

# Topology Approach for Crude Oil Price Forecasting of Particle Swarm Optimization and Long Short-Term Memory

Marina Yusoff<sup>1</sup>, Darul Ehsan<sup>2</sup>, Muhammad Yusof Sharif<sup>3</sup>, Mohamad Taufik Mohd Sallehud-din<sup>4</sup>

Institute for Big Data Analytics and Artificial Intelligence (IBDAAD), Kompleks Al-Khawarizmi,  
Universiti Teknologi MARA (UiTM), 40450 Shah Alam, Selangor<sup>1</sup>

College of Computing, Informatics and Mathematics, Kompleks Al-Khawarizmi,  
Universiti Teknologi MARA (UiTM), 40450 Shah Alam, Selangor<sup>2</sup>

College of Computing, Informatics and Mathematics, Universiti Teknologi MARA Shah Alam, Malaysia<sup>3</sup>  
PETRONAS Research Sdn Bhd, Jln Ayer Hitam, Kawasan Institusi Bangi, 43000 Bandar Baru Bangi, Selangor<sup>4</sup>

**Abstract**—Forecasting crude oil prices hold significant importance in finance, energy, and economics, given its extensive impact on worldwide markets and socio-economic equilibrium. Using Long Short-Term Memory (LSTM) neural networks has exhibited noteworthy achievements in time series forecasting, specifically in predicting crude oil prices. Nevertheless, LSTM models frequently depend on the manual adjustment of hyperparameters, a task that can be laborious and demanding. This study presents a novel methodology incorporating Particle Swarm Optimization (PSO) into LSTM networks to optimize the network architecture and minimize the error. This study employs historical data on crude oil prices to explore and identify optimal hyperparameters autonomously and embedded with the star and ring topology of PSO to address the local and global search capabilities. The findings demonstrate that LSTM+starPSO is superior to LSTM+ringPSO, previous hybrid LSTM-PSO, conventional LSTM networks, and statistical time series methods in its predictive accuracy. LSTM+starPSO model offers a better RMSE of about +0.16% and +22.82% for WTI and BRENT datasets, respectively. The results indicate that the LSTM model, when enhanced with PSO, demonstrates a better proficiency in capturing the patterns and inherent dynamics data changes of crude oil prices. The proposed model offers a dual benefit by alleviating the need for manual hyperparameter tuning and serving as a valuable resource for stakeholders in the energy and financial industries interested in obtaining dependable insights into fluctuations in crude oil prices.

**Keywords**—Crude oil; deep learning; Particle Swarm Optimization; Long Term-Short Memory; forecasting

## I. INTRODUCTION

As one of the most significant commodities in the world, crude oil is responsible for the energy consumption. It is the foundation for daily items, from plastics to transportation fuels. Considering that fluctuations in crude oil prices significantly influence economies worldwide, price forecasting can help reduce the risks of oil price volatility [1]. Predictive methods in oil and gas operations can boost efficiency, lower costs, and reduce environmental impact from a good forecasting model [2]. Machine learning researchers and developers face challenges when working with large datasets and diverse data types, primarily because of noisy and unclean data [3]. Several pre-processing methods have been

developed to address this issue, with specific methods yielding favourable outcomes. Hence, the choice of pre-processing techniques would depend upon the data's characteristics and quality. Typically, benchmark datasets such as ready data do not necessitate extensive pre-processing tasks [4]. However, the most significant challenge is effectively managing substantial quantities, especially the time series data, which requires the development of a more understandable model. The abovementioned difficulties are relevant to the oil and gas data, especially in real-time data monitoring. Further investigation is required to effectively tackle the obstacles and determine the practicability of these methodologies using benchmark and real-life data.

Recently, there has been an increasing preference for incorporating predictive analytics in the oil and gas sector. Machine learning methods and their diverse applications in the oil and gas sector encompass multiple areas, such as pipeline prediction [5], well-log formation [6], and crude oil price forecasting [7].

Due to its effectiveness in predictive analytics, the Long Short-Term Memory (LSTM) model is widely utilized in various oil and gas industry sub-fields and other engineering and finance-related disciplines. Prior studies have examined various methodologies on LSTM, including LSTM with optimization and CNN with LSTM. For instance, CNN and LSTM address two instances of degradation prediction in offshore operation platforms for natural gas treatment plants and seawater injection pumps for oil [8]. Compared to a single LSTM model, the performance of the CNN with the LSTM model is superior, exhibiting a notable enhancement of 15.5% in precision.

Furthermore, the performance of LSTM has also been documented in reputable studies on time series [9], [10]. Yang et al. [9] employed the LSTM model to forecast short-and long-term production events in shale gas wells. The method performed better than the ARIMA, Arps, and Duong methods. In another study, Song et al. [11] used LSTM, a feature extraction and optimization model that incorporates feature engineering and parameter optimization and exhibits the lowest mean absolute error (MAE) value compared to other

models such as BPNN, LSTM, and random forest as reported by Dyer et al. [12]

One of the recent challenges in oil and gas is predicting crude oil prices. The demand for crude oil price prediction has increased due to crude oil's complex and highly unpredictable characteristics of crude oil [13]. Several methods were proposed and evaluated using benchmark datasets. For instance, LSTM and Henry gas solubility optimization (CHGSO) technique to estimate crude oil prices using West Texas Intermediate (WTI) and Brent Crude Oil Time Series COTS datasets [14] and Hybrid Wavelet Transform (WT) Bidirectional Long Short-Term Memory Network (BiLSTM)-Attention-CNN. WT-BLAC performs well for WTI data with R2, RMSE, MAPE, and MAE of 0.97, 2.25, 1.18, and 2.63, respectively. Furthermore, it evaluated a similar dataset using ensemble and ANN but with a different range of time series data [15]. The models have acceptable and significant interpretability in time series prediction of crude oil futures prices.

Support vector regression model did another forecasting crude oil prices, which are infamous for being unpredictable and have been fine-tuned using a genetic algorithm [2]. They use a ten-year daily dataset from NASDAQ and key economic input features. A study by Shahbazbegian et al. [10] divided time series into sub-series by the proposed hybrid model, which employs a multifaceted approach to capture distinct characteristics, LSTM is combined with the Markov switching model to forecast volatile and fluctuating sub-series.

After integrating these predictions using a linear combination, a comprehensive estimation of the time series for WTI crude oil prices is generated. The proposed method's respective RMSE and MAPE values are 4.18 and 0.03. Furthermore, He et al. [16] found that a hybrid forecasting model based on multi-modal features for price trends and employing the variational mode decomposition algorithm, extraction of data features with multiple modes, and time series employment of analysis provides acceptable performance. More research on crude oil forecasting solutions is still in demand. LSTM and its variants have great potential to obtain better forecasting results by embedding an appropriate optimization method. This paper focuses on using LSTM embedded with one of the popular computational optimizations, PSO. PSO is chosen due to its ease of implementation, high precision, and fast convergence [17], [18]. The applied forecasting model for oil and gas power transformers obtained an acceptable solution with PSO [19], [20]. Hence, we improve the LSTM by integrating with PSO.

PSO star and ring topology, as well as a new particle representation that could improve the accuracy performance of crude oil forecasting, are embedded. The ring star topology and CHGSO\_LSTM method, LSTM, and statistical time series techniques on benchmark crude oil price data compared the experimental results. Benchmark crude oil price data setting is the same as tested for CHGSO\_LSTM by [14]. The rest of the paper follows the organization of the section as follows. Section II describes preliminaries on PSO and LSTM. The material and method for the proposed solution are in Section III. The computational results and discussion is

mentioned in Section IV and V respectively. Finally, Section VI concludes the paper.

## II. PRELIMINARIES

### A. Particle Swarm Optimization

In 1995, James Kennedy and Russell Eberhart introduced the PSO algorithm as a powerful population-based optimization technique [17], [18]. PSO has gained popularity among scientists and researchers due to its ease of implementation, high precision, and rapid convergence. The PSO algorithm is renowned for exploring and exploiting the search space effectively, rendering it suitable for various applications [17], [19]. PSO algorithm is a metaheuristic optimization technique that employs a population of particles to iteratively adjust their positions and velocities to find the best solution to a given problem. Before implementing PSO, particle representation must be designed carefully for the proper objective function [20]. The particle representation is an essential element of the PSO design for ensuring the algorithm's efficiency. Particle representation is a mechanism for encoding problem-solving solutions. Its ability to determine the properties of individual particles is used to map feature elements. By assigning an appropriate representation to each particle, PSO could facilitate efficient solutions [20], [21].

PSO can systematically investigate various regions within the search space while leveraging the search process to enhance and optimize a viable solution. The search strategies employed in the PSO algorithm are affected by the parameters, namely the acceleration constants ( $C_1$  and  $C_2$ ) and the inertia weight, as discussed by Shi and Eberhart [22]. Eq. (1) and Eq. (2) denote the velocity and position formulas adapted from the canonical PSO [18], [22].

$$V_{i(new)} = WV_i + C_1r_1(PBest_i - X_i) + C_2r_2(Pbest_i - X_i) \quad (1)$$

$$X_{i(new)} = X_i + V_{i(new)} \quad (2)$$

where,

$V_{i(new)}$  = new velocity of the  $i^{th}$  particle

$V_i$  = current velocity of the  $i^{th}$  particle

$X_i$  = current position of the  $i^{th}$  particle

$X_{i(new)}$  = new position of the  $i^{th}$  particle

$W$  = inertia weight

$C_1$  and  $C_2$  = acceleration coefficient

$r_1$  and  $r_2$  = random function in the range of [ 0, 1]

$PBest_i$  = position of the personal best of the  $i^{th}$  particle

$Gbest_i$  = position of the global best derived from all particles in the swarm

The cognitive component, represented as  $C_1r_1(PBest_i - X_i)$  encompasses various factors such as the acceleration coefficient, a random function, and the difference between the personal best position,  $PBest_i$ , and the current position,  $X_i$ .

The difference between the previous position,  $X_i$  and the personal best position  $PBest_i$  can be observed. The social component,  $C_2 r_2(Gbest_i - X_i)$ , incorporates the acceleration coefficient, a random function, and the disparity between the previous position,  $X_i$ , and the global best position,  $Gbest_i$ . The specific component signifies the historical performance of the particle, which is obtained from the combined version of all particles.

Another PSO strategy is topology. Topology is another PSO technique for exploiting and exploration. Topology controls how particles interact and exchange information, enabling them to jointly explore and seek the best outcome, such as the star, ring, and square. It establishes how information is shared and how interactions take place among particles to find the best solution.

### B. Long Term-Short Memory

LSTM can effectively capture long-range dependencies within sequential data [23]. This characteristic renders it highly appropriate for natural language processing, speech recognition, and time series analysis. Due to its feedback connections and capacity to acquire knowledge of long-time features from time series data, the LSTM network demonstrates significant efficacy in processing and predicting sequential data. The ability to capture extensive dependencies in sequential data is a considerable advantage of LSTM in deep learning [23].

LSTM is a deep learning technique demonstrating remarkable efficiency in capturing extensive dependencies within sequential data. LSTM is accomplished by employing memory cells alongside various gating mechanisms, including input, forget, and output gates. The architecture of the LSTM is depicted in Fig. 1. These mechanisms enable the extended short-term memory network to preserve and strategically discard information over duration, thereby facilitating its ability to accurately capture the interdependencies inherent in the dataset.

A neural network model's performance with dense and LSTM units depends on several variables, including the problem's complexity. Increasing the number of dense units in a neural network can enhance the model's ability to discover the patterns and relationships in the data when dealing with complex problems. Similarly, increasing the number of LSTM units would improve its ability to capture long-term dependencies and remember previous information over time. This is especially important in tasks involving sequential data, such as time series analysis.

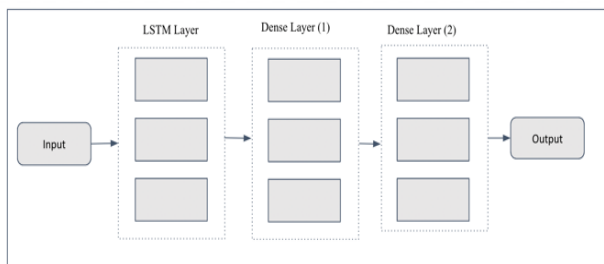


Fig. 1. LSTM architecture.

Furthermore, dense units can transform input data into a higher-dimensional space, increasing the model's ability to separate and classify. However, it is crucial to exercise caution when selecting the appropriate number of dense and LSTM units to prevent the data's overfitting or underfitting. Furthermore, the structure and architecture of the neural network model can influence the effectiveness of dense and LSTM units. Considering the specific problem, it is recommended to experiment with different combinations of dense and LSTM units to find the optimal configuration that yields the best performance and generalization on the given task. The selection and number of dense and LSTM units during the construction of neural network models can impact the performance of the models [24] [25]. The configuration and selection of dense and LSTM units in a neural network model can significantly impact its performance.

## III. MATERIALS AND METHODS

This section elaborates on the description of the materials, data sources, and research methodologies used. The proposed approach captures the steps to see the performance of enhancement of LSTM with PSO models on crude oil forecasting. The approach includes data acquisition, pre-processing, construction of the proposed methods, and evaluation. We propose two variants of the LSTM+PSO model, including the LSTM+starPSO and LSTM+ringPSO models, and compare them with ARIMA, SARIMAX, and LSTM. Fig. 2 demonstrates the overview of the proposed methodology. In addition, we introduced a particle representation or solution mapping for the PSO. Detailed steps are elaborated in the following sub-sections.

### A. Data Acquisition

This study uses two different datasets: the WTI Crude Oil dataset [26] and the BRENT Crude Oil dataset [27]. Brent Crude refers to the assemblage of oil extracted from the North Sea's seabed, whereas WTI Crude denotes the amalgamation of oil obtained from land in the United States. WTI and BRENT are widely recognized benchmarks in the oil and gas industry. Specifically, the price of BRENT oil is commonly utilized as a reference point for the light oil market in Africa, Europe, and the Middle East. The datasets used for this study were obtained from the FRED website, a publicly accessible economic data repository owned by the Federal Reserve Bank of St. Louis. The website provides daily frequency data on WTI and BRENT crude oil prices from the early 1990s.

Nevertheless, the scope of this study is limited to the utilization of data solely from the period spanning from January 4, 2000, to April 15, 2021. The WTI dataset comprises 5409 objects, while the BRENT dataset contains 5438 objects. Both datasets share the same features: date, price, open, high, low, volume, and percentage change. A recent finding shows that the same dataset was used as a main part of a study by Altan and Karasu [28], which used two different forecasting methods, PSO+LSTM and CHGSO+LSTM. It was reported that the CHGSO+LSTM approach performed better than the LSTM method. The dataset's narrative was interestingly explored by using other LSTM variants.

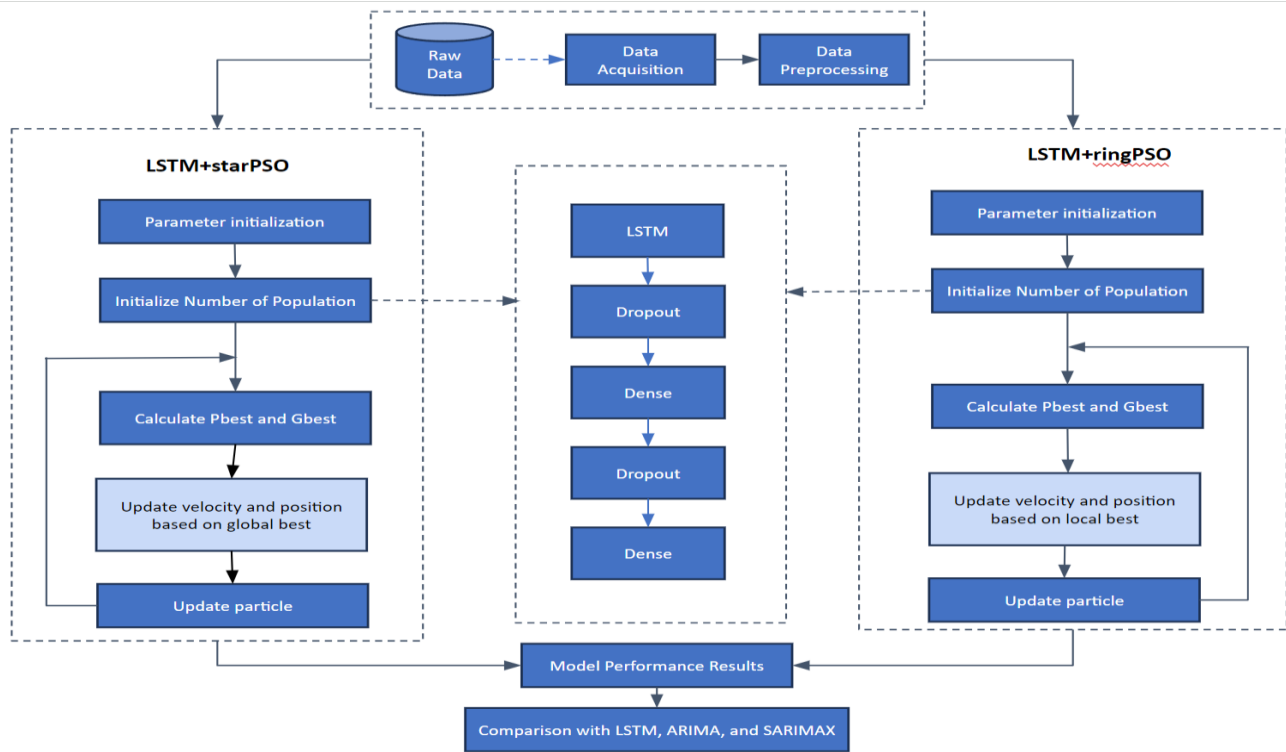


Fig. 2. Flow of methodology.

### B. Data Cleaning

From the data behavior perspective, the two datasets, WTI and BRENT, exhibit no missing values. Therefore, no missing value procedure is imposed. However, the identification of outliers is required. The interquartile range (IQR) method can detect outliers by utilizing the interquartile range Data distribution should be determined in the interquartile range within Q1 and Q3 or between the 25th and 75th percentiles. An outlier is any data point that lies outside a predefined range, typically defined as below the 25th percentile and above the 75th percentile by about 1.5 times. This careful method of outlier discovery and eradication improves the overall data quality and ensures that the dataset is used for subsequent forecasting. Eq. (1) is the IQR formula [29].

$$IQR = Q3 - Q1 \quad (3)$$

where, Q1 is the first quartile, and Q3 is the third quartile.

### C. Proposed Method

This section explains an enhancement method of LSTM with PSO in forecasting crude oil. The initial part of the method construction is the identification of particle representation. A new particle representation is proposed to adhere to the LSTM architecture and its parameters. The aim is to find the best position of the particle that can give an optimal or near-optimal solution. It is represented by the particle's item, namely, lookback, LSTM unit, Dense Unit, and learning rate. The representation consists of discrete and continuous values shown in Fig. 3.

Lookback	LSTM Unit	Dense Unit 1	Dense Unit 2	Learning Rate
[3 -10]	[64 - 256]	[10 -100]	[10 -100]	[0.01 - 0.001]

Fig. 3. Particle representation.

The LSTM-PSO algorithm incorporates two distinct topologies: the star and the ring. Consequently, two hybrid methodologies are proposed, specifically LSTM+starPSO and LSTM+ringPSO. Algorithm 1 outlines the procedural steps involved in the implementation of LSTM+starPSO. The LSTM+ringPSO follow similar steps, except Step 9. The algorithm commences by initializing the population of particles or swarm size. It is followed by initializing various parameters, including the number lookback, LSTM unit, dropout, dense unit, and learning rate. The Step 4 involves initializing the inertia weight,  $w$ , and acceleration constants  $C_1$  and  $C_2$ . Steps 5 and 6 involve the initialization of the minimum value of velocity (Vmin), the maximum value of velocity (Vmax), the minimum position (Pmin), and the maximum value of position (Pmax). The subsequent step involves determining the setting for formulating the objective function and the iteration number, denoted as  $i$ . The ninth step involves the uploading of input data. Step 10 involves the implementation of the LSTM+starPSO algorithm, while Step 11 entails the computation of the Pbest and Gbest values for each particle. The updated characteristics of the particle are outlined in Step 14.

---

**Algorithm 1: LSTM+starPSO**

---

```
1 Begin
2 Initialize the number of particles
3 Initiate number lookback, LSTM unit, dropout, dense unit,
  learning rate
4 Declare C1, C2 and W
5 Initialize Vmin, and Vmax
6 Initialize Pmax and Pmin.
7 Set the objective function based on RMSE
8 Set iteration number, i
9 Load datasets
10 Execute LSTM+starPSO
11 Calculate Pbest and Gbest value for each particle
12 Do
13 For each particle
14     Update features of the particle
15     Calculate the new velocity value,  $V_{(new)}$ 
16     Calculate new position,  $D_{(new)}$ 
17     Calculate Pbest(new)
18     Calculate Gbest(new)
20 While (stopping condition is reached)
21 End
```

---

The new velocity value for each particle is determined in Step 15 by applying Eq. (1). Eq. (2) is utilized to update the new position, denoted as P(new), in Step 16. Ultimately, Pbest(new) and Gbest(new) values are established by considering the fitness value assigned to the given problem. The iteration process commences at Step 12 and continues until Step 20, during which each particle's current velocity and position are updated. The iteration will continue until it meets the specified stopping condition.

#### D. Performance Measure

This study has used two essential empirical measurements to evaluate and compare the effectiveness of the LSTM and PSO+LSTM models. The performance metrics, Mean Absolute Percentage Error (MAPE) and Root Mean Squared Error (RMSE) are the cornerstone indicators used to evaluate the precision and dependability of these models. RMSE is a well-known statistical metric that expresses the variance between the predicted values produced by the models and the actual observed values. A lower RMSE value signifies greater accuracy and precision, as the model's predictions closely match the observed data.

Second, by evaluating the relative error as a percentage of the actual values, MAPE provides an insightful perspective on the performance of the models. MAPE averages out the absolute percentage differences between the predicted and actual values. This metric is beneficial for Assessing how well the models can predict values roughly equivalent to the actual data points and approximately proportional to them. A lower MAPE indicates that the models make more accurate predictions with minor relative errors in applications.

## IV. COMPUTATIONAL RESULTS

### A. Parameter Setting

The proposed method encompasses two distinct parameter setting categories: Particle Swarm Optimization (PSO) and Long Short-Term Memory (LSTM). In the Particle Swarm Optimization (PSO) context, a population size for initializing particles is selected from a set of values, namely {10, 20, 30}. The value of the iteration variable, denoted as  $i$ , is adjusted to a value of 30. The importance of C1 and C2 remains consistently equal to 2. The lower and upper bounds for the inertia weight are 0.4 and 0.9, respectively. The particle representation set by PSO mapping determines the selection of random values for parameters such as lookback, LSTM unit, dense unit, and learning rate in the context of LSTM. The values are selected randomly during the execution of the program. The lookback parameter is randomly selected from 3 to 10, while the LSTM unit is chosen from 64 to 256. The density unit and learning rate values are specified within the range of [10, 100] and [0.01, 0.01], respectively.

### B. Computational Results using LSTM Based on the Number of Lookback

In assessing the impact of lookback on LSTM performance, we conducted experiments using varying lookback values, ranging from 3 to 10, as indicated in Table I. The objective is to determine an appropriate value for the lookback parameter and identify the optimal RMSE and MAPE values. As shown in Table I, the utilization of WTI in LSTM models yields diverse RMSE and MAPE values, indicating variations in performance. Two lookbacks, specifically 5 and 8, stand out due to their comparable RMSE and MAPE results. The root mean square error (RMSE) for Lookback 5 is calculated to be 2.6604, with a MAPE of 4.6256.

On the other hand, Lookback 8 has an RMSE of 2.7761 and a MAPE of 4.2628. Based on the analysis, it is observed that the MAPE of the lookback 8 model is significantly lower than that of the lookback 5 model, with a difference of -0.35. Additionally, the model with RMSE of lookback value equal to 8 is slightly higher than that of lookback equal to 5, with a difference of +0.11. Consequently, the model for lookback equal to 8 is deemed optimal for the WTI dataset. The results obtained on the BRENT dataset indicate a different outcome, highlighting the prominence of a particular lookback value of 7.

### C. Comparison Results of Different Methods

The forecasting results on the datasets WTI and BRENT are summarized in Table II and Table III. We incorporate LSTM with starPSO and ringPSO and tabulate the results from the recent finding by [28] and the conventional LSTM and two statistical time series methods, ARIMA and SARIMAX. The best results are highlighted in bold-face type. LSTM+starPSO provides better performance for the two datasets. In Table II, LSTM+starPSO with LSTM units of 212, dense unit of 77, and learning rate of 0.0083, lookback equals to 8 demonstrates the superior performance compared to other methods with RMSE of about 1.7512. We can see from Table III on the BRENT dataset a better forecasting

performance offered by LSTM+starPSO, where both RMSE and MAPE are minimized. The performance seems better than CHGSO\_LSTM, which reported 1.7540 for WTI and 0.8453 of the RMSE for BRENT. In terms of the number of population, it shows that 20 is acceptable for both datasets. According to the results, although the data is univariate, the statistical models ARIMA and SARIMAX are ineffective compared to LSTM and its variants in forecasting future oil prices.

Every PSO requires an objective function or criterion that the PSO seeks to optimize. In this case, RMSE from the LSTM result is used as the objective function, which PSO tries to minimize the RMSE value. According to our findings, the ideal lookback range and learning rate are [6, 7] and 0.008,

respectively. LSTM+starPSO performs relatively similarly to the CHGSO-LSTM model [28] for the WTI and BRENT datasets in RMSE and MAPE. For the WTI dataset, +0.16% and -8.56% are obtained for RMSE and MAPE. A similar trend can be seen with the BRENT dataset, with +22.82% and +18.7% in RMSE and MAPE. On the other hand, LSTM+ringPSO performs slightly lower than LSTM+starPSO, where the result is -3.63% and -13.74% for RMSE and MPE for WTI and +20.7% and +21% for RMSE and MAPE for BRENT. PSO uses the position and velocity update method to find the best RMSE value. LSTM+starPSO outperforms CHGSO-LSTM with RMSE and MAPE with 11% and 5.7% reduction. The same goes for the BRENT dataset, where we see a 39% and 42.8% reduction in RMSE and MAPE, respectively.

TABLE I. COMPUTATIONAL RESULTS USING CONVENTIONAL LSTM BASED ON THE NUMBER OF LOOKBACK

Dataset		WTI		BRENT	
Split	Lookback	RMSE	MAPE	RMSE	MAPE
60:40	3	<b>2.1566</b>	<b>3.4105</b>	<b>2.4901</b>	<b>2.6015</b>
	4	3.3722	4.7473	8.1081	7.1307
	5	4.2722	6.8263	7.2351	6.2737
	6	4.2524	6.3045	3.1759	3.0359
	7	10.7975	17.5782	5.2858	6.1048
	8	5.5147	8.2872	6.1838	7.7876
	9	5.2320	9.1245	4.1238	4.2867
	10	4.7698	5.5458	4.9142	4.9342
70:30	3	5.6011	10.4097	3.9737	5.3860
	4	4.1777	7.7009	3.3143	3.7884
	5	4.5288	7.8300	<b>2.7562</b>	<b>3.2520</b>
	6	5.6229	10.4402	3.1277	3.7722
	7	<b>1.9410</b>	<b>3.4709</b>	3.9283	4.7050
	8	3.2464	6.2048	5.4484	6.5811
	9	4.6569	8.9385	4.2039	5.3158
	10	4.3710	7.5679	2.7731	3.5558
85:15	3	2.6810	5.0949	3.0697	4.0615
	4	6.0544	10.7058	3.5370	4.9132
	5	2.6604	4.6256	3.2284	4.9966
	6	2.7652	5.0243	4.8779	7.9213
	7	3.2381	5.7339	<b>2.1756</b>	<b>2.9873</b>
	8	<b>2.7761</b>	<b>4.2628</b>	4.0058	6.2856
	9	3.3066	6.1230	2.7391	4.0462
	10	3.8408	7.4215	2.6276	3.9131
90:10	3	2.4417	4.8107	3.2680	3.7026
	4	5.1032	9.8931	2.5472	3.2090
	5	4.7821	8.8877	<b>1.8540</b>	<b>2.1971</b>
	6	<b>2.1066</b>	<b>4.2554</b>	4.6973	6.7682
	7	3.0119	6.0918	2.5806	3.2530
	8	3.4707	7.1888	3.5000	4.8287
	9	4.4182	8.6690	2.0140	2.5106
	10	4.0567	8.1125	2.3070	2.8610



TABLE II. CRUDE OIL FORECASTING MODEL PERFORMANCE

Performance Measure	CHGSO_LSTM [28]	LSTM+starPSO (Proposed method)			LSTM+ringPSO			LSTM	ARIMA	SARIMAX
		20	10	30	20	10	30			
Population	20	20	10	30	20	10	30	-	-	-
RMSE	1.7540	<b>1.7512</b>	2.0601	1.9486	1.8199	2.0962	1.9889	1.9410	12.8125	15.0574
MAPE	2.7570	<b>3.0150</b>	3.5933	3.3466	3.1964	3.7775	3.4424	3.4709	24.8355	29.5014
Learning Rate	0.01	<b>0.0083</b>	0.0529	0.0076	0.0062	0.0470	0.0043	0.01	0.01	0.01
LSTM units	200	<b>212</b>	70	145	183	207	154	200	200	200
Dense Units	50	<b>77</b>	12	84	59	59	71	50	50	50
Lookback	[3, 10]	<b>8</b>	7	7	6	4	5	7	-	-

TABLE III. BRENT CRUDE OIL FORECASTING MODEL PERFORMANCE

Performance Measure	CHGSO_LSTM [28]	LSTM+star_PSO			LSTM+ringPSO			LSTM	ARIMA	SARIMAX
		20	10	30	20	10	30			
Swarm Size	20	20	10	30	20	10	30	-	-	-
RMSE	1.8453	<b>1.5024</b>	2.0601	2.8145	<b>1.5283</b>	1.8231	<b>1.7265</b>	1.8540	14.9476	16.2943
MAPE	2.4525	<b>2.0660</b>	<b>1.9954</b>	3.6217	<b>2.0269</b>	<b>2.0359</b>	2.2522	2.1971	25.1301	27.4191
Learning Rate	0.01	0.0066	0.0083	0.0011	0.0054	0.0051	0.0057	0.01	0.01	0.01
LSTM Unit	200	183	106	219	185	209	224	200	-	-
Dense Unit	50	51	60	38	65	27	14	50	-	-
Lookback	[3, 10]	10	3	5	7	7	7	5	-	-

## V. DISCUSSION

### A. Effect of LSTM and Dense Units

In LSTM, dense units can facilitate the transformation of the input data into a higher-dimensional space, increasing the model's ability to separate and classify. On the other hand, the number of dense and LSTM units should be chosen carefully to avoid overfitting or underfitting the data. Therefore, network architecture design, including the composition of dense and LSTM units, is significantly important [30]. It is advisable to experiment with different combinations of dense and LSTM units to find the optimal architecture that yields the best performance of forecasting accuracy. This paper explores using stochastic particle features to determine the most suitable number of dense and LSTM units and the incorporation of dense layers within an LSTM network. Interestingly, the LSTM+starPSO model, which used 212 LSTM units and 77 for dense units on WTI datasets, showed how sufficient dense and LSTM units reduce the overfitting problem and increase forecasting accuracy. In the context of the BRENT dataset, the model architecture consisted of 183 LSTM units and 51 dense units.

### B. Effect of Lookback

Lookback in time series forecasting establishes how much historical data the LSTM, LSTM+starPSO, and LSTM+ringPSO models should consider when making predictions for the following time step. Depending on several variables, including the data patterns, the effect of lookback on LSTM can be significant. Determining the most effective lookback period is contingent upon the unique attributes of the time series data [31]. More than eight lookback numbers appear required for accurate prediction in the best

performance models, which offer little but longer-term dependencies. With such a small number of lookbacks, more is needed.

### C. Effect of Learning Rate

The role of the learning rate in the LSTM forecasting model is to assist in effective converging and achieving good performance. We represent the particle with the range of learning rate of [0.01, 0.001]. It is randomly chosen within these ranges during the LSTM+starPSO and LSTM+ringPSO execution. The small learning rate value is randomly chosen at about 0.006 to 0.008 for both datasets using LSTM+starPSO, meanwhile about 0.004 to 0.005 when using LSTM+ringPSO. However, the use of a small learning rate value has a significant effect on the forecasting results. It is evident that the choice of a small learning rate in LSTM models for time series forecasting aims to achieve a better convergence and generalization in the exploitation and exploration of the search space [32].

### D. Effect of PSO Topology in LSTM

Ring topology is a local-based focal point of particles. It attracts particles to the best particle in its corresponding neighborhood. In our experiment, three swarm sizes are used: 10, 20, 30. For instance, each particle has 29 neighborhoods when using a swarm size equal to 30. However, due to the various particle positions in the search space, the nearest particle of each particle considers the local neighborhood involved in the local search. Each particle's local surroundings consist of a fixed number of other particles. It differs from star topology, where all particles within the swarm share information with and are influenced by the particle with the highest performance [33]. Star topology promotes global exploration by encouraging particles to move towards the

optimal solution discovered by any swarm member. LSTM with star topology has demonstrated that each particle in the searching space is attracted to the best particle of the swarm. It obtained the best forecasting accuracy performance for WTI and BRENT datasets. The global searching by star topology [34] achieves the objective function of minimizing the RMSE value.

## VI. CONCLUSION

In this study, an enhancement of LSTM incorporated with PSO addresses the challenges of forecasting the daily time series crude oil price data. The proposed method comprises two main steps. At the PSO, particle mapping is designed together with topology to achieve a dynamic LSTM architecture and improve PSO searching capabilities for exploration and exploitation. With this method, more accurate forecasting is obtained. Experimental findings show that compared with the recent CHGSO\_LSTM, the suggested LSTM+starPSO offers the most performing methods. It is a better outcome compared to LSTM+ringPSO and conventional methods. However, more experimental work could be conducted by embedding ensembles and executing feature engineering strategy and hyperparameter tuning.

## ACKNOWLEDGMENT

The authors would like to acknowledge the Petroleum Nasional Berhad (PETRONAS) Academia Collaboration Grant, Institute for Big Data Analytics and Artificial Intelligence (IBDAAI), and Universiti Teknologi MARA for the financial support provided to this research project.

## REFERENCES

- [1] Z. Xu, M. Mohsin, K. Ullah, and X. Ma, "Using econometric and machine learning models to forecast crude oil prices: Insights from economic history," *Resources Policy*, vol. 83, p. 103614, 2023, doi: 10.1016/J.RESOURPOL.2023.103614.
- [2] S. Lipsa and R. K. Dash, "GASVR- A Model to Predict and Analyze Crude Oil Price," in *2022 2nd Asian Conference on Innovation in Technology, ASIANCON 2022*, 2022. doi: 10.1109/ASIANCON55314.2022.9908764.
- [3] H. Son and Y. Jang, "Partial convolutional LSTM for spatiotemporal prediction of incomplete data," *IEEE Access*, vol. 8, pp. 164762–164774, 2020, doi: 10.1109/ACCESS.2020.3022774.
- [4] P. Sohrabi, H. Dehghani, and R. Rafie, