ISSN No:-2456-2165

Multivariate Petroleum Price Prediction Model with CNN-LSTM using Attention Mechanism

[1]Muhammad Bilyaminu Abdullahi
Department Of Computer Sciences
Federal Polytechnic, Bali araba, Nigeria

[2]Kabiru Ibrahim Musa
Department of Management Information Technology
Abubakar Tafawa Balewa University Bauchi, Nigeria

[3]Abdulsalam Ya'u Gital
Department Of Mathematical Sciences
Abubakar Tafawa Balewa University Bauchi, Nigeria

[4]Auwal Nata`ala; [5]Abubakar Umar Lawan
Department Of Computer Sciences
Federal Polytechnic, Bali Taraba, Nigeria

[6]Ismail Zahraddeen Yakubu
Department of Computing Technologies
SRM Institute of Science and Technology Kattankulathur, Chennai, India, 603203

Abstract:- As one-third of the world's energy consumption, crude oil is vital to the global economy, yet because of its volatility and complexity, it is still difficult to estimate its price.Although machine learning models might enhance forecasts, they are not impervious to unanticipated shocks, geopolitical events, and uncertainty in the world economy. Few studies have employed hybrid models to increase prediction accuracy, despite a large body of research on machine-learning models' potential to improve forecasting. The current approach for predicting petroleum prices forecasts for a short time horizon (10 days) by ignoring outside data that could enhance the prediction performance. Though useful for many in the oil and gas sector, short-term petroleum price forecasting has some limitations and challenges of its own, including limited accuracy, volatility and uncertainty, and a potential inability to fully account for the unpredictable effects of government policies on petroleum prices. In the oil and gas sector, medium- to long-term predictions may offer more consistent and trustworthy direction for strategic planning. This degree of selective attention is also lacking in current skip connection-based forecasting algorithms. When the network is producing predictions, attention mechanisms enable it to choose focus on distinct segments of the input sequence. As a result, the skip connection increases the model's computational complexity, necessitates a large amount of memory, adds noise and redundancy, and needs to be carefully designed and tuned to fit the network architecture and data domain. In order to increase the accuracy of petroleum price predictions, this study suggests combining the benefits of long short-term memory (LSTM), CNN, and attention connection. The proposed model outperformed the classical skip base CNN-LSTM algorithm, which came in second place with an MAE and RMSE of 0.0231 and 0.0297, and skip base CNN-GRU, which achieved the highest MAE and RMSE of 0.0236 and 0.0318, respectively, according to experimental results on MATLAB 2022a. The proposed model also achieved the lowest MAE and RMSE values of 0.0175 and 0.0199.

***Keywords**:- Attention; Deep Learning; Convolutional Neural Network; Long Short-Term Memory; Machine Learning; Petroleum Price and Forecasting.*

I. INTRODUCTION

Petroleum, the world's most traded commodity, is the primary energy source for economic activity. Petroleum is vital to civilization and industry, and it also contributes significantly to global politics and international relations since it meets a large portion of the world's energy needs. It is still the world's most important energy source, accounting for 33% of total primary energy use [1]. Its importance is demonstrated by the fact that 94% of all energy used is utilized in the transportation sector, where it is dominant. Fuel prices have long been a source of anxiety for people in all social classes.Petroleum price fluctuations directly affect the country's economy by affecting consumers, producers, retailers, and the public and private transportation sectors. It also indirectly affects inflation, stock market performance, commodity prices, and other key facets of contemporary man's life cycle [1].

Furthermore, governments can boost economic stability and effectively prepare for significant fluctuations in oil prices through reliable forecasting The literature [2] reports that the oil market is extremely turbulent. In addition to supply and demand, additional variables that impact crude oil prices include stock prices, political unrest, economic activity, and more. Because oil is chaotic and unpredictable, it is therefore very difficult to predict, even if oil price prediction is very crucial [2].

As a result, numerous methods have been put out in the literature to forecast the price of gasoline [3]. To solve this issue, both machine learning and conventional statistical models have been used [4[Using machine learning to predict petroleum prices is a challenging endeavor that calls for extensive data processing and domain expertise. Nonetheless, it is feasible to forecast future oil prices by analyzing previous data using machine learning models.

Developing a machine learning model to forecast petroleum prices can be done in a number of ways. Using time series analysis techniques like Seasonal Autoregressive Integrated Moving Average (SARIMA) or Autoregressive Integrated Moving Average (ARIMA) is one popular approach [5]. By identifying patterns and trends in the previous data, these models are able to predict future prices.Another strategy is to forecast future prices using machine learning methods like neural networks, support vector regression (SVR), and random forests. Numerous variables, including supply and demand, weather patterns, geopolitical events, and economic indicators, can affect oil prices and are included in these models. However, these strategies frequently proven to be unsuccessful due to the crude oil market's generally unpredictable nature.

Furthermore, unforeseen or abrupt fluctuations in oil prices brought on by peculiar occurrences might significantly impair the forecasts' accuracy. Recent research has investigated the application of hybrid models to forecast the specifics of changes in crude oil prices in such circumstances.The application of hybrid models to enhance the prediction of crude oil prices based on outside data has increased. Nevertheless, research employing these models has concentrated on forecasting crude oil prices through a single-step, univariate strategy utilizing traditional deep-learning techniques. The application of hybrid models to enhance the prediction of crude oil prices based on outside data has increased. Nevertheless, research employing these models has concentrated on forecasting crude oil prices through a single-step, univariate strategy utilizing traditional deep-learning techniques [6].

You will need to compile a sizable dataset of past oil prices together with other data points, such stock market indices, geopolitical events, and oil production and consumption statistics, in order to create an efficient petroleum price prediction model. In addition, the data must be meticulously cleaned and preprocessed before being incorporated into your model. It is crucial to remember that a variety of intricate and unpredictable factors affect petroleum prices, therefore any forecasts produced by a machine learning model should be considered estimations rather than exact forecasts. Furthermore, the price of petroleum is highly volatile and can vary quickly in response to unforeseen circumstances or modifications in the market. Therefore, additional factors must be included, such as text-related data from news articles and social media sources, to overcome uncertainties with this kind of prediction [6]. These texts can offer comprehensive and contextual information on how oil prices fluctuate in a more descriptive manner. Thus, this research employs a hybrid model using CNN and Long Short-Term Memory (LSTM) with an attention connection mechanism to improve the forecast accuracy.

II. RELATED WORK

This section aims to provide a review of the existing literature related to the paper's specific topic on the prediction of the crude oil price and the use of artificial neural networks or other artificial intelligence algorithms. There were a number of researchers that showed the ability of machine learning techniques in studying the predictions in the price of crude oil. In relation to that, [7] introduced Crude Oil Price Prediction Based on a Dynamic Correcting Support Vector Regression Machine. For the authors to make the forecasting result more accurate, the HRGA is applied to the optimized parameters of ε-SVR. The predicted result is very good. But in cases where the number of features for each data point exceeds, the SVM will underperform.

[8]suggested Particle Swarm Optimization-Based Intelligent Prediction System for Gas Metering System Training Neural Networks. This study presents a comparison of the performance of PSONN and GANN with pure ANN. The findings demonstrate that the suggested PSONN model produces encouraging outcomes in terms of gas measurement prediction accuracy. With the expert integration system, IPS can be expanded for other industry applications in the future.

Introduced a novel method for predicting crude oil prices based on stream learning. The researchers compared our stream learning model to three other widely used oil price prediction models in order to assess its forecasting accuracy. The experiment results demonstrate that over a range of forecast time horizons, the stream learning model achieves the highest accuracy in terms of both mean squared prediction error and directional accuracy ratio. Nevertheless, large amounts of data processing and storing can be challenging due to the memory and processing resources that are typically limited in streaming learning algorithms.

[10] made Crude Oil Price Forecasting available. XGBoosting has repeatedly shown itself to be a successful prediction technique for problems including both classification and regression. However, XGBboosting can be memory-intensive, particularly when dealing with large datasets, which makes it less appropriate for systems with constrained memory. [11]provided a random wavelet neural network-based accuracy evaluation and global crude oil price forecast using synchronization. The empirical findings show that the suggested model is helpful in increasing forecast precision and has a greater accuracy in terms of changes in the price of crude oil. On the other hand, Wavelet Neural Networks can be prone to overfitting, which happens when the model is too complicated and fits the training data too closely, leading to poor performance on the new dataset. This can make it challenging to read and understand the results of these networks.

[12] introduced a predictive model for estimating petroleum consumption using a machine learning approach, the result obtained revealed that the two machine learning models: LR and RFR outperformed the ARIMA model with lower values of prediction accuracy in terms of MAE, MAPE, RMSE, and R2. However, the prediction can be improved if the parameters of the machine learning models were tuned.

[2] proposed Oil price forecast using deep learning and ARIMA, the results show that a convolution neural network yields the best performance with a relatively simple structure and lesser training time. However, exploring the ensemble of several different models than improve the crude oil price prediction accuracy. Moreover, it will also be interesting to combine traditional trend and seasonality models while applying neural networks specifically to model the residuals to improve the accuracy of long-term prediction.

[13] presented a crude oil price prediction using complex network and deep learning algorithms, the experiments show that the proposed model has higher accuracy, and is more robust and reliable. However, the paper only considers crude oil prices in Nigeria, without necessarily considering other factors such as financial market, economic growth, dollar exchange rate, demand and supply etc. The model proposed in this work is built based on monthly data, which restricts the prediction horizons to months.

[14] constructed a Crude oil price prediction model with long short-term memory deep learning based on prior knowledge data transfer. The authors adopt several evaluation methods to compare the predictive ability of Type I, Type II, and Type III data. The comparison results of MAE, RMSE, MAPE, SMAPE, TIC, and R show that Type III predicting performance is better than Type I and Type II. However, LSTMs are unable to handle temporal dependencies that are longer than a few steps. The researchers found that when they trained an LSTM on a dataset with long-term dependencies (e.g., 100 steps), the network struggled to learn the task and generalize to new examples.

[15] proposed Oil Price Predictors using a Machine Learning Approach. After analyzing the results and comparing the accuracy of the model first, we can conclude that oil prices in 2019-2022 will have a slight upward trend and will generally be stable. However, the use of other machine learning algorithms or the identification of new factors of influence for obtaining more accurate predictions.

[16]Using machine learning to generate a crude oil price forecast, the author deduces that the AR model has the highest MSE value. The MSE value of MA changes significantly when compared to AR. Ultimately, the ARIMA model has the lowest MSE value, as we can see. The relationships and interdependencies between various variables, as well as the impact of outside variables on the time series, including marketing, competition, or events, are not, however, captured by ARIMA models.

(17) established a multi-step-ahead crude oil price forecasting model using an extreme learning machine optimized by the particle swarm optimization algorithm and a two-layer decomposition technique. The empirical results demonstrate that the prediction model suggested in this paper increases the accuracy of crude oil price predictions. The trend of crude oil price series, however, is affected by multidimensional and complex factors. However, other influencing factors that affect crude oil price changes can be considered to further improve the overall forecasting effect.

[18] predict the Price of Crude Oil and its Fluctuations Using Computational Econometrics: Deep Learning, LSTM, and Convolutional Neural Networks. However, the use of advanced techniques such as optimization and photogrammetry in the oil industry to study multiple aspects and characteristics.

[19]utilized The authors have shown that the AdaBoost-LSTM and AdaBoost-GRU models outperform the benchmarking models as expected, and the empirical results demonstrate that the AdaBoost-GRU is superior to all models studied in this research. Crude oil price prediction: A comparison between AdaBoost-LSTM and AdaBoost-GRU for improving forecasting performance. As a result, predicting crude oil prices using the AdaBoost-GRU ensemble-learning model of all other models shows promise. However, AdaBoost employs a regressive learning boosting method. High-quality data is therefore required. Additionally, it is highly susceptible to noise and outliers in the data, therefore these elements must be removed before utilizing the information. Moreover, compared to the XGBoost algorithm, it is far slower.

[20] introduced a time series Analysis of Nigerian Monthly Crude Oil Prices. This price rise may take the Nigerian government out of unchartered waters and the worst recession in 40 years. This will give the Nigerian government the chance to revive the economy only if they diversify the economy, create job opportunities, combat insurgency, and put in place zero-tolerance measures against corruption and mismanagement of public funds. But ARIMA models require a lot of data preprocessing and tuning, as you need to check the stationarity, autocorrelation, and partial autocorrelation of the data, and find the optimal values of the parameters using trial and error or grid search.

[21] presented a process of modeling the Nigerian Bonny Light crude oil price: the power of fuzzy time series, The fuzzy time series model is hereby recommended as the best model for estimating the price of the Nigerian Bonny Light crude oil. But fuzzy time series require subjective decisions, especially in the fuzzification stage. In some fuzzy time, series methods, membership values are ignored, but the membership values are vital in a fuzzy inference system.

[22] proposed The Optimal Machine Learning Modeling of Brent Crude Oil Price, the findings indicate that the ANN and DNN models are highly accurate in forecasting the price of Brent crude oil, and it is advised that they be used consistently to both closely monitor the price of Brent crude oil and to anticipate future prices. However, the DNN model lacks interpretability, which makes it challenging to comprehend how the model generates predictions. In addition, the model can be computationally expensive, needing a significant amount of

ISSN No:-2456-2165

time and computing power to train and optimize the network.

[23] proposed a Multi-Step Crude Oil Price Prediction Based on LSTM, it was concluded that the proposed hybrid approach is promising for crude oil price forecastingsurpassing all other contestants. However, alternative signal decomposition techniques can also be utilized. Additional datasets on crude oil prices and state-of-the-art metaheuristics can be employed for validation. It is also possible to employ stacked LSTM models and LSTM coupled with layers of a convolutional neural network (CNN). All of this would be far too much to apply in a single study, though, and future research in this challenging field is probably going to cover every topic that has been addressed.

[24] proposed Artificial Intelligence-Based Prediction of Crude Oil Prices Using Multiple Features under the Effect of the Russia–Ukraine War and COVID-19 Pandemic, There is a comparison between the outcomes produced by deep learning and machine learning algorithms. Finally, our approach can accomplish high-performance estimate with an average mean absolute error value greater than 0.3786. For more precise forecasts, we intend to create DL networks. Furthermore, one approach whose inference system is a collection of fuzzy rules with learning capabilities is the neuro-fuzzy inference system (ANFIS). This approach should be seen as one that can aid in improving prediction, along with other algorithms like teaching-learning-based optimization (TLBO).

[25] proposed modeling liquefied petroleum gas prices in Nigeria using time series machine learning models, concerning the result of the MAPE, the naive models produced lower MAPE values for most states compared to the NNETAR and ARIMA models, implying that the naive model produces valuable predictions of the price of 12.5kg refilling LPG for most states in Nigeria. However, the model required significant trial and error to develop and optimize, which can be time-consuming and require a lot of data.

[26] predict crude oil prices in Nigeria with machine learning models, the NNETAR model yielded the optimal forecast performance for crude oil prices in Nigeria as it had the lowest RMSE and the MAE among all other models it was concluded that the proposed hybrid approach is promising for crude oil price forecasting, outscoring all competitors. However, by incorporating crude oil price interventions over the years to understand the price changes.

Recently in 2023, [6] proposed petroleum price prediction with CNN-LSTM and CNN-GRU Using Skip-Connection, the proposed models both showed significantly better performance than the existing vanilla LSTM and GRU models. However, the research ignored external influential factors that can improve the forecast accuracy.

III. METHODOLOGY

This section is subdivided into five subsections that describe the: (1) multivariate input, (2) CNN-based architecture, (3) LSTM/GRU-based architecture, (4) attention connection technique, and (5) output layer to enhance the prediction accuracy of petroleum prices. The proposed framework is presented in Fig. 1.

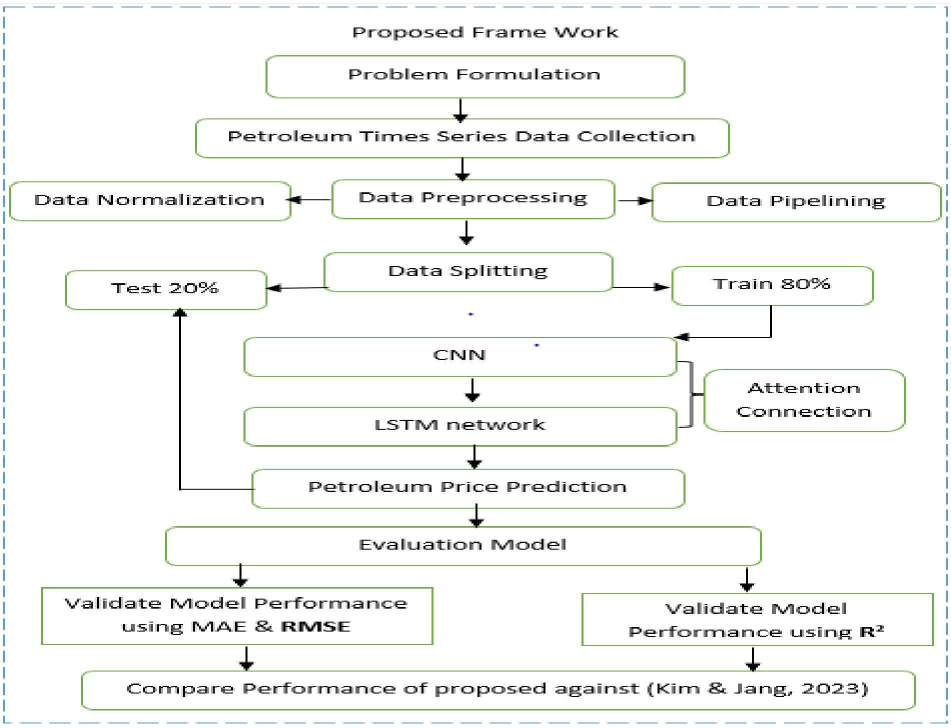

Fig. 1: Framework for the proposed study

The description of each of the stages is further elaborated in the following subsections.

A. Multivariate Input Sequences

Seven variables make up our data, which are a multivariate time-series sequence: (1) regular petroleum price; (2) premium petroleum price; (3) diesel price; (4) WTI crude oil price; (5) Brent oil price; (6) Dubai oil price; and (7) heating oil price. The sequential data of the variable is represented by each row in our data, which is a vector whose length varies based on the number of previous timesteps (t) in our input layer. By using more historical timesteps in the data, the sample size is expanded, which improves prediction accuracy. Afterward, we feed our data into the CNN layers.

B. CNN

CNN models have performed admirably across a number of picture datasets. The convolutional layers and pooling layers make up the CNN architecture. The convolutional layers are specifically made to filter and extract useful features from images, like channels, color, and resolution. For each convolutional layer, there is a convolution kernel, or filter, which functions as a small window and sequentially slides from left to right over a sub-region to extract the essential but minimal aspects of a given image through complex convolution operations. A rectified linear unit (ReLU), a nonlinear activation function, and a pooling layer come next.In order to extract unique features from the raw input image data, the CNN subsamples, removes particular values from the convolved features, and shrinks the dimensions of the matrix in each pooling layer. For instance, the most important feature properties of that kernel are represented by the max-pooling approach, which computes the largest value from each window shift. Therefore, before the fully linked layer is employed, the pooling layer generates matrices of various dimensions that are subsequently processed layer by layer to obtain the final pooled output value. The robust properties of our time-series data can be captured by a CNN architecture without requiring iterative extension of the matrix's dimensions. After the CNN layers, we then process our data using an RNN, which considers the long-term features of the data.

C. LSTM

The vanishing gradient problem stated earlier was suggested to be resolved with LSTM. As seen in Figure 2, an LSTM contains three gates. Information can be optionally let via the gates. They consist of a pointwise multiplication operation × and a sigmoid neural net layer σ. The first is the forget gate (ft), which checks to see if the preceding cell's state Ct-1 is reflected as follows:

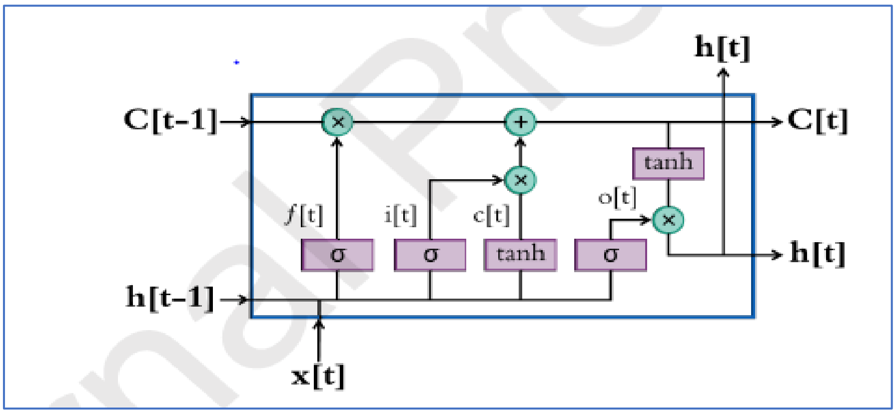

Fig. 2: The Structure of a Long Short-term Memory Cell.

