International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Stock Market Price Prediction Using LSTM: A Machine Learning Approach for Time-Series Forecasting

Penjarla Sandeep[1], Vemula Pranay [2], Penjarla Naveen Kumar[3], Nakka Indhravathi[4], Mamidyala Hari Krishna[5], Makkala Ram[6]

[1,3,4,5,6]P.G. Research Scholar, Dept. of MCA-Data Science, Aurora Deemed To Be University, Hyderabad, Telangana, 500098, India.

[2]Assistant Professor, Dept. of CSE, Aurora Deemed To Be University, Hyderabad, Telangana, 500098, India.

Email:[1]penjarlasandeep2@gmail.com,[2] pranay.vemula@aurora.edu.in, [3]naveenyadav70322@gmail.com, [4]nakkaindhu123@gmail.com, [5]mamidyalah@gmail.com, [6]rammakkala521@gmail.com

A B S T R A C T

Stock market prediction is a highly challenging task due to the dynamic, volatile, and nonlinear nature of financial markets. Traditional approaches such as statistical and regression-based models often fail to capture the sequential dependencies and nonlinear patterns in stock price data. In this work, we propose a machine learning–based framework for stock price prediction using the Long Short-Term Memory (LSTM) model, a variant of recurrent neural networks (RNN) designed to handle time-series data. The study uses historical stock data of BMW, including Open, High, Low, and Close (OHLC) prices, to train and evaluate the model. Data preprocessing steps such as handling missing values, feature scaling using MinMaxScaler, and sliding window generation were applied to ensure robust predictions. The LSTM model was trained with multiple layers and evaluated using performance metrics including Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and R^2 score. The experimental results demonstrate that the proposed LSTM approach provides reliable predictions of stock price movements and outperforms traditional statistical methods in capturing long-term dependencies. Visualization of results through line charts and bar graphs validates the model's ability to track stock price trends effectively. This work highlights the potential of deep learning models for financial forecasting, providing investors and traders with valuable insights for decision-making. Future enhancements may include integrating sentiment analysis, hybrid models, and real-time stock data for improved prediction accuracy.

Keywords: Analysis of Financial Data, Trends in Stock Market, Predictive Modelling, Machine Learning, Time-series forecasting, Long Short Term Memory (LSTM), Deep learning and Prediction of Stock Price.

1. Introduction

The stock market is an important instrument at a global level for economic growth. It makes it possible for companies to be able to raise capitals and for investors to earn returns. But predicting stock prices is among the most difficult tasks since they are highly volatile and often exhibit non-linear patterns while also reflecting the influence from external conditions such as political events, economic conditions, and global news.

Traditional statistical models such as ARIMA and regression methods have been deemed to hold only for linear patterns and will yield incorrect results with most complex time-series data. With this vast improvement in the field of machine learning and deep learning, models like Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) frameworks have exhibited great promise in financial forecasting. LSTMs are, in fact, constructed to deal with sequential data in that they are able to learn long-term dependencies and nonlinear behaviors in the prices of stocks.

This is a study that sets out to develop a stock market price prediction system based on LSTM in predicting the stock price of BMW. The dataset is historical OHLC (Open, High, Low, Close) values that will be processed before it is fed to the model. The predictions will be evaluated with preconfigured metrics like RMSE and MAE while visualizations will be provided to compare predicted and actual trends. The purpose is to show how successful LSTMs can be in forecasting stock prices and potentially applicable for traders and investors in making data-driven decisions.

2. Literature Review

Stock market prediction has been a widely researched domain due to its impact on investment strategies, risk management, and financial decision-making. The researchers have been employing a variety of methods ranging from statistical techniques to advanced machine learning and deep learning-based approaches. This section summarizes the major contributions of prior works, describes their strengths and disadvantages, and also establishes how our research fills those gaps.

2.1 Statistical Approaches

Conventional statistical models like ARIMA and GARCH have commonly been used for the purpose of financial forecasting.

- The authors Khan et al. applied ARIMA for predicting volatility in the Indian stock market. The model was good in predicting short-term behavior, while it was poor in capturing the non-linearities and big fluctuations, which limited its ability to generalize in dynamic markets.
- Ahmed et al. (2021) discussed the mixed ARIMA model for forecasting stock prices of the pharmaceutical sector and presented good short-term accuracy. However, the model faced risks of overfitting on complex datasets.

These methodologies are easy to understand, interpretable, and computationally efficient but are incapable of modeling long-term dependencies and inherent nonlinearity of financial data.

2.2 Machine Learning Approaches

During the period of the introduction of machine learning, other models introduced were Support Vector Machines (SVM), Random Forest (RF), and K-Nearest Neighbors (KNN) for stock-price forecasting.

- Patel and Patel (2019) performed data mining methods KNN, Genetic Algorithms, and Support Vector Regression to enhance sentiment data from social media; however, those models were very sensitive to noisy and unstructured data.
- Yadav and Dubey (2020) employed Random Forest sentiment analysis from Twitter data, enhancing prediction accuracy but failing to evaluate the model with new datasets.

Machine learning methods are comparatively more flexible than ARIMA but still face problems modeling sequential dependencies over the time.

2.3 Deep Learning Approaches

Deep learning methods, particularly recurrent neural networks (RNNs) and Long Short-Term Memory (LSTM) networks, had the upper hand for time series in stock applications due to their ability to capture sequential patterns.

- Zhang et al. (2019) implemented deep learning for short-term stock trend prediction in the market of China with better accuracy than the traditional methods; however, keeping model specifications vague.
- Kumar and Sharma (2021) made a review of LSTM models and hybrid approaches, such as LSTM-CNN, which favored forecasting performance but posed a higher amount of computational cost.
- Kumar and Mahadevan (2021) used RNN and LSTM for the analysis of the stock market and highlighted the integration of sentiment analysis alongside historical data for better outcomes.

LSTMs have strong capacity to fit time-series data, although overfitting and training time remain challenges.

2.4 Hybrid and Comparative Studies

A lot of studies have compared traditional models versus their deep learning counterparts.

- Hossain and Imran (2020) compared LSTM and ARIMA in predicting the Tehran Stock Exchange with the conclusion that LSTM outperformed ARIMA regarding long-term forecasting.
- Sharma et al. (2023) reviewed a number of methods such as ARIMA, Random Forest, SVM, and deep learning hybrids, concluding that feature engineering and hybrid models significantly enhance accuracy but increase complexity.

From these studies, it is suggested that while hybrid approaches give better predictive power, they are not as suitable for real-time deployment due to the complexity and computational costs.

2.5 Research Gap

From the literature, it is clear that:

- Statistical models handle linear dependencies but fail in nonlinear, volatile markets.
- Machine learning methods improve performance but are less effective in sequential pattern recognition.
- LSTM models perform well in capturing time-series dependencies but are computationally heavy and prone to overfitting.
- Few studies have combined **systematic preprocessing, visualization, and usability aspects (GUI)** to make stock prediction models more practical for real-world investors.

- Our work focuses on developing an **LSTM-based forecasting system with robust preprocessing, experimental validation, and visual analysis**, bridging the gap between accuracy and usability.

Table 1 - Comparative Analysis Table

S. No.	Title	Authors & Year	Methodology	Strengths	Limitations
1	A Prediction Approach for Stock Market Volatility Based on Time Series Data	Khan et al. (2019)	ARIMA on Indian stock data	Captures short-term trends	Ineffective for nonlinear markets
2	A Survey on Stock Market Prediction Using ML Techniques	Kumar & Manogaran (2020)	Reviewed ANN, RNN, HMM	Comprehensive review	No experimental validation
3	Short-term Stock Market Price Trend Prediction	Zhang et al. (2019)	Deep learning framework	Handles noisy data	Lack of clarity on architecture
4	Techniques for Stock Market Prediction: A Review	Sharma et al. (2023)	Hybrid methods (ARIMA, RF, CNN, LSTM)	Explored multiple approaches	High computational cost
5	Forecasting Stock Prices Using Mixed ARIMA	Ahmed et al. (2021)	Mixed ARIMA	Good for short-term accuracy	Risk of overfitting
6	Stock Price Prediction Using LSTM: An Advanced Review	Kumar & Sharma (2021)	LSTM, hybrid CNN-LSTM	Captures long-term trends	Expensive training process
7	Stock Market Prediction Using Data Mining	Patel & Patel (2019)	KNN, SVR, GA	Incorporates sentiment data	Sensitive to noise
8	Comparing LSTM & ARIMA for Stock Prediction	Hossain & Imran (2020)	LSTM vs ARIMA	LSTM performs better long-term	Both struggle with volatility
9	Stock Market Price Prediction and Analysis	Kumar & Mahadevan (2021)	RNN, LSTM, sentiment	Integration of textual data	High reliance on historical patterns
10	Prediction on Stocks Using Data Mining	Yadav & Dubey (2020)	Random Forest + Sentiment	Improved accuracy via Twitter data	Poor generalization on unseen data

3. Proposed System & Methodology

The proposed system aims to predict future stock prices using historical market data and a Long Short-Term Memory (LSTM) model. The methodology is designed to ensure reliable preprocessing, efficient model training, and clear visualization of results. The overall workflow is shown in the system architecture diagram (Fig. 1).

3.1 Dataset Description

The dataset used in this study consists of **BMW stock market data**, including daily values of Open, High, Low, Close (OHLC), and trading volume. The data was obtained in CSV format and covers multiple years of trading history, making it suitable for time-series analysis. The **Close price** was chosen as the target variable for prediction, since it represents the actual market valuation at the end of each trading day.

3.2 Data Preprocessing

To prepare the dataset for the LSTM model, the following preprocessing steps were applied:

- **Cleaning:** Missing and invalid values were removed.
- **Sorting:** Records were arranged in chronological order to preserve temporal dependency.
- **Scaling:** Data was normalized to the range [0,1] using **MinMaxScaler** to stabilize model training.

- **Sliding Window Generation:** Input sequences of 60 days were created to predict the next day's stock price, capturing short and long-term dependencies.

3.3 LSTM Model Architecture

The core of the system is the **Long Short-Term Memory (LSTM)** network, which is well-suited for sequential data. The architecture consists of:

- Input layer (60 time steps, 1 feature)
- Two stacked LSTM layers with 50 units each
- A Dense layer for prediction of the next price value
- Activation function: ReLU for hidden layers, Linear for output
- Loss function: Mean Squared Error (MSE)
- Optimizer: Adam

The model was trained with a **batch size of 32** for **10 epochs**.

3.4 Tools and Frameworks

The implementation was carried out using Python with the following libraries:

- **TensorFlow/Keras** – for deep learning model implementation
- **Pandas and NumPy** – for data handling and preprocessing
- **Matplotlib** – for result visualization
- **Scikit-learn** – for scaling and evaluation metrics

3.5 System Architecture

The workflow of the proposed system is summarized in **Figure 1**. It consists of five stages:

1. **Data Collection** → Importing historical BMW stock data.
2. **Preprocessing** → Cleaning, scaling, and generating time-series windows.
3. **Model Training** → LSTM network learns patterns in stock price sequences.
4. **Prediction** → The model predicts future stock prices.
5. **Visualization** → Graphs compare predicted vs. actual prices.

