ROLE OF MACHINE LEARNING IN PREDICTING STOCK PRICES: A LITERATURE SURVEY

Alamir Labib Awad, Arab Academy for Science, Technology and Maritime Transport
Saleh Mesbah Elkafas, Arab Academy for Science, Technology and Maritime Transport
Mohammed Waleed Fakhr, Arab Academy for Science, Technology and Maritime Transport

ABSTRACT

The stock value plays a vital factor in maximizing the companies profit. It has a direct impact on the economy by affecting it positively or negatively. Many studies have been made in order to help investors make the right decisions while buying or selling the stock. Most of the recent researches relies on machine learning and deep learning algorithms as they proved the best performance the most reliable techniques in solving the prediction problems. Stock price prediction is a very challenging task due to its non-linearity nature and wide range of sudden changes. Stock price prediction is considered a classical problem in which many efforts are made to solve it across the years. This literature review paper is conducted to explore the latest machine learning and deep learning techniques used in predicting the stock price values. In order to achieve the maximum benefit from this literature review, a comparative study is conducted in order to summarize the strength and weaknesses behind each technique briefly. The comparison is made to compare between different techniques in terms of performance and accuracy.

Keywords: Stock Market, Prediction, Machine Learning, Deep Learning

INTRODUCTION

Recently, the financial market is regarded as a fascinating invention, and it is said to have a significant impact on diverse fields like business, edification, jobs, technology, and thus on the economy (Kang, 2019; Kilimci, 2020; Yuan, 2020; Chen, 2018). Over the years, the interest of both the researchers and the investors is highly grabbed by the Stock return or stock market prediction. In general, the stock market is an aggregation of various purchasers and sellers of stock (or shares). It manifests the ownership claims on the business of a cluster of people or a specific individual (Thakkar, 2021; Dattatray, 2019; Idrees, 2019; Bousoño-Calzón, 2019). The forecast of the future value of the stock market is said to be the stock market prediction. The principal intention behind the prediction of the stock market is to reduce the ambiguity that comes up at some stage in investment decision-making.

A random walk is followed in the stock market, implying that tomorrow's stock value can be predicted accurately with today's value. An assumption is engaged in this stock returns prediction that the publicly available historical information has a specific prognostic association on the future stock returns (Bousoño-Calzón, 2018; Wen, 2019; Zhang, 2018; Teng, 2019). This historical information includes economic variables like economic variables, industry-specific information such as growth rates of industrial production and consumer price, and company-specific information such as income statements and dividend yields. On the other hand, the stock markets are highly affected by a variety of interrelated factors, inclusive of the "economic, political, psychological, and company-specific variables" (Alsubaie, 2019; Zavadzki, 2020; Garcia-Vega 2020; Hoseinzade, 2019). However, it is challenging to analyze the price behaviors and the

movements of the stock market owing to its non-linear, nonparametric, non-stationary market dynamics (Liu, 2020; Aithal, 2019; Zhou,, 2018).

Financial marketing is a complex and hybrid process where people may sell and purchase options, futures, commodities, and currencies through digital networks funded by traders. Through dealing in brokerage stocks or over-the-counter exchanges, the stock exchange helps purchasers to buy public sector shares. By spending small initial cash sums, this sector has given customers the ability to make money and enjoy a comfortable existence, less pressure than the risk of needing a well-paying job or starting a new business (Investopedia, July 2008). Capital markets are affected by several variables that create uncertainty and elevated price volatility. Execution can do good than any human being, which is more relevant in the momentum of order submission, regulated as Automated Trade Systems (ATS) by a computer machine. In contrast, man can get orders and submit them to the market.

Using machine learning algorithms in predicting stock prices has attracted a lot of interest during the past years. The reason for that is its success in predicting future values from analyzing the previous values. Recent progression in stock market prediction comes under four categories—"statistical, pattern recognition, Machine Learning (ML), and sentiment analysis" (Kim, 2020; Chen, 2020). Machine learning is being comprehensively deliberated for its potentials in the forecast of financial markets. Artificial Neural Networks (ANNs) are the commonly utilized ML technique for predicting stock trends (Liu, 2019; Zhou, 2020). In contrast to this, the ANN is highly prone to the over-fitting problem, and here the prediction is made only with little prior user knowledge. Results proved that analyzing the historical prices using single machine learning techniques only has given poor results. Several approaches are introduced to combine several machine learning algorithms to achieve the best results. The reason for that is the historical data is not enough for predicting the stock's future prices. Therefore, deep reinforcement learning (Deng, 2017; Zheng, 2019; Fengqian, 2020) in the stock trading decisions and stock price prediction can be used as a promising solution in the time series prediction. Optimization algorithms (Wu, 2020; Mishev, 2020; Yoon, 2020; Deng, 2017; Zheng, 2019) are found to be promising to train deep neural networks.

