard Deviation	Mean	Median	Best	Worst
FF+ Ensemble Classifier	2.5604	1.824	0.034952	0.022165	5.5353
GWO+ Ensemble Classifier	1.9845	1.4465	0.030314	0.019518	4.126
MFO+ Ensemble Classifier	1.2057	0.88414	0.01898	0.012836	2.4739
WOA+ Ensemble Classifier	1.3505	0.98443	0.020605	0.013399	2.8097
CNN	26.96	19.873	0.48792	0.035663	50.012
MM-HPA[6]	33.026	24.223	0.52402	0.33338	67.456
Ensemble Without FS	1.8196	1.3292	0.028075	0.018488	3.7674
RDAGW+ Ensemble Classifier	2.3691	1.7294	0.036441	0.023843	4.9113

TABLE 9. Statistical analysis of RDAGW model over existing techniques for SABIC company at window size = 7.

Methods	Standard Deviation	Mean	Median	Best	Worst
FF+ Ensemble Classifier	16.752	12.137	0.27718	0.15893	35.418
GWO+ Ensemble Classifier	20.842	14.726	0.38093	0.17915	45.667
MFO+ Ensemble Classifier	15.467	10.555	0.30924	0.11599	34.823
WOA+ Ensemble Classifier	16.868	12.213	0.27475	0.15751	35.695
CNN	16.963	12.501	0.30459	0.021828	31.693
MM-HPA[6]	18.193	13.188	0.30503	0.17557	38.446
Ensemble Without FS	20.416	14.604	0.36683	0.18714	44.093
RDAGW+ Ensemble Classifier	1.4115	0.90117	0.029128	0.0090114	3.253

TABLE 10. Statistical analysis of RDAGW model over existing techniques for SABIC company at window size = 15.

Methods	Standard Deviation	Mean	Median	Best	Worst
FF+ Ensemble Classifier	5.8778	4.1057	0.076727	0.046205	12.974
GWO+ Ensemble Classifier	13.737	10.129	0.22511	0.14125	27.584
MFO+ Ensemble Classifier	6.4151	4.3789	0.10209	0.048692	14.417
WOA+ Ensemble Classifier	13.485	9.9748	0.22826	0.13862	26.527
CNN	25.73	18.969	0.46998	0.03187	47.31
MM-HPA[6]	2.5392	1.7664	0.040643	0.020892	5.6346
Ensemble Without FS	9.2467	6.7662	0.14508	0.085922	18.984
RDAGW+ Ensemble Classifier	19.694	12.507	0.28679	0.10685	45.357

TABLE 11. Statistical analysis of RDAGW model over existing techniques for SABIC company at window size = 30.

Methods	Standard Deviation	Mean	Median	Best	Worst
FF+ Ensemble Classifier	16.988	12.349	0.25524	0.15891	35.569
GWO+ Ensemble Classifier	17.657	13.085	0.30247	0.18827	34.468
MFO+ Ensemble Classifier	25.739	16.937	0.40624	0.16519	58.777
WOA+ Ensemble Classifier	19.521	13.942	0.31108	0.18021	42.187
CNN	28.005	20.646	0.5122	0.034202	51.427
MM-HPA[6]	278.24	205.93	4.7531	2.7177	543.81
Ensemble Without FS	21.986	14.803	0.28769	0.1437	49.695
RDAGW+ Ensemble Classifier	16.163	11.956	0.27307	0.16735	31.841