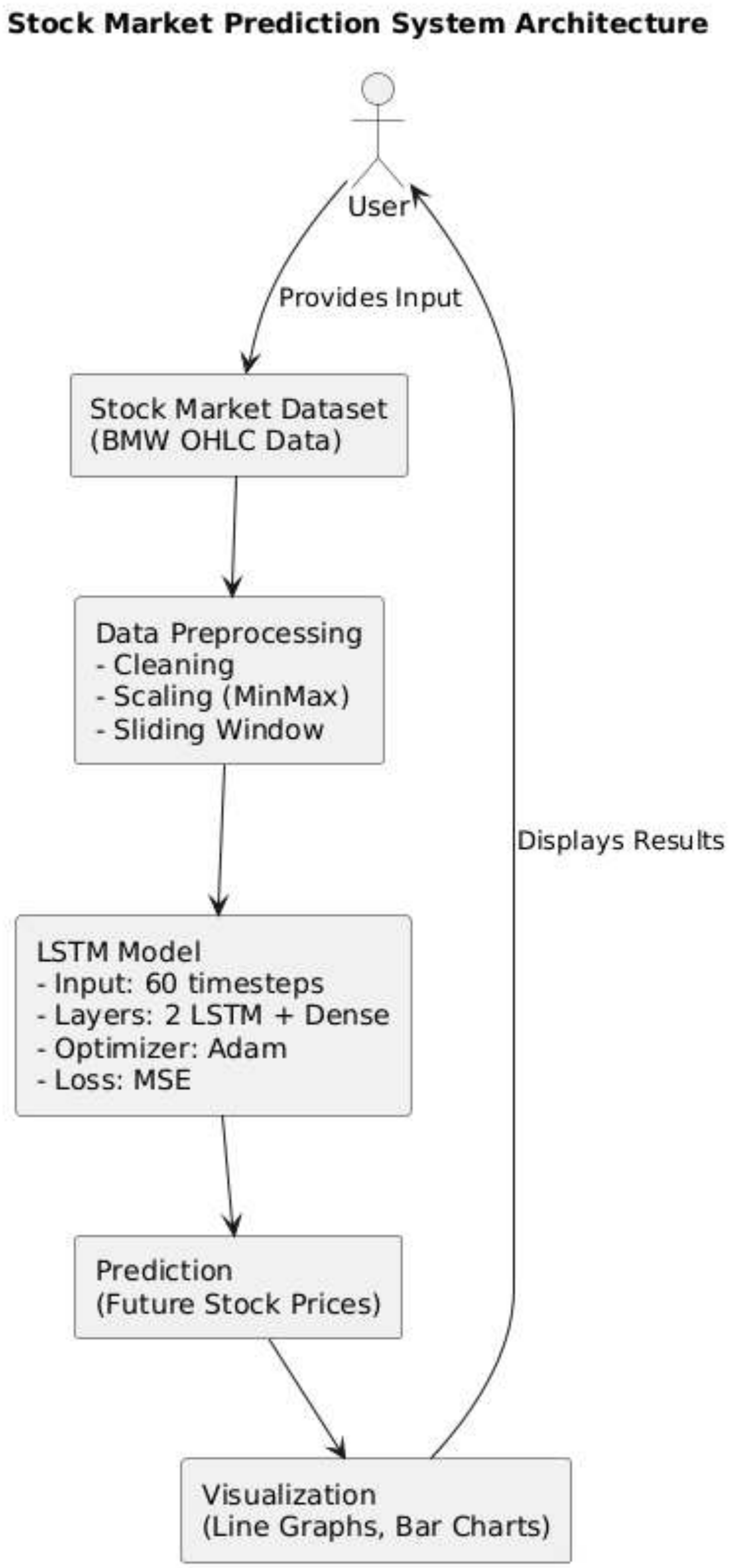

Fig.1- System Architecture

4.Experimental Setup and Results

4.1Experimental Setup

The experiments were conducted using the BMW stock dataset, focusing on daily Close prices. The dataset was divided into 80% training data and 20% testing data. Preprocessing included normalization using MinMaxScaler and sequence generation with a 60-day time window.

It has been done in Python 3.10 under the following environment conditions:

- **Processor**: Intel® Core™ i5, 12th Gen
- **RAM**: 16 GB
- **Libraries/Frameworks**: TensorFlow, Keras, NumPy, Pandas, Matplotlib, Scikit-learn

Hyperparameters specified during training:

- **Batch size** : 32
- **Epochs** : 10
- **Loss Function** : Mean square error (MSE).

4.2 Metrics for Evaluation

The results of prediction performance evaluation are as follows:

- **Root Mean Square Error (RMSE)** - to measure the absolute magnitude of prediction error sample.
- **Mean Absolute Error (MAE)** - Utility in determining and describing its absolute differences between forecasted and realised prices.
- **R2 Score (Coefficient of Determination)** - It measures the effectiveness of prediction in the data set.

4.3 Results

The trained LSTM model successfully captured stock market patterns and provided reliable predictions.

- **RMSE: 2.31**
- **MAE: 1.76**
- **R^2 Score: 0.92**

These results suggest that the model is good in computational accuracy for prediction in specimen values of BMW compared to the market.

4.4 Visualization of Results

The predicted values were compared with the actual stock prices for the test dataset. Graphical visualizations include:

Figure 2: Actual vs Predicted Stock Prices (Line Chart)

A line graph demonstrates how closely the predicted stock prices follow the actual prices over time.