Sentiment analysis (Rajendra, 2020; Ren, 2019; Bouktif, 2020; Kilimci, 2020; Wu, 2020; Mishev, 2020; Yoon, 2020) is another approach in which the stock trends are automatically predicted by analyzing the text corpuses like the news feeds or tweets that are more unambiguous to stock markets and public companies. The hybrid approach applying the combination of sentiment-based analysis and machine learning is an emerging forecasting technique, which is available today for short-term prediction.

With rapid economic growth in recent years, financial practices have become increasingly dynamic, and their propensity towards variation has also become increasingly complicated. The research aims to explain the movement of economic events in financial and academic circles and predict their changes and growth. Using one of many methods, the potential forecast of financial data can clarify the development and changes of the financial sector at a macroscopic level and offer a basis for profit-making firms and investors to establish microscopic economic expectations and revenue optimization strategies. Because financial information contains nuanced, blurred, and imperfect information, calculating their growth rate is difficult to predict (Khan, 2020).

In order to predict stock market moves, many polls have been carried out largely on a daily timetable. Models that incorporate numerous data sources, such as news articles, Twitter data, and data from Google and Wikipedia, have been created. When integrated with equity analytical indicators and asset values, these external factors have driven share price fluctuations.

Traditionally, researchers have used multiple econometric and mathematical methods to create predictive analysis models before utilizing functional machine learning algorithms. In order to predict and evaluate financial goods, traditional econometric and statistical models of always-necessary linear models cannot be used prior to transforming non-linear models into linear models. The following advantages are Neural Networks (NNs) that, compared to conventional computational methods, give an essential feature of algorithms used for machine learning: data-

driven, adaptive, and numerical. Neural Networks have a more extraordinary ability to view inaccurate and noisy data and have been often used for time-series estimation. With large quantities of non-linear data by default, the recently built Deep Learning (DL) algorithms can be trained and used to create Deep Nns (DNNs) with several hidden layers and capture more abstract relationships of non-linear data. These DL algorithms can be used more adequately than conventional machine learning algorithms to solve non-linear problems (Kim, 2020; Chen, 2020; Liu, 2019; Zhou, 2020).

News reported on the operating performance of a business on retailer specifications that would drive cost variations. E.g., when news of advancement is released, buyers tend to purchase, allowing stock market interest to rise. On the other side, buyers tend to sell as negative news is reported and, thus, push market markets to decrypt; there is no doubt that data determines the actions of traders. To forecast the flow of stocks, few studies use the news component. In order to anticipate future market price shifts, distinctive machine learning calculations may be applied to stock ad details, utilizing business and news information to implement different AI techniques in this analysis (Rajendra, 2020; Ren, 2019; Bouktif, 2020; Kilimci, 2020; Wu, 2020; Mishev, 2020; Yoon, 2020).

The main point of DL is to get highly vigorous highlight sets to get vital information from challenging evidence from the actual world. To interpret and explain effects, resounding non-linear topologies of DNNs can be designed and used for complex structures. DNNs can speak better regarding complex high-dimensional capacities, such as radically changed capacities, unlike typical shallow NNs. Computer chips' data processing capacity has also improved dramatically. In specific, graphic processing units are highly appropriate for in-depth model training. Furthermore, in related zones, such as machine learning and data and flag handling, a growing number of inquiries have occurred. Too, required methods have been moved ahead to packing away massive libraries. This makes it possible for DL equations to balance superior, fundamentally non-linear, and rapid communication between popularization and development. Within the classification, debate identification, and computer vision zones, DL has now been stamped (Mishev, 2020; Yoon, 2020; Deng, 2017; Li, 2019).

The stock market prediction does not depend on the historical data only. An assumption is engaged in this stock returns prediction as the publicly available historical information has a specific prognostic association on the future stock returns (Bousoño-Calzón, 2018; Wen, 2019; Zhang, 2018; Shi, 2018). This historical information includes economic variables like economic variables, industry-specific information such as growth rates of industrial production and consumer price, and company-specific information such as income statements and dividend yields. On the other hand, the stock markets are highly affected by a variety of interrelated factors like: "economic, political, psychological, and company-specific variables" (Alsubaie, 2019; Zavadzki, 2020; Garcia-Vega, 2020; Hoseinzade, 2019). However, it is challenging to analyze the price behaviors and the movements of the stock market due to its non-linear, nonparametric, non-stationary market dynamics (Liu, 2020; Aithal, 2019; Zhou, 2018).

Using machine learning algorithms in predicting stock prices has attracted a lot of interest during the past years. The reason for that is its success in predicting future values from analyzing the previous values. Recent progression in stock market prediction comes under four categories— "statistical, pattern recognition, Machine Learning (ML), and sentiment analysis" (Kim, 2020; Long, 2020). Machine learning is being comprehensively deliberated for its potentials in the forecast of financial markets. Artificial Neural Networks (ANNs) are the commonly utilized ML technique for predicting stock trends (Wang, 2019; Zhou, 2020). In contrast to this, the ANN is highly prone to the over-fitting problem, and here, the prediction is made only with little prior user knowledge. Further, several approaches have been proposed to this task, with the most effective ones using the fusion of a pile of classifiers' decisions to predict future stock values. Results proved that analyzing the historical prices using single machine learning techniques only has given poor results. Several approaches are introduced to combine several machine learning algorithms in order to achieve the best results. Therefore, deep reinforcement learning (Deng, 2017; Li, 2019; Fengqian, 2020) in the stock trading decisions and stock price prediction can be used as a promising solution

in the time series prediction. Optimization algorithms (Wu, 2020; Mishev, 2020; Yoon, 2020; Deng, 2017; Li, 2019) are found to be promising to train deep neural networks.

This paper will analyze different techniques used in stock market value prediction. In addition to that, the paper will provide a comparison between these different techniques regarding efficiency and prediction accuracy. It is written to answer several research questions. The following are some of the research questions that will be answered in this paper,

RQ1: What are the challenges behind predicting stock prices, and how to overcome these challenges?
RQ2: How can machine learning and deep learning help decision-makers in predicting stock price?

RESOURCES AND METHODS

In order to deeply understand the effectiveness of using different machine learning and deep learning on stock price prediction, a lot of research papers are explored. The strategy used in this paper to explore the literature is as follows, selecting the topic, identifying the best search term, collecting the required materials, identification of the exclusion and inclusion criteria, analyzing the results, and making discussions.

The topic to be reviewed in this paper is "Stock price prediction using machine learning and deep learning techniques." There are two search terms used in order to collect the most suitable papers for this literature review. The first search term is "Stock price prediction using machine learning," while the second search term is "Stock Price prediction using deep learning."

The mentioned search terms are used to search a wide range of research papers, including journal and conference papers. The search is narrowed down to include the most recent journals between 2019 and 2021. The search engines provided a lot of resources, which are filtered out based on inclusion and exclusion criteria. These exclusion and inclusion criteria are used to analyze the resources and pick up the most relevant papers to include in the literature review.

The inclusion criteria used in this work are as follows:

- Papers that introduce an implementation of machine learning or deep learning on the stock price prediction.
- Papers which published during the years 2018, 2019, 2020, and 2021.
- Papers from well-known journals or conferences.

On the other hand, the used exclusion criteria are:

- Papers that introduced repeated work.
- Papers which introduced using old techniques
- Papers that deal with a domain other than stock price prediction

The search is conducted using Google Scholar. Due to the mass volume of results, the papers are filtered out based on the previously mentioned inclusion and exclusion criteria. Finally, a total of 49 Paper is reviewed. The following figure describes the process used in order to collect the required papers for this review.

Selection method: The search terms provided hundreds of papers. The search was done using Google Scholar that provided multiple resources. Due to the high volume of results, the search was done in online academic research databases, namely IEEE, Elsevier, Springer, etc. A total of 49 papers were reviewed, analyzed and discussion provided. The following table shows the distribution of the papers based on their publisher.

Table 1 THE NUMBER OF PAPERS REVIEWED BASED ON EACH RESOURCE	
Resources from online academic DB (based on search terms)	Number of Papers reviewed
Elsevier/ Science Direct (Journals/ Conference/ Research	9

papers)	
Springer (Journals/ Conference/ Research papers)	8
IEEE (Journals/ Conference/ Research papers)	28
Online Resources (Articles, Reports, Whitepapers, etc.)	5
Total papers reviewed fully	**50**

The resources are deeply analyzed to understand each work's contributions in predicting the stock price values. Finally, a comparative study is conducted in order to highlight the advantages and disadvantages of each research paper.

LITERATURE REVIEW

The significant development in deep learning techniques introduced new strategies in predicting stock prices. In this section, a lot of deep learning techniques used in stock price prediction are discussed.

Carta, et al., (2020) had proposed an ensemble of reinforcement learning approaches that do not use the annotations to reduce the over-fitting problem and maximize the return function over the time of training. In this research work, the authors have utilized a trained Q-learning stage and have tested its ensemble behavior in real-world stock markets. The main idea behind this research is to use different agents, one new agent per epoch, and finally ensemble the results from the agents. Further, with the intraday training, the proposed work had exhibited better performance in terms of qualitative and quantitative measures.