VIII. CONCLUSION

This paper introduces a new stock market prediction model that includes three major phases: feature extraction, optimal feature selection, and prediction. The features like the statistical features and SOTI features were extracted. The classifiers in the prediction phase were trained with the extracted features. To acquire best-predicted results, it is more crucial to select the most relevant features. The optimal features were selected from the extracted features by RDAGW. As the major contribution, the prediction process was made via an ensemble-based classification model that includes SVM, RF1, RF2, and optimized NN, respectively. The final predicted results were acquired from the Optimized NN, therefore to increase the prediction accuracy of NN; its weights were fine-tuned via RDAGW. The proposed hybrid optimization model was the conceptual amalgamation of two standard optimization models: RDA and GWO. On observing the 30th window size for the ALINMA Company dataset, the RDAGW model is 90%, 87.5%, 92.5%, 60%, 97.5%, and 98.75% better than existing approaches like FF+ Ensemble Classifier, GWO+ Ensemble Classifier, MFO+ Ensemble Classifier, WOA+ Ensemble Classifier and MM-HPA, respectively in terms of MAPE. Similar to this, the RDAGW model had recorded the least value for every variation in the window size. Therefore, it is suggested as an appropriate technique for close price prediction of the Saudi stock market effectively.

REFERENCES

[1] K. Tissaoui and J. Azibi, "International implied volatility risk indexes and Saudi stock return-volatility predictabilities," *North Amer. J. Econ. Finance*, vol. 47, pp. 65–84, Jan. 2019.

[2] Y. Trichilli, M. B. Abbes, and A. Masmoudi, "Predicting the effect of Googling investor sentiment on Islamic stock market returns: A five-state hidden Markov model," *Int. J. Islamic Middle Eastern Finance Manage.*, vol. 13, no. 2, pp. 165–193, Feb. 2020.

[3] A. Oueslati and Y. Hammami, "Forecasting stock returns in Saudi Arabia and Malaysia," *Rev. Accounting Finance*, vol. 17, no. 2, pp. 259–279, May 2018.

[4] S. S. H. Chowdhury, M. A. Rahman, and M. S. Sadique, "Stock return autocorrelation, day of the week and volatility: An empirical investigation on the Saudi Arabian stock market," *Rev. Accounting Finance*, vol. 16, no. 2, pp. 218–238, May 2017.

[5] N. T. Hung, "Stock market volatility and exchange rate movements in the Gulf Arab countries: A Markov-state switching model," *J. Islamic Accounting Bus. Res.*, vol. 11, no. 9, pp. 1969–1987, Aug. 2020.

[6] S. R. Polamuri, K. Srinivas, and A. K. Mohan, "Multi model-based hybrid prediction algorithm (MM-HPA) for stock market prices prediction framework (SMPPF)," *Arabian J. Sci. Eng.*, vol. 45, no. 12, pp. 10493–10509, Dec. 2020.

[7] X. Zhang, Y. Li, S. Wang, B. Fang, and P. S. Yu, "Enhancing stock market prediction with extended coupled hidden Markov model over multi-sourced data," *Knowl. Inf. Syst.*, vol. 61, no. 2, pp. 1071–1090, Nov. 2019.

[8] Y. Chen and Y. Hao, "A feature weighted support vector machine and K-nearest neighbor algorithm for stock market indices prediction," *Expert Syst. Appl.*, vol. 80, pp. 340–355, Sep. 2017.

[9] C. Bousoño-Calzón, H. Molina-Bulla, J. J. Escudero-Garzás, and F. J. Herrera-Gálvez, "Expert selection in prediction markets with homological invariants," *IEEE Access*, vol. 6, pp. 32226–32239, 2018, doi: 10.1109/ACCESS.2018.2846878.

[10] 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, doi: 10.1109/ACCESS.2019.2901842.

[11] A. A. Hasan and A. C. Fong, "Sentiment analysis based fuzzy decision platform for the Saudi stock market," in *Proc. IEEE Int. Conf. Electro/Inf. Technol. (EIT)*, Rochester, MI, USA, May 2018, pp. 0023–0029.

[12] W. Mensi, "Global financial crisis and co-movements between oil prices and sector stock markets in Saudi Arabia: A VaR based wavelet," *Borsa Istanbul Rev.*, vol. 19, no. 1, pp. 24–38, Mar. 2019.

[13] S. Sharif, "How foreign investors influence stock markets? The Saudi Arabian experience," *Middle East Develop. J.*, vol. 11, no. 1, pp. 105–123, Jan. 2019.