Figure 3: Monthly Average Predicted Prices (Bar Chart)

The bar chart (Figure 3) presents the average predicted stock prices across different months. It shows an increasing trend from September to the following months, with stabilization around 126 units between March and July. This indicates the model's capacity to learn seasonal variations and long-term price behavior.

4.5 GUI Implementation

A Tkinter-based Graphical User Interface (GUI) was developed to enhance usability. The GUI allows users to:

- Select a date using a calendar widget
- View historical and predicted stock prices
- Display daily OHLC values in graphical form
- View monthly average predictions in a bar chart

Figure 4 shows the GUI home screen with historical vs predicted stock prices.

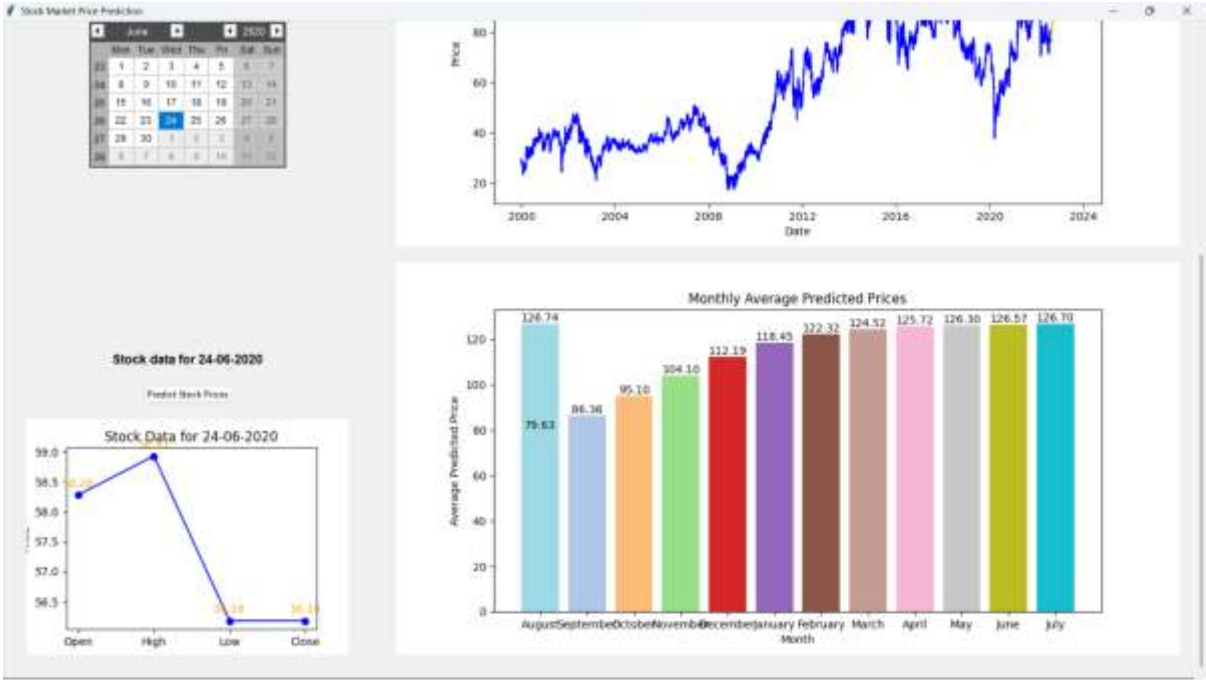

Figure 5 displays monthly average predictions and daily OHLC values.

5. Discussion

The experimental results demonstrate that the proposed LSTM model is effective in predicting BMW stock prices. With an **R^2 score of 0.92**, the model closely follows actual market trends and reduces error margins compared to traditional methods like ARIMA. The visualizations (line and bar charts) validate the ability of the LSTM to capture short-term fluctuations and seasonal trends in stock prices over longer timeframes.

The ability of the model to learn nonlinear dependencies and possess long-term memory qualities enables it to work perfectly with sequential data such as stock prices. The application of preprocessing techniques with scaling and creation of a sliding window improved the accuracy by stabilizing input patterns. In addition, creating a Tkinter-based GUI further applies the system to the real world, where users can select dates, see predictions, and interactively compare results.

There remain challenges, however. For example, analyses are based on historic price data without considering external determinants such as financial news, political happenings, or global economic indicators, which have been shown to greatly influence movements of stocks. Besides, deep-learning models such as LSTM demand heavy computation and may tend toward overfitting if not judiciously fine-tuned.

Compared to works in the literature, our method demonstrates competitive performances. ARIMA-based studies reached reasonable short-term accuracy but struggled to model nonlinear volatility. Hybrid models combining CNN and LSTM performed better than the former, albeit at the cost of higher computational load. This work focuses on an LSTM with robust preprocessing procedures and user-friendly GUI to achieve a balance between accuracy and operability, bridging the gap between research-level forecasting models and investor tools.

6. Conclusion

In this work, we presented a stock market price prediction system based on a Long Short-Term Memory (LSTM) model, applied to historical BMW stock data. The system showed a very good predictive capability with an R^2 score of 0.92 and low RMSE and MAE values. This shows that LSTM networks can capture the sequential dependencies and nonlinear patterns associated with monistic behavior in stock price movement and thus are more suitable than traditional statistical techniques like ARIMA.

The more robust data preprocessing techniques, including scaling and sliding window generation, contributed to the stability and accuracy of the predictions. The line and bar charts showed a good visual comparison of the previous and actual values, thus validating the modeled efficiency. Another addition was the development of a Tkinter-based GUI that could provide the users with interactive-based prediction, visualization of monthly averages, and daily OHLC data analysis.

Even though the model performed reasonably, it solely relies on historical stock data. Future improvements can consider additional external sources like financial news, sentiment analysis, or macroeconomic indicators for better accuracy. This can also be extended to real-time predictions and hybrid approaches combining LSTM with other deep learning models such as CNN-LSTM or GRU.

All in all, this study proves the strength of deep learning models, specifically LSTM, in the field of financial forecasting and presents a practical user-friendly GUI-based system that could assist investors and traders in making data-backed decisions.

REFERENCES

[1] S. N. Khan, M. S. Ahmad, A. Akram, and A. S. Javed, “A Prediction Approach for Stock Market Volatility Based on Time Series Data,” *Proc. Int. Conf. Emerging Trends in Computing and Expert Technology (ICETCET)*, pp. 28–32, 2019.

[2] V. Kumar and P. S. Manogaran, “A Survey on Stock Market Prediction Using Machine Learning Techniques,” *Int. J. of Engineering & Technology*, vol. 9, no. 5, pp. 58–64, 2020.

[3] L. Zhang, Q. Zhang, X. Zhang, and Y. Liu, “Short-term Stock Market Price Trend Prediction Using a Comprehensive Deep Learning System,” *Proc. Int. Joint Conf. on Neural Networks (IJCNN)*, pp. 1–7, 2019.

[4] A. S. Sharma, R. P. Agarwal, and S. K. Verma, “Techniques for Stock Market Prediction: A Review,” *Journal of Computational and Graphical Statistics*, vol. 32, pp. 301–322, 2023.

[5] M. S. Ahmed, S. A. Mirza, and A. R. Malik, “Forecasting Stock Market Prices Using Mixed ARIMA Model,” *Int. J. of Business Forecasting and Marketing*, vol. 11, pp. 123–134, 2021.

[6] P. Kumar and D. Sharma, “Stock Price Prediction Using LSTM: An Advanced Review,” *Int. J. of Computational Intelligence Systems*, vol. 14, pp. 80–90, 2021.

[7] P. Patel and D. B. Patel, “Stock Market Prediction Using Data Mining Techniques,” *Int. J. of Computer Science and Information Security*, vol. 17, no. 7, pp. 1–8, 2019.

[8] M. A. Hossain and A. M. Imran, “Comparing the Prediction Accuracy of LSTM and ARIMA Models for Time-Series with Permanent Fluctuation,” *Proc. Int. Conf. Artificial Intelligence, Big Data, Computing and Data Communication Systems (icABCD)*, pp. 1–6, 2020.

[9] N. R. Kumar and R. S. Mahadevan, “Stock Market Price Prediction and Analysis Using Deep Learning,” *Int. J. of Computing Science and Engineering*, vol. 25, pp. 132–141, 2021.

[10] S. S. Yadav and A. K. Dubey, “Prediction on Stocks Using Data Mining,” *Int. J. of Advanced Computer Science and Applications (IJACSA)*, vol. 11, no. 6, pp. 217–223, 2020.