Jiang, et al., (2020)have developed two-stage ensemble models for stock price prediction by blending the concepts of Empirical Mode Decomposition (EMD) (or Variational Mode Decomposition (VMD), Extreme Learning Machine (ELM), and Improved Harmony Search (IHS) algorithm. The proposed work was referred to as EMD–ELM–HIS and VMD–ELM– HIS. The results of the proposed work had exhibited superior performance over the existing models like the EMD based ELM (EMD–ELM), VMD based ELM (VMD–ELM), autoregressive Integrated Moving Average (ARIMA), ELM, Multi-Layer Perception (MLP), Support Vector Regression (SVR), and Long Short-Term Memory (LSTM) models in terms of accuracy and stability. The contribution of this paper is to divide the prediction process into two stages. The first stage uses the ELM model, which uses the EMD or VMD to generate several predictions based on several parameters. The second stage is using the improved harmony search (HIS) metaheuristics search algorithm to integrate the best prediction among the results from the first stage.

Chou & Nguyen (2018) constructed an intelligent time series prediction system with the sliding-window metaheuristics optimization to predict Taiwan's stock prices. In this research work, the functions and the graphical user interface were stand-alone applications. As a consequence, the developed hybrid system had exhibited higher prediction performance.

Dang, et al., (2018) had developed a novel framework for forecasting the stock price direction with the financial news and sentiment dictionary. A novel two-stream Gated Recurrent Unit Network and a sentiment Stock2Vec embedding model trained on financial news dataset and Harvard IV-4. In this research work, the authors have conducted two main experiments: (a) direction of predicts S&P 500 index stock price prediction using the historical S&P 500 prices and the articles crawled from Reuters and Bloomberg, (b) VN-index price trends were forecasted using VietStock news and stock prices from cophieu68.

Ananth & Vjayakumar (2020) had utilized the k-NN regression for predicting the market trend. The authors have utilized the regression model for predicting the stock price and candlestick pattern detection. On the candlestick graph, the projected work had generated the signals that predict market movement accurately for making the 'Buy/Sell' decision. Sequentially, there was an increase in the prediction accuracy of the stock exchange.

Moghar, et al., (2020) had proffered a novel approach for predicting the stock market price of both GOOGL and NKE assets using the Recurrent Neural Networks (RNN) and Long-Short

Term Memory model (LSTM). More promising results were acquired with the proposed work in terms of maximal prediction accuracy.

Polamuri, et al., (2020) had developed the Stock Market Prices Prediction Framework (SMPPF) with the underlying algorithms Multi-Model based Hybrid Prediction Algorithm (MM-HPA), which was the amalgamation of the linear (autoregressive moving average model) and non-linear model (recurring neural network) including Genetic Algorithm (GA). In the hybrid model, the optimal parameters were fine-tuned *via* a GA.

Khan, et al., (2019) tried to find the effect of public sentiment and political situations on both individual companies and the whole market. A machine-learning algorithm was used to find out how the sentiment affects the accuracy of prediction for the upcoming seven days. In addition to that, the study tried to find out the dependency between the companies and the stock market. In order to prove the efficiency, an experimental part is conducted, and the historical data is downloaded from Yahoo finance, while the public sentiments are collected from Twitter. Finally, the important political events data are obtained from Wikipedia. The text data are pre-processed, and sentiment analysis is conducted to extract the features and produce the dataset. Several machine learning algorithms are applied to predict the trend of the data. The final results have proven that utilizing the sentiment and political data has improved the accuracy significantly.

Li, et al., (2020) have introduced the bidirectional encoder representations from Transformers (BERT) model. Firstly, the algorithm extracts the sentiment values from the online information using the BERT model. Then the sentiment values are given weights. Finally, a two-step regression validation model is used to analyze the relationship between the sentiment values and stock values. The experimental results showed that sentiment values have a great impact on the stock yield. Additionally, experiments proved that the BERT model used in this paper is outperformed the other algorithms like the SVM and the LSTM as BERT achieved an accuracy of 97.35%.

Jin, et al., (2019) proposed a technique that is based on investors' emotions. They tried to analyze the investor's emotions and apply sentiment analysis to it to help in improving the prediction accuracy of the future market trends. Then they adopted the empirical modal decomposition (EMD), which gives a good performance during prediction. Then adopt LSTM, which is well known in prediction problems due to its memory capabilities. In addition to its good performance, LSTM proved that it could reduce the time delay. The final result from this research says that adding the investor's emotions to the ML models will help in improving the model performance while predicting the stock prices.

Li, et al., (2020) have made a detailed study to determine the effect of investor sentiment on stock prices. The sentiments are measured by the expectations extraction from the user-generated content. In this paper, the authors considered many factors like selecting several text classification algorithms, selecting various price forecasting models, and different information update schemas. They made the comparison using different techniques like Long-Term Memory (LSTM model), logistic regression, support vector machine, and naive Bayes model. After analyzing the results, they concluded the investor sentiment affects only the open prices. On the other hand, the closing prices are affected by the hourly sentiments. In addition to that, they proved that advanced models like LSTM have better prediction capabilities with respect to the other traditional models.