[14] X. Zhang, S. Qu, J. Huang, B. Fang, and P. Yu, "Stock market prediction via multi-source multiple instance learning," *IEEE Access*, vol. 6, pp. 50720–50728, 2018.

[15] M. A. Finta, B. Frijns, and A. Tourani-Rad, "Volatility spillovers among oil and stock markets in the US and Saudi Arabia," *Appl. Econ.*, vol. 51, no. 4, pp. 329–345, Jan. 2019.

[16] V. S. H. Wong and S. E. Massah, "Recent evidence on the oil price shocks on Gulf cooperation council stock markets," *Int. J. Econ. Bus.*, vol. 25, no. 2, pp. 297–312, May 2018.

[17] G. Al Gahtani, C. A. Bollino, S. Bigerna, and A. Pierru, "Estimating the household consumption function in Saudi Arabia: An error correction approach," *Appl. Econ.*, vol. 52, no. 11, pp. 1259–1271, Mar. 2020.

[18] J. Lee, R. Kim, Y. Koh, and J. Kang, "Global stock market prediction based on stock chart images using deep Q-network," *IEEE Access*, vol. 7, pp. 167260–167277, 2019, doi: 10.1109/ACCESS.2019.2953542.

[19] S. S. Alotaibi, "Optimization insisted watermarking model: Hybrid firefly and Jaya algorithm for video copyright protection," *Soft Comput.*, vol. 24, pp. 14809–14823, Mar. 2020.

[20] 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, doi: 10.1109/ACCESS.2020.2969293.

[21] L. Chen, Z. Qiao, M. Wang, C. Wang, R. Du, and H. E. Stanley, "Which artificial intelligence algorithm better predicts the Chinese stock market?" *IEEE Access*, vol. 6, pp. 48625–48633, 2018, doi: 10.1109/ACCESS.2018.2859809.

[22] S. Bouktif, A. Fiaz, and M. Awad, "Augmented textual features-based stock market prediction," *IEEE Access*, vol. 8, pp. 40269–40282, 2020, doi: 10.1109/ACCESS.2020.2976725.

[23] M. Nabipour, P. Nayyeri, H. Jabani, S. Shahab, and A. Mosavi, "Predicting stock market trends using machine learning and deep learning algorithms via continuous and binary data; a comparative analysis," *IEEE Access*, vol. 8, pp. 150199–150212, 2020, doi: 10.1109/ACCESS.2020.3015966.

[24] 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, doi: 10.1109/ACCESS.2019.2895252.

[25] C. Bousoño-Calzón, J. Bustarviejo-Muñoz, P. Aceituno-Aceituno, and J. J. Escudero-Garzás, "On the economic significance of stock market prediction and the no free lunch theorem," *IEEE Access*, vol. 7, pp. 75177–75188, 2019.

[26] A. M. Fathollahi-Fard and M. Hajiaghaei-Keshteli, "Red Deer algorithm (RDA); a new optimization algorithm inspired by Red Deer's mating," in *Proc. Int. Conf. Ind. Eng.*, 2016, pp. 33–34.

[27] S. Mirjalili, S. M. Mirjalili, and A. Lewis, "Grey wolf optimizer," *Adv. Eng. Softw.*, vol. 69, pp. 46–61, Mar. 2014.

[28] M. Obthong, N. Tantisantiwong, W. Jeamwatthanachai, and G. Wills, "A survey on machine learning for stock price prediction: Algorithms and techniques," in *Proc. 2nd Int. Conf. Finance, Econ., Manage. IT Bus.*, vol. 1. Prague, Czech Republic: Vienna House Diplomat Prague, May 2020, pp. 63–71, doi: 10.5220/0009340700630071.

[29] M. M. Beno, I. R. Valarmathi, S. M. Swamy, and B. R. Rajakumar, "Threshold prediction for segmenting tumour from brain MRI scans," *Int. J. Imag. Syst. Technol.*, vol. 24, no. 2, pp. 129–137, Jun. 2014, doi: 10.1002/ima.22087.

[30] V. Tejaswini and D. D. Susitra, "Hybrid PSO-WOA for solving ORPD problem under unbalanced conditions," *J. Comput. Mech., Power Syst. Control*, vol. 2, no. 2, pp. 10–20, Apr. 2019.

[31] K. S. G. Devi, "Hybrid genetic algorithm and particle swarm optimization algorithm for optimal power flow in power system," *J. Comput. Mech., Power Syst. Control*, vol. 2, no. 2, pp. 31–37, Apr. 2019.

[32] R. Kumar, "Hybrid cat swarm and crow search algorithm to solve the combined economic emission dispatch model for smart grid," *J. Comput. Mech., Power Syst. Control*, vol. 2, no. 3, pp. 10–18, Jul. 2019.

[33] R. P. Nair and P. Kanakasabapathy, "Hybrid PSO-BF algorithm for economic dispatch of a power system," *J. Comput. Mech., Power Syst. Control*, vol. 2, no. 4, pp. 28–37, Oct. 2019.

[34] G. Rajeshkumar and T. P. Sujatha, "Optimal positioning and sizing of distributed generators using hybrid MFO-WC algorithm," *J. Comput. Mech., Power Syst. Control*, vol. 2, no. 4, pp. 19–27, Oct. 2019.

[35] *ATR*. Accessed: Jun. 5, 2020. [Online]. Available: https://www.tradingtechnologies.com/xtrader-help/x-study/technical-indicator-definitions/average-true-range-atr/

[36] *EMA*. Accessed: Jun. 5, 2020. [Online]. Available: https://www.tradingtechnologies.com/xtrader-help/x-study/technical-indicator-definitions/exponential-ema/

[37] *RSI*. Accessed: Jun. 5, 2020. [Online]. Available: https://www.tradingtechnologies.com/xtrader-help/x-study/technical-indicator-definitions/relative-strength-indicator-rsi/

[38] *ROC*. Accessed: Jun. 5, 2020. [Online]. Available: https://www.tradingtechnologies.com/xtrader-help/x-study/technical-indicator-definitions/rate-of-change-roc/

[39] *The Datas of ALINMA Company, DAR ALARKAN Company, and SABIC Company*. Accessed: Jan. 25, 2021. [Online]. Available: https://www.tadawul.com.sa/wps/portal/tadawul/markets/equities/market-watch/market-performance/!ut/p/z1/pZDLCsIwEEU_KTcP27oMqTYlsVDUWrORrCSg1YX4_UbBhaCx4t0NnMPMXOJIT9zgr2HvL-E0-EOcty7bNdoqjYIZaFZCZjNjSjVhWHKySQJKEDfGZ0wVdCpgYXMagQp1uxAcFf3Rr-omh2yl7ubdKvr8Px8j78eHSHz3XRK5N_gAUhW9Am86SAJ4rkh8cT6uY3qEOsgbhnvh9A!!/p0/IZ7_NHLCH082K85800A6HDE3EP2292=CZ6_NHLCH082K0H2D0A6EKKDC520S3=LA0=/

SAUD S. ALOTAIBI received the bachelor's degree in computer science from King Abdulaziz University, Jeddah, Saudi Arabia, in 2000, the master's degree in computer science from King Fahd University, Dhahran, Saudi Arabia, in May 2008, and the Ph.D. degree in computer science from Colorado State University, Fort Collins, Co, USA, in August 2015. From January 2009 to 2017, he worked as the Vice Deputy of the IT Center, Umm Al-Qura University, Mecca, Saudi Arabia. Then, he was designated as the Vice COE of the Smart Campus. He is currently the Dean of the College of Computer and Information Systems. He is also an Associate Professor of Computer Science with Umm Al-Qura University. His current research interests include emotional intelligence, data mining, natural language processing, machine learning, deep learning, computer networks, wireless sensor networks, and network security.