As seen in Figure 8, the forget gate (f (t)) chooses whether to retain or discard the data from the cell state. A logistical function yields a result for f (t) of 0 or 1, respectively, denoting the abandonment or maintenance of the current cell state at time step t. where the weight, output, input, bias, and x, y, and z are, in that order. The symbol for the activation function is ¾. However, as Eq. 4 illustrates, the input gate determines which input values are permitted to be stored in the cell state. where g (t) is the new candidate value, c(t) is the new cell state, and i (t) is the signal (0 or 1) that governs the updating process.Furthermore, the LSTM is a potent machine learning model that can capture both long-term dependency and non-linear relationships in a complex dataset since the output gate (o(t)) is in charge of releasing the stored information to the subsequent neurons. Compared to an RNN, the state Ct-1 of the preceding cell can be altered less in an LSTM. Consequently, one benefit of LSTM is that it can more accurately reflect a cell's initial state.

D. Attention Connection

In this research, the classic LSTM with Attention was proposed. The ability of a model to concentrate on particular components inside the data—in this case, the hidden state outputs of LSTMs—is referred to as attention in machine learning. To obtain state-of-the-art outcomes, the suggested design consists of an attention layer inserted between traditional LSTM layers. An interesting example is given by deep learning models with attention. They frequently provide interpretability in the form of attention heat maps along with overall performance gains. They are also frequently faster than employing a traditional LSTM. In order to anticipate the target time series, this seeks to identify pertinent information from among the many aspects of the electrical time series data.

Different trading information has varying priorities in petroleum datasets; certain information is superfluous, while other information could be crucial. We also need to prioritize important features and eliminate unnecessary features when predicting the energy index. As a result, an energy forecast model can be established and decisions can be guided by the useful information from various time periods. The study suggested an attention link to optimize the input feature sequence in the prediction of petroleum time series, drawing inspiration from the previously mentioned data.

We therefore use attention connections for the layers of CNN and LSTM. As a result, we add connections to the hybrid model and incorporate them into our proposed CNN-LSTM hybrid model. At present, A common technique found in many deep learning architectures is attention connections. Thus, we implement an altered version of the attention layer immediately before the fully connected layers, much like in the skip architecture. We present CNN-LSTM-Attention, a proposed model that comprises of two max-pooling layers with a pool size of 3 × 2, two convolutional 1-dimensional (1D) layers of 512 × 512 filters with a kernel size of 3, and two LSTM layers with 512 units each. Then, to concatenate the data and create the attention connection layer, we employ flattened layers as both our last encoder layer and the first input layer of the LSTM layers. Ultimately, the output is generated by applying the activation function. The newly obtained model can pay more attention to the specific input feature sequence, extract the key feature sequences effectively, and eliminate the influence of the redundant feature sequences because of the attention weight. Theoretically, the prediction accuracy can be better.

E. Dataset Collection and Description

In the study, the data for training the model are taken from a dataset was extracted from the official website of South Korea's national oil corporation Available online: https://www.opinet.co.kr/user/dopospdrg/dopOsPdrgSelect. do (accessed on 29 December 2021) and the official Yahoo Finance Site Available online: https://finance.yahoo.com/ (accessed on 2 January 2022) using a Python-based application programming interface. The National Oil Corporation website provides South Korea's daily petrol oil prices, including regular, premium, and diesel prices, while the Yahoo Finance site provides a range of stock market prices for stockholders and investors, including the daily stock prices, exchange traded funds, exchange rates, and oil prices.

F. Experimental Setup and Choice of Metrics

In this experiment, the simulation approach was adopted to run several computer simulations. This is necessary to evaluate the performance of the model against other models with different algorithms on MATLAB 2021a. The system was implemented with MATLAB version 2021a on a DELL laptop System Corei (TM) i7-4200U, 2.4GHz CPU and 6GB RAM. The system is experimented with using the dataset specified in the design.

Testing the Neural Network is a vital step in the design process. In this research, a Standard Statistical forecasting metric was used to evaluate the network performance in other to select the model with the best performance. The developed network is tested and evaluated by using the Mean Square Error and Root Mean Square Error (RMSE) in the MATLAB Neural Network (NN-tools) package. Also, mean average percentage error (MAPE) and correlation error (R^2) are used to ascertain the model's prediction accuracy. The following mathematical equations are used to calculate the first two evaluation metrics:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - y| \tag{1}$$

$$RMSE = \sqrt{\frac{\Sigma(y_i - \hat{y}_i)^2}{n}} \tag{2}$$

where n is the number of observations, y_i is the i_{th} observed value, $\acute{Y}_i$ is the corresponding predicted value, and y is the true value (or ground-truth value). RMSE is simply the root of the mean squared error. We calculated how well these sets of data are related to each other. Finally, the coefficient of determination (R^2) provides us with an overall guideline for tracking the goodness of fit of our proposed model throughout the next n-days of prediction. It is calculated as follows:

$$1 - \frac{\Sigma(y_i - \hat{y}_i)^2}{\Sigma(y_i - \bar{y}_i)^2} \tag{3}$$

where y_i is the vector of actual values, Ýi is the vector of predicted values, and $\bar{y}_i$ is the mean value of the y values.

IV. RESULTS

To determine whether model produces a more accurate forecast utilizing this research, the accuracy of the predicted forecasts of the proposed model will be compared with the other models. RMSE, MSE, and correlation error (R). Evaluating a petroleum forecasting model is crucial to ensure its accuracy and reliability in predicting petroleum prices or related metrics. Now, we'll give the forecasted result based on all of the algorithms, rank each model according to its performance criterion, and talk about the results. The primary objective of the simulation's generated graphs is to provide the reader with a summary of the trend's continuation rather than a direct analysis. The proposed model was implemented using the parameter settings depicted in table 2. 70% of the data was used for training and 30% was used for testing. Training the proposed deep learning model with 70% of the data and reserving 30% of the data for testing is a common practice in machine learning and deep learning for several important reasons: The primary purpose of reserving a portion of the data for testing is to evaluate the model's performance on unseen data. It allows us to assess how well the model generalizes to data it has not been exposed to during training. This split helps in understanding the trade-off between bias and variance. The training data is used to fit the model, and the testing data is used to evaluate how well it generalizes. If the model performs well on the training data but poorly on the test data, it may be overfitting (high variance). Conversely, if it performs poorly on both, it may be underfitting (high bias). Therefore, splitting the data into training and testing sets is a fundamental practice in machine learning and deep learning to ensure that models are developed, evaluated, and fine-tuned effectively. It helps strike a balance between fitting the data well and ensuring that the model can make accurate predictions on new, unseen data. Table 1 depict the summary of the overall parameters use for the implementation of the system.

Table 1: Parameter Settings

Parameter	**Settings**
Normalization	MinMax
Input Layer	1
Output Layer	1
Train ratio	70
Validation	30
Input Size	1
Num Responses	1
Convolutional Layer	4
Max Pool Layer	1
Num Hidden Units	200
Sequence Input Layer	Input Size
LSTM Layer	4
Attention Layer	1
Fully Connected Layer	1
Regression Layer	1

A. Evaluation of the Predictors

Evaluating a petroleum forecasting model is an iterative process that involves refining the model based on feedback and performance metrics. It's crucial to take into account the particular requirements. and objectives the forecasting task and to select appropriate evaluation methods accordingly. It is important to choose appropriate evaluation metrics based on the specific task and goals of the forecasting model. In this research, we the common metrics for petroleum forecasting which include:

- **Mean Absolute Error (MAE):** Calculates the mean absolute difference between the values that were expected and those that were not. It offers a sense of forecast accuracy and is simple to comprehend.
- **Root Mean Square Error (RMSE):** Similar to MAE but gives more weight to larger errors. It helps penalize larger deviations from actual values.
- **Coefficient of Determination (R-squared or R^2):** Measures the proportion of variance in the target variable explained by the model. It provides an indication of how well the model fits the data.

Finally, the results are compared with the baseline model to compare the forecasting model's performance in predicting the next value as the current one. From the results of our simulation, we obtain the following plots which will demonstrate how perfectly fit is our model. To validate the developed network, the coefficient of determination (R-squared or R2) is used to show how accurately the proposed trained model fits the dataset.

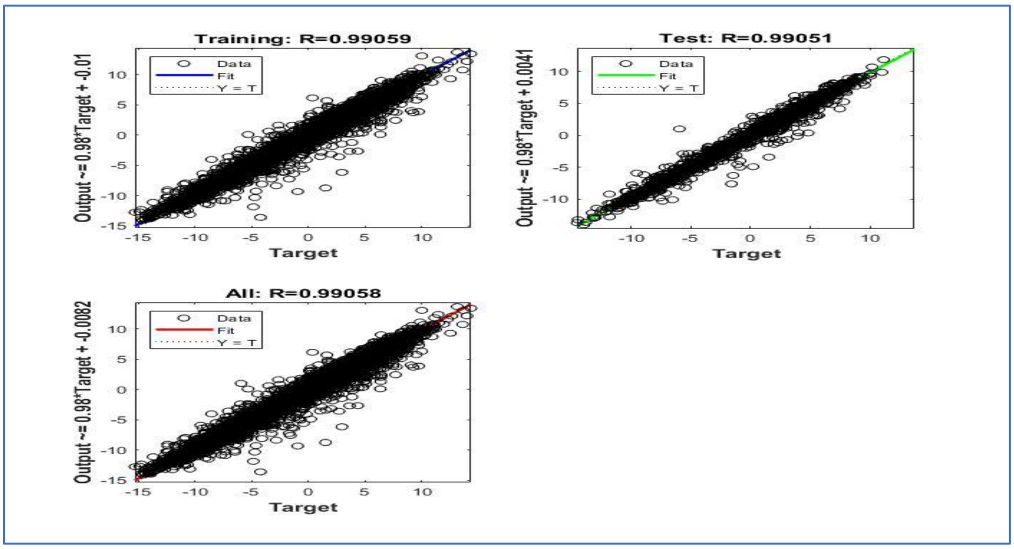

Fig. 3: Coefficient of determination (R-squared or R2)

From Figure 3. It was observed that the value of R is close to 1 (good) which demonstrate that the model prediction is close to the actual dataset. If it was close to zero (bad) then it shows that the model completely fails in making a correct prediction. In our own case, the model prediction was close to the actual test which demonstrate a good prediction by the proposed model as shown in table 2 below. Figure 4. display the error autocorrelation function.

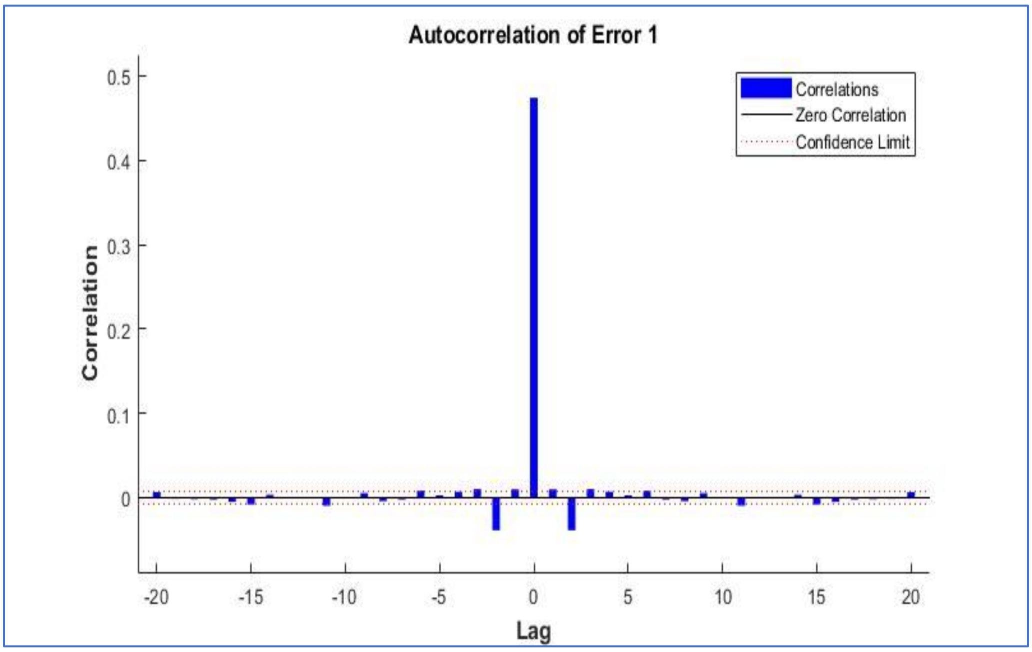

Fig. 4: Auto correlation error for the proposed model

Figure 4 shows how the prediction errors are related in time. Idyllically, for perfect prediction model, there should be only one nonzero value of the autocorrelation function. These nonzero values should occur at zero lag; this is also called mean square error. Such an autocorrelation function would imply complete un-correlation of predicted errors with each other. Base on the graph as shown in Figure 4 the correlations, except for the one at zero lag, fall approximately within the 98% confidence limits around zero, thus the model appears to be suitable. If there were significant correlation in the prediction errors, then it's possible to improve and enhance the prediction accuracy maybe by changing the neural network structure or increasing the number of delays in the network.

B. Result Presentation

Table 2 illustrates how the performances of the proposed attention base CNN- LSTM model performed when compared with those of the Skipp base CNN-LSTM for long term forecast horizon using MAE, RMSE and R2.

Table 2: Performance Comparison with existing method for long term forecast (≥ month)

Model Algorithm	**Connection**	**Variable**	**MAE**	**RMSE**	**R**
Proposed CNN-LSTM	Attention	Multivariate	0.0175	0.0199	0.9906
CNN-LSTM	Skip	Multivariate	0.0231	0.0297	0.9884
CNN-GRU	Skip	Multivariate	0.0236	0.0318	0.9855

For the RMSE and MAE, the smaller the value, the better the performance. In our own case, the model prediction error was very small which demonstrate a good prediction by the proposed model as shown in table 3. This is further supported by the explanations in the subsections below. For more intuitive discussion, each of the predictive accuracy is depicted in the subsequent figures.]

C. Result Analysis

In this subsection we present an intuitive discussion using graphical representation of the results achieved after simulating the method. The performance of the proposed model is compared with CNN-LSTM and CNN-GRU with skipp connection. The analysis is presented for the three metrics used for the evaluation in the proceeding section. For the RMSE and MAE, the smaller the value, the better the performance. Hence, we will evaluate the proposed model with other popular algorithms from the baseline method.

V. MAE

MAE measures the average absolute difference between the predicted values and the actual target values. Lower MAE values indicate better prediction accuracy. Figure 5 depict the MAE score for the proposed attention base hybrid CNN-LSTM algorithms as compare to the other baseline method (Skip base Hybrid CNN-LSTM and Skip base Hybrid CNN-GRU) both with multivariate inputs.

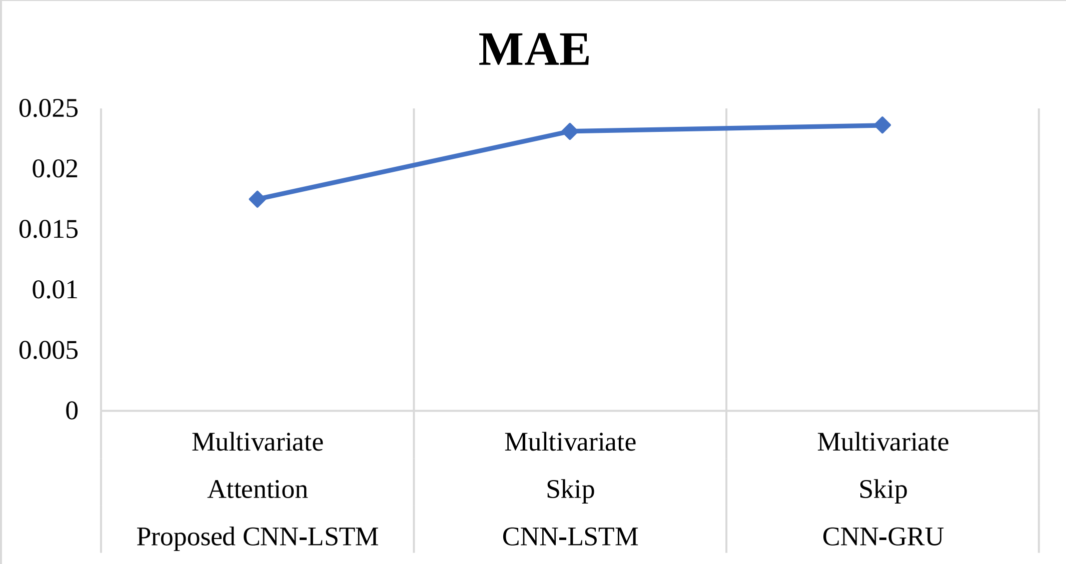

Fig. 5: MAE for all method

From figure 5. The proposed model achieved the lowest MAE value of 0.0175 which the translate to better performance compare to the classical skip base CNN-LSTM algorithm which came second with an MAE of 0.0231and skip base CNN-GRU which attain the highest MAE of 0.0236 respectively. Thus, the proposed model excel at aligning source and target sequences and focusing on relevant predictive feautures in the petroleum price datasets.this makes the attention mechanisms beneficial in few-shot learning scenarios, where the model is required to adapt to new tasks with limited training data. The attention connection helps the model concentrate on relevant information during both training and inference. However, MAE is a useful metric for assessing prediction accuracy, however, it should be considered alongside other evaluation metrics, such as Root Mean Square Error (RMSE). Hence, the RMSE score is further evaluated in the next subsection.

VI. RMSE

RMSE is similar to MAE but gives more weight to larger errors. It calculates the square root of the mean of the squared differences between predicted and actual values. For the RMSE, the smaller the value, the better the performance. Figure 6 depict the RMSE score for the proposed attention base hybrid CNN-LSTM algorithms as compare to the other baseline method (Skip base Hybrid CNN-LSTM and Skip base Hybrid CNN-GRU) both with multivariate inputs.

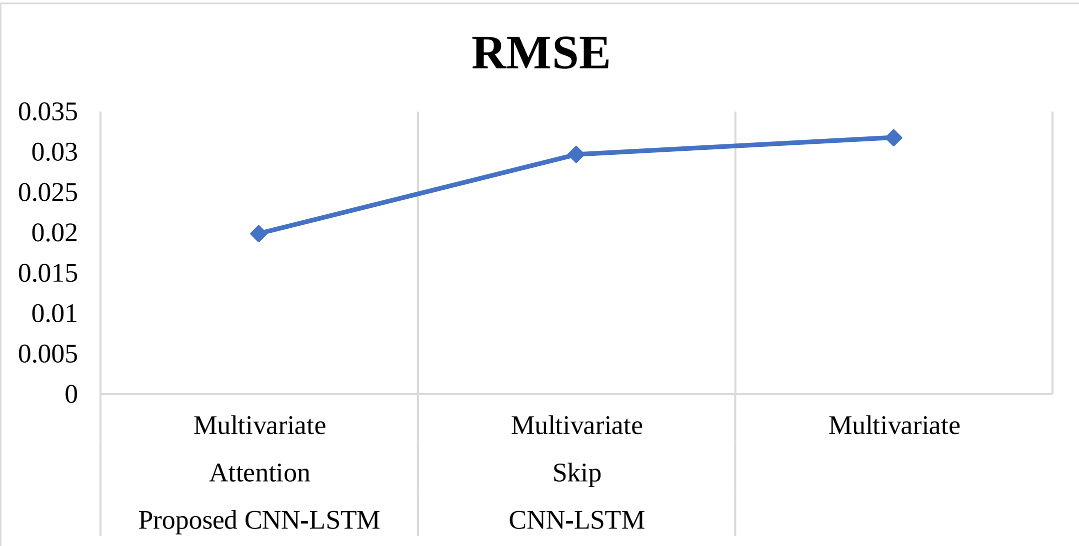

Fig. 6: RMSE for all method

From figure 6. The proposed model achieved the lowest RMSE value of 0.0199 which then translate to better performance compare to the classical skip base CNN-LSTM algorithm which came second with an RMSE of 0.0297 and skip base CNN-GRU which attain the highest RMSE of 0.0318 respectively. While MAE and RMSE are useful metric for assessing prediction accuracy, it should be considered alongside R-squared (R^2), to provide a comprehensive understanding of the model's performance. This is further discussed in the proceeding section.

VII. ERROR AUTO CORRELATION (R2)

R-squared measures the proportion of the variance in the target variable that is explained by the model. It provides an indication of how well the model fits the data, with values closer to 1 indicating a better fit. Figure 7 depict the auto correlation for the proposed attention base hybrid CNN-LSTM algorithms as compare to the other baseline method (Skip base Hybrid CNN-LSTM and Skip base Hybrid CNN-GRU) both with multivariate inputs.

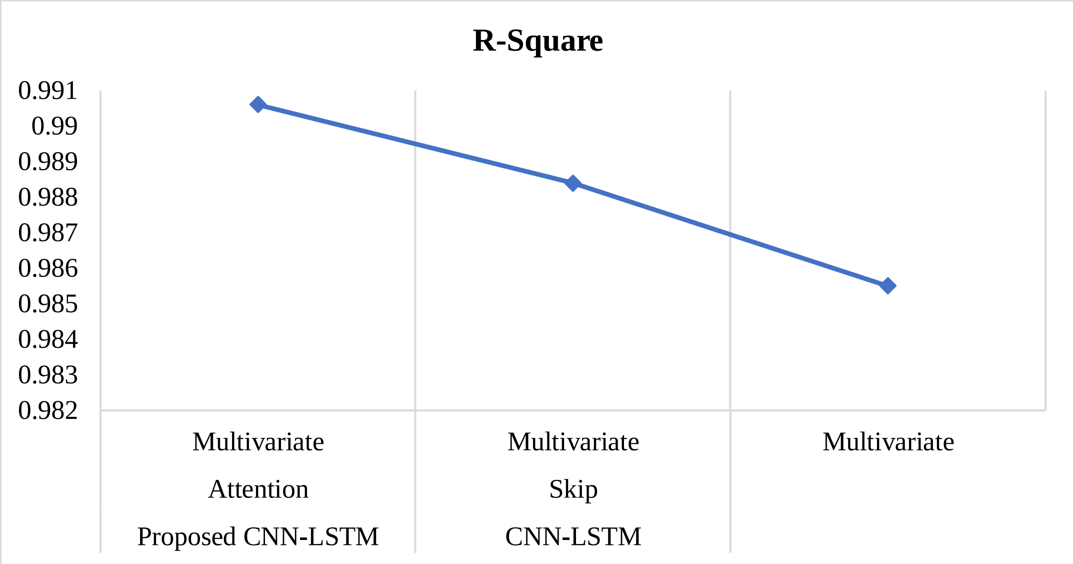

Fig. 7: auto correlation error R^2 for all methods

Hence, from Figure 8. The proposed model achieved a perfect fit of 0.9906. this is better fitness score compare to the other classical models (Skip base CNN-LSTM with 0.9884 and Skip base CNN-GRU with 0.9855). In general, Figure 8 depict the overall predictive performance of the three-method used for MAE and RMSE.

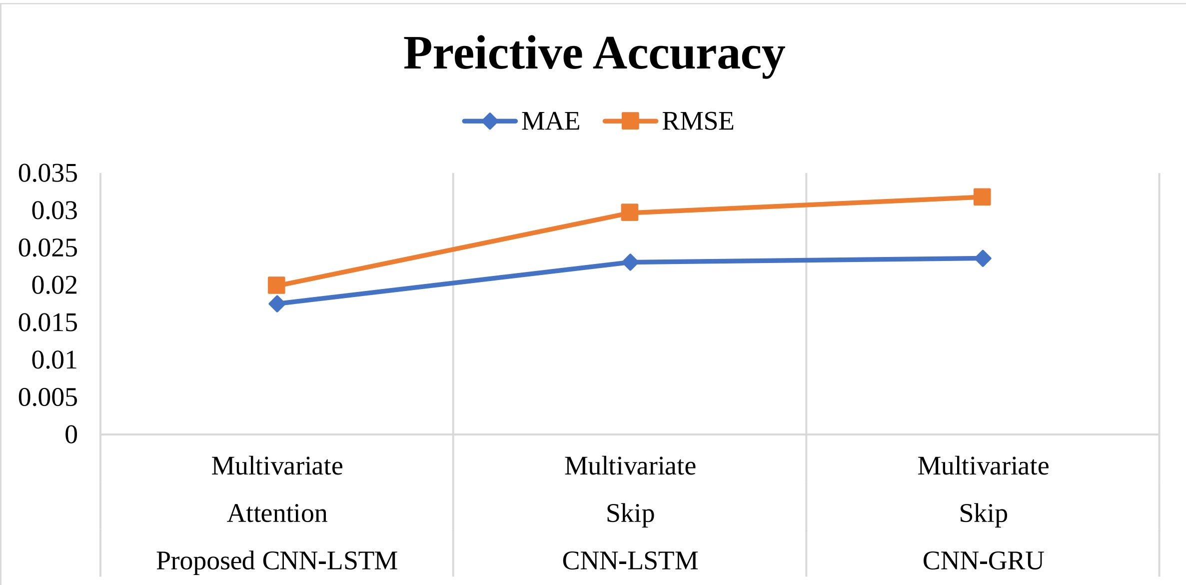

Fig. 8: Result Summary for MAE and RMSE

Therefore, from the results presented and analyzed above, it may be observed that the suggested model had the lowest MAE and RMSE worth of which translate to better performance compare to the classical skip base CNN-LSTM algorithm which came second and the skip base CNN-GRU which, when it came to RMSE and MAE on the model, was the least performing benchmark petroleum forecasting datasets. This result suggest that skip connections are primarily designed to address the vanishing gradient problem by enabling gradients to flow more easily during training. While they are effective in some cases, attention mechanisms can offer an alternative solution by allowing the model to explicitly weigh the importance of different inputs, potentially reducing the impact of vanishing gradients. Attention mechanisms are integral to many sequence-to-sequence tasks including speech recognition and machine translation.They allow the model to generate the output sequence while focusing on various segments of the input sequence, resulting in more accurate and coherent predictions.

Furthermore, it is also clear that the proposed hybrid model attains the highest level of performance by obtaining the best score in terms of model fitness (R^2). This research results have further demonstrated the superiority of integrating attention connection on hybrid algorithm over conventional skip connection.Thus, the proposed attention base CNNLSTM achieved the lowest validation MAE and RMSE values, demonstrating a clear improvement in forecasting accuracy when compared to the other models that were evaluated consequently, our deep learning model that we have proposed shows superiority in forecasting future petroleum prices using historical data with superior precision compared to the most advanced algorithms demonstrating the powerful strength of hybrid models base on attention mechanism in revealing nonlinear and complex patterns in big data.

Hence, the result shows that attention mechanisms in the CNN-LSTM models permit the network to make predictions by focusing on distinct segments of the input stream. This is particularly useful in scenarios where different parts of the input have varying levels of importance or relevance. Skip connections, on the other hand, do not provide this level of selective attention. Thus, the result from the attention mechanisms can provide interpretable insights into which parts of the input sequence the model is focusing on when making predictions. This interpretability can be valuable for understanding model decisions, especially in applications where transparency is crucial.

VIII. CONCLUSION

Petroleum price prediction is an ongoing task, and models should be retrained and updated regularly to adapt to changing market conditions and data patterns. Accurate petroleum price prediction has practical applications in various industries, including energy trading, risk management, transportation, and policy-making. It helps stakeholders make informed decisions and mitigate risks associated with price fluctuations. Petroleum price prediction using hybrid deep learning algorithms combines the strengths of deep learning with other modeling techniques and external features to provide more accurate forecasts in a complex and volatile market. Hence this research proposed to enhance theprediction precision of the petroleum price forecasting hybrid design that combines the advantages of attention connection, CNN and LSTM algorithm. Based on MATLAB 2022a experimental data, the suggested model yielded the lowest MAE value of 0.0175 which the translate to better performance compare to the classical skip base CNN-LSTM algorithm which came second with an MAE of 0.0231and skip base CNN-GRU which attain the highest MAE of 0.0236 respectively. Thus, the proposed model excels at aligning source and target sequences and focusing on relevant predictive features in the petroleum price datasets. This makes the attention mechanisms beneficial in few-shot learning scenarios, where the model is required to adapt to new tasks with limited training data. At 0.0199, the suggested model produced the lowest RMSE value which then translate to better performance compare to the classical skip base CNN-LSTM algorithm which came second with an RMSE of 0.0297 and skip base CNN-GRU which attain the highest RMSE of 0.0318 respectively. These models have the potential to improve risk management and decision-making for businesses and organizations operating in the petroleum industry.

ACKNOWLEDGMENT

We appreciate the expert support of our supervisors Dr. K. I. Musa and Prof. A. Y. Gital for their guidance towards the success of this research.

REFERENCES

[1]. Weldon, K.a., et al., Petroleum prices prediction using data mining techniques--A Review. arXiv preprint arXiv:2211.12964, 2022.

[2]. Guo, J. Oil price forecast using deep learning and ARIMA. in 2019 International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI). 2019. IEEE.

[3]. Xu, Z., et al., Using econometric and machine learning models to forecast crude oil prices: Insights from economic history. Resources Policy, 2023. 83: p. 103614.

[4]. An, J., Oil price predictors: Machine learning approach. 2019.

[5]. Arun Kumar, K., et al., Comparative analysis of Gated Recurrent Units (GRU), long Short-Term memory (LSTM) cells, autoregressive Integrated moving average (ARIMA), seasonal autoregressive Integrated moving average (SARIMA) for forecasting COVID-19 trends. Alexandria engineering journal, 2022. 61(10): p. 7585-7603.

[6]. Kim, G.I. and B. Jang, Petroleum Price Prediction with CNN-LSTM and CNN-GRU Using Skip-Connection. Mathematics, 2023. 11(3): p. 547.

[7]. Shu-rong, L. and G. Yu-lei. Crude oil price prediction based on a dynamic correcting support vector regression machine. in Abstract and applied analysis. 2013. Hindawi.

[8]. Rosli, N., R. Ibrahim, and I. Ismail, Intelligent prediction system for gas metering system using particle swarm optimization in training neural network. Procedia Computer Science, 2017. 105: p. 165-169.

[9]. Gao, S. and Y. Lei, A new approach for crude oil price prediction based on stream learning. Geoscience Frontiers, 2017. 8(1): p. 183-187.

[10]. Gumus, M. and M.S. Kiran. Crude oil price forecasting using XGBoost. in 2017 International conference on computer science and engineering (UBMK). 2017. IEEE.

[11]. Huang, L. and J. Wang, Global crude oil price prediction and synchronization based accuracy evaluation using random wavelet neural network. Energy, 2018. 151: p. 875-888.

[12]. Folorunso, S., A. Taiwo, and O. Olabanjo, A predictive model for estimating petroleum consumption using machine learning approach. Lautech journal of engineering and technology, 2018. 12(2): p. 80-87.

[13]. Bristone, M., R. Prasad, and A.A. Abubakar, CPPCNDL: Crude oil price prediction using complex

network and deep learning algorithms. Petroleum, 2020. 6(4): p. 353-361.
[14]. Cen, Z. and J. Wang, Crude oil price prediction model with long short term memory deep learning based on prior knowledge data transfer. Energy, 2019. 169: p. 160-171.
[15]. An, J., A. Mikhaylov, and N. Moiseev, Oil price predictors: Machine learning approach. International Journal of Energy Economics and Policy, 2019. 9(5): p. 1.
[16]. Shambulingappa, H., Crude oil price forecasting using machine learning. International Journal of Advanced Scientific Innovation, 2020. 1(1): p. 1-11.
[17]. Zhang, T., et al., Multi-step-ahead crude oil price forecasting based on two-layer decomposition technique and extreme learning machine optimized by the particle swarm optimization algorithm. Energy, 2021. 229: p. 120797.
[18]. Assaad, R.H. and S. Fayek, Predicting the price of crude oil and its fluctuations using computational econometrics: deep learning, LSTM, and convolutional neural networks. Econometric Research in Finance, 2021. 6(2): p. 119-137.
[19]. Busari, G.A. and D.H. Lim, Crude oil price prediction: A comparison between AdaBoost-LSTM and AdaBoost-GRU for improving forecasting performance. Computers & Chemical Engineering, 2021. 155: p. 107513.
[20]. Taiwo, A., Time Series Analysis of Nigerian Monthly Crude Oil Price. FUPRE Journal of Scientific and Industrial Research (FJSIR), 2021. 5(1): p. 1-12.
[21]. Bartholomew, D.C., et al., Modeling the Nigerian Bonny light crude oil price: the power of fuzzy time series. Open Journal of Modelling and Simulation, 2021. 9(4): p. 370-3900.
[22]. Paul, C., et al., The optimal machine learning modeling of Brent crude oil price. Quarterly Journal of Econometrics Research, 2021. 7(1): p. 31-43.
[23]. Jovanovic, L., et al., Multi-step crude oil price prediction based on lstm approach tuned by salp swarm algorithm with disputation operator. Sustainability, 2022. 14(21): p. 14616.
[24]. Jahanshahi, H., et al., Artificial intelligence-based prediction of crude oil prices using multiple features under the effect of Russia–Ukraine war and COVID-19 pandemic. Mathematics, 2022. 10(22): p. 4361.
[25]. Ogundunmade, T. and A. Adepoju, Modelling Liquefied Petroleum Gas Prices in Nigeria Using Time Series Machine Learning Models. Mod Econ Manag, 2022. 1(5): p. 3.
[26]. OGUNDUNMADE, T.A., A. Adepoju, and A. Allam, PREDICTING CRUDE OIL PRICE IN NIGERIA WITH MACHINE LEARNING MODELS. 2022: LAP LAMBERT ACADEMIC PUBL.