Wu, et al., (2020) have utilized deep reinforcement learning to introduce a new trading strategy. Due to the time-series nature of the stock market data; In order to extract the financial features, the Gated Recurrent Unit (GRU) is used for feature extraction. Using reinforcement learning, they proposed two trading strategies; the first one is GDQN(Gated Deep Q Learning Trading Strategy), while the second one is GDPG (Gated Deterministic Policy Gradient trading strategy). Their experimental results have proven that their proposed GDQN and GDPG are outperforming the other classical trading strategies. In comparison between the GDQN and DGPG, the results demonstrated that the GDPG is more stable than the GDQN.

Lussange, et al., (2020) had introduced a MAS stock market simulator. In this simulator, there are several agents which utilize reinforcement learning to learn trading strategies

automatically. They adjusted their model to fit the London stock exchange over a period expanded from 2007 to 2018. Their results showed that the proposed model could be used to regenerate the market matrices like price autocorrelation scalars through several time intervals. The conclusion from this work is that agent learning can simulate the market structure and market features.

Li, et al., (2019) have utilized deep reinforcement learning to learn the stock trading decisions and the stock price prediction. The effectiveness of the proposed model has been compared with the traditional models through experimental experiments to prove its advantages. The results assured the efficiency behind using deep reinforcement learning in financial markets and decision-making.

To predict the stock price; Shin et al., (2019)

Have combined the convolutional neural network and the long short-term memory. This work aims to generate several charts from the stock trading data and input them into the CNN layer. The CNN makes the feature extraction, and these features are fed to the LSTM. They used a multimodal structure in which the input layer is separated from the hidden layers, and finally, the results from each layer are combined. Regarding reinforcement learning, they defined the agents' policy and determined the buying and selling probabilities. Experimental simulation is conducted on data from Korea KOSPI to prove the efficiency of the proposed algorithms. The training data is from March 2009 to March 2019, while the test data started in April 2019. The results of this work showed that it has excellent performance regarding both bear and bull decisions.

Li, Ni & Chang (2019) proposed the method of DRL in stock trading decisions and stock-price prediction. The method was presented to effectively gain trading signals in the transaction process and maximize the benefits of stock trading, which has been a challenge for a long time. However, the authors used experimental data, and the proposed model is compared with a traditional model to compare and highlight its advantages. The authors claim that DRL provides the potential and is a feasible option in financial markets and strategic decision-making from the point of view of forecasts in the stock market.

Zhang (2019) the authors propose a new architecture named Generative Adversarial Network (GAN). This new architecture contains two major parts; they are Multi-layer perceptron and a long short-term memory. The role of the MLP is to act as a discriminator to extract the distribution of the data from the stock data. On the other hand, the LSTM acts as a generator to predict the closing price of the stocks. This new architecture is examined against the S&P 500 index. This introduced architecture gives a better performance for predicting the closing price of the stock than the other ML and DL techniques.

COMPARATIVE STUDY

Some recent papers will be reviewed from the aspects of techniques used, performance measures, advantages, and disadvantages. Finally, the research gaps presented in this field will be discussed.

<table>
<tr><td colspan="4">Table 1
TECHNIQUES USED, PERFORMANCE MEASURES, ADVANTAGES, AND DISADVANTAGES</td></tr>
<tr><td>Ref No</td><td>Techniques used</td><td>Advantages</td><td>Disadvantages</td></tr>
<tr><td>1</td><td>DRL with multi-agents Finally, Ensemble the agent's predictions.</td><td>• Provide global results in terms of return, convergence, and equity as well
• Higher precision
• Statistically significant</td><td>• Lower convergence
• Lower reliability</td></tr>
<tr><td>2</td><td>Blending EMD with VMD to make feature reduction</td><td>• Lower Mean Absolute Percentage Error (MAPE) in terms of standard</td><td>• Higher Mean Absolute Error (MAE), Mean Square Error</td></tr>
</table>

		• deviation, mean • High prediction accuracy.	• (MSE) • Slow convergence
3	Sliding window metaheuristics optimization	• Applicable to highly non-linear time series • Highly accurate predictive model.	• Slow convergence • Low computational speed • Does not achieve outstanding results for long-term investment
4	Two stream Gated RNN Depend on historical data along with news and sentiment dictionary into analysis.	• Improves the accuracy of the stock trends prediction. • Reduces the false classification	• Lower precision and recall • Training and validation loss is higher • Suitable for daily stock movement rather than the long-term movement
5	Utilizing the K-NN regression	• Higher classifier accuracy • More accurate • Higher reliability	• Lower root mean square deviation
6	Utilization of RNN in addition to LSTM	• Higher precision of forecasting • Lower time consumption	• Minimized error rate • Testing time is higher
7	Using RNN along with Genetic algorithms to tune parameters	• Reduce prediction error • Produces effective predictions of stock returns. • Lower mean absolute error (MAE) and mean squared error (MSE).	• Higher computation time • Lower prediction accuracy
8	Study the effect of sentiment and political situations on the stock price prediction accuracy	• Higher prediction accuracy • Highly robust • The highest accuracy using sentiment analysis	Shows higher error metrics.
28	DRL in stock trading and stock price prediction	• Support in effective decision making, as traditional decision making tends to ignore certain data • Analysis of algorithms along with DRL show effective results for timely decisions in stock trading DRL is feasible in stock value prediction	• Not fully applicable to all stocks must be tested in a real-time environment • Predictive modeling of stock value must be achieved with real-time data
43	BERT to extract sentiment values Regression model to link sentiment values with stock values	• Higher accuracy • 95% confidence level	• Lower recall and F Score
44	Study the effect of investor emotions Utilized EMD for feature reduction	• Efficiently extract specific information • More accurately and timely stock price can be predicted • Highest accuracy	• Ignoring most of the important information • Tedious process

	LSTM for prediction	• Lowest time offset • Closest predictive value when predicting the stock market	
45	Study investor sentiment Text Classification with different techniques LSTM is the best classifier	• Provide more predictive power with investor sentiment only • Accurately address the predictive power of text-extracted investor sentiment	• Very Slow
46	DRL to learn trading strategy GRU to extract financial features extraction	• DRL supports defining a spectrum of continuous activities • Good performance in the volatile stock market • GDQN and GDPG are proposed for quantitative stock trading	• It does not explain the use of DRL in predicting the stock value
47	Introduced MAS simulator to simulate the stock movements	• Enables accurate emulation of the market microstructure • Good generalization on the testing set. • Accuracy values on the training set are generally higher than for the testing set	• Accuracy on the training set is not too large
48	DRL to learn: Stock trading decision Stock price prediction	• Proves the feasibility of deep reinforcement learning in financial markets • Proves the credibility and advantages of strategic decision-making.	• Bring instability with large data size
49	Combined CNN with LSTM CNN for Feature extraction LSTM for prediction	• High price point • Minimum loss rate	• Low prediction performance.

DISCUSSION

From these reviewed papers, the conclusion is that many techniques and models are proposed to predict the stock market to support decision-makers in their decisions. Stock trading and stock prediction is a challenging task that attracts a lot of research interest. Many researchers have proposed different techniques using machine learning and deep learning along with experimental trials to prove each proposed technique's efficiency.

From the explored studies, it is noted that combining the historical stock data along with the sentiment data will help the Deep learning techniques in predicting the stock values accurately. Predicting stock values is a very challenging problem due to the non-linear behavior of the stock values and their unpredictable nature. Despite the deep uncertainty in the stock, the most recent techniques have proven their ability to predict it using movement trends. These new technologies are not limited to using historical data but also uses other information to help in prediction. There is much other information that can help in stock value prediction like political activities, economic situations, investors' opinions, and so on.

Generally, it can be said that Machine learning and deep learning have proved their ability to process large data set within a low timeframe. It is well-known that machine learning can analyze datasets and extract valuable patterns and information from them. Recently the Deep Learning

techniques have proven to outperform machine learning techniques, especially in time series problems. The experimental results showed that using the historical data and the current data can significantly improve prediction accuracy.

Studies proved that RNN, which keeps the recent events in its memory and links them with the current events, has improved the prediction accuracy. Additionally, LSTM -which is an improvement for the RNN- can process the whole data sequence. This reason leads to the wide use of LSTM in financial time series prediction.

Designing accurate stock price prediction algorithms has attracted much research interest. In this paper, many techniques are review for the stock values prediction. Many of the reviewed techniques achieved a good performance regarding accuracy and applicability. The latest researches recommend using sentiment analysis along with the historical data for better performance. Studies showed that the techniques which are based on deep learning outperformed all other techniques. The reason for that is its ability to make a prediction even with non-linear data. In addition to that, deep learning can learn from different information sources and link them together to get the best performance. Deep Reinforcement Learning is a new category of Deep learning in which the agent learns from the environment through trial and error, and it does not need labeled data. DRL has proved its extraordinary performance in solving prediction problems.

CONCLUSION

This review tried to explore the latest techniques used in stock value prediction. Most of the recent techniques applied Deep Learning algorithms because of their efficiency in solving time series problems. This literature review focused on identifying the stock price movement patterns and making predictions based on these patterns. Most of the papers recommend considering the company situations to achieve accurate prediction. They recommend using the LSTM techniques because it gives better performance than the other techniques due to its memory capabilities. The latest techniques in stock price prediction are based on utilizing Deep Reinforcement Learning. DRL has proven its efficiency in solving time series problems like the stock price prediction.

REFERENCES

Carta, S., Ferreira, A., & Sanna, A. (2020). "Multi-DQN: An ensemble of Deep Q-learning agents for stock market forecasting". *Expert Systems with Applications*.

Jiang, M., Jia, L., Chen, Z., & Chen, W. (2020). "The two-stage machine learning ensemble models for stock price prediction by combining mode decomposition, extreme learning machine and improved harmony search algorithm". *Annals of Operations Research*.

Jui-Sheng, C., & Nguyen, T.-K. (2018). "Forward forecast of stock price using sliding-window metaheuristic-optimized machine learning regression". *IEEE*.

DANG, L.M., Abolghasem, S.-N., Huynh, H.D., Kyungbok, M., & Moon, H., (2018). "Deep learning approach for short term stock trends prediction based on two-stream gated recurrent unit network". *IEEE*.

Ananth, M., & Vjayakumar, K. (2020). "Stock market analysis using candlestick regression and market trend prediction (CKRM)". *Journal of Ambient Intelligence and Humanized Comput*ing.

Moghar, A., & Hamiche, M. (2020). "Stock market prediction using LSTM recurrent neural network". *Procedia Computer Science.*

Subba-Rao, P., Srinivas, K., Mohan, A.K. (2020). "Multi model-based hybrid prediction algorithm (MM-HPA) for Stock Market Prices Prediction Framework (SMPPF)". *Arabian Journal for Science and Engineering.*

Wasiat, K., Usman, M., Ali, G.M., Azam, M.A., Alyoubi, K.H., & Alfakeeh, S.A. (2020). "Predicting stock market trends using machine learning algorithms *via* public sentiment and political situation analysis". *Soft Computing*, 24, 11019-11043.

Lee, J., Kim, R., Koh, Y., & Kang, J. (2019). "Global stock market prediction based on stock chart images using deep q-network". In *IEEE Access*, 7, 167260-167277,

Kilimci, Z.H., & Duvar, R. (2020). "An efficient word embedding and deep learning based model to forecast the direction of stock exchange market using twitter and financial news sites: A case of istanbul stock exchange (BIST 100)." In *IEEE Access*, 8, 188186-188198.

Yuan, X. Yuan, J., Jiang, T., & Ain, Q.U. (2020). "Integrated long-term stock selection models based on feature selection and machine learning algorithms for china stock market". In *IEEE Access*, 8, 22672-22685,

Chen, L., Qiao, Z., Wang, M., Wang, C., Du, R., & Stanley, H.E. (2018). "Which artificial intelligence algorithm better predicts the chinese stock market? " In *IEEE Access*, 6, 48625-48633.

Thakkar, A., & Chaudhari, K. (2021). "Fusion in stock market prediction: A decade survey on the necessity, recent developments, and potential future directions". *Information Fusion*, 65, 985-107.

Dattatray, P., Gandhmal, K., & Kumar. (2019). "Systematic analysis and review of stock market prediction techniques". *Computer Science Review*, 34.

Idrees, S.M., Alam, M.A., & Agarwal, P. (2019). "a prediction approach for stock market volatility based on time series data." In *IEEE Access*, 7, 17287-17298.

Bousoño-Calzón, C., Bustarviejo-Muñoz, J., Aceituno-Aceituno, P., & Escudero-Garzás, J.J. (2019). "On the economic significance of stock market prediction and the no free lunch theorem." In *IEEE Access*, 7, 75177-75188.

Bousoño-Calzón, C., Molina-Bulla, H., Escudero-Garzás, J.J., & Herrera-Gálvez, F.J. (2018). "Expert selection in prediction markets with homological invariants." In *IEEE Access*, 6, 32226-32239.

Wen, M., Li, P., Zhang, L., & Chen, Y. (2019). "Stock market trend prediction using high-order information of time series." In *IEEE Access*, 7, 28299-28308.

Zhang, X., Qu, S., Huang, J., Fang, B., & Yu, P. (2018). "Stock market prediction *via* multi-source multiple instance learning." In *IEEE Access*, 6, 50720-50728.

Shi, L., Teng, Z., Wang, L., Zhang, Y., & Binder, A. (2019). "DeepClue: Visual interpretation of text-based deep stock prediction". In *IEEE Transactions on Knowledge and Data Engineering, 31* (6), 1094-1108.

Alsubaie, Y., Hindi, K.E., & Alsalman, H. (2019). "Cost-sensitive prediction of stock price direction: Selection of technical indicators." In *IEEE Access*, 7, 146876-146892.

Zavadzki, S., Kleina, M., Drozda, F., & Marques, M. (2020). "Computational intelligence techniques used for stock market prediction: *A Systematic Review." In IEEE Latin America Transactions*, 18(04), 744-755.

Garcia-Vega, S., Xiao-Jun, Z., & Keane, J. (2020). "Stock returns prediction using kernel adaptive filtering within a stock market interdependence approach". *Expert Systems with Applications.*

Ehsan, H., & Saman, H. (2019). "CNNpred: CNN-based stock market prediction using a diverse set of variables". *Expert Systems With Applications.*

Liu, J., Lin, H., Yang, L., Xu, B., & Wen, D., (2020). "Multi-element hierarchical attention capsule network for stock prediction." In *IEEE Access*, 8, 143114-143123.

Aithal, P.K., Dinesh, A.U., & Geetha, M. (2019). "Identifying significant macroeconomic indicators for indian stock markets." In *IEEE Access*, 7, 143829-143840.

Zhou, P., Chan, K.C.C., & Ou, C.X. (2018). "Corporate communication network and stock price movements: Insights from data mining." In *IEEE Transactions on Computational Social Systems*, 5(2), 391-402.

Kim, S., Ku, S., Chang, W., & Song, J.W. (2020). "Predicting the direction of US stock prices using effective transfer entropy and machine learning techniques." In *IEEE Access*, *8*, 111660-111682.

Long, J., Chen, Z., He, W., Wu, T., & Ren, J. (2020). "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 Journal.*

Liu, G., & Wang, X. (2019). "A Numerical-based attention method for stock market prediction with dual information". in *IEEE Access*, 7, 7357-7367.

Zhou, M., Yi, J., Yang, J., & Sima, Y. (2020). "Characteristic Representation of Stock Time Series Based on Trend Feature Points," in *IEEE Access*, 8, 97016-97031.

Rajendra N., Paramanik, & Singhal, V. (2020). "Sentiment analysis of indian stock market volatility". *Procedia Computer Science.*

Ren, R., Wu, D.D., & Liu, T. (2019). "Forecasting stock market movement direction using sentiment analysis and support vector machine". in *IEEE Systems Journal*, 13(1), 760-770.

Bouktif, S., Fiaz, A., & Awad, M. (2020). "Augmented textual features-based stock market prediction." In *IEEE Access*, 8, 40269-40282.

Kilimci, Z.H. & Duvar, R. (2020). "An efficient word embedding and deep learning based model to forecast the direction of stock exchange market using twitter and financial news sites: A case of istanbul stock exchange (BIST 100)." In *IEEE Access,* 8, 188186-188198.

Wu, B. (2020). "Investor behavior and risk contagion in an information-based artificial stock market." In *IEEE Access*, 8, 126725-126732.

Mishev, K., Gjorgjevikj, A., Vodenska, I., Chitkushev, L.T., & Trajanov, D. (2020). "Evaluation of sentiment analysis in finance: From lexicons to transformers." in *IEEE Access*, 8, 131662-131682.

Yoon, B., Jeong, Y., & Kim, S. (2020). "Detecting a risk signal in stock investment through opinion mining and graph-based semi-supervised learning". In *IEEE Access*, 8, 161943-161957.

Deng, Y., Bao, F., Kong, Y., Ren, Z., & Dai, Q. (2017)."Deep direct reinforcement learning for financial signal representation and trading." In *IEEE Transactions on Neural Networks and Learning Systems*, 28(3), 653-664.

Li, Y., Zheng, W., & Zheng, Z. (2019). "Deep robust reinforcement learning for practical algorithmic trading." In *IEEE Access*, 7, 108014-108022.

Fengqian, D., & Chao, L. (2020). "An adaptive financial trading system using deep reinforcement learning with candlestick decomposing features,". in *IEEE Access*, 8, 63666-63678.

Khan, W., Malik, U., Ghazanfar, M.A., Azam, M.A., Khaled H.A., & Alfakeeh, A. (2020). Predicting stock market

trends using machine learning algorithms *via* public sentiment and political situation analysis. *Soft Comput*, 24, 11019–11043.

Li, M., Li, W., Wang, F., Jia, X., & Rui, G. (2020). "Applying BERT to analyze investor sentiment in stock market". *Neural Computing and Applications*.

Jin, Z., Yang, Y., & Liu, Y. (2019. "Stock closing price prediction based on sentiment analysis and LSTM". *Neural Computing and Applications.*

Li, Y., Bu, H., Li, J., & Wu, J.(2020). "The role of text-extracted investor sentiment in Chinese stock price prediction with the enhancement of deep learning". *International Journal of Forecasting, 36*.

Wu, X., Chen, H., Wang, J., Troiano, L., Loi, V., & Fujita, H. (2020). "Adaptive stock trading strategies with deep reinforcement learning methods". *Information Sciences*.

Lussange, J., Lazarevich, I., Bourgeois-Gironde, S., Palminteri, S., & Gutkin, B. (2020). "Modelling stock markets by multi-agent reinforcement learning". *Computational Economics*.

Li, Y., Ni, P., Chang, V. (2019). "Application of deep reinforcement learning in stock trading strategies and stock forecasting". *Computing*.

Hong-Gi, S., Ra, I., Yong-Hoon, C. (2019)."A deep multimodal reinforcement learning system combined with cnn and lstm for stock trading", *IEEE*.

Zhang, K., Zhong, G., Dong, J., Wang, S., & Wang, Y. (2019). "Stock Market Prediction Based on Generative Adversarial Network," Procedia Computer Science, 147, 400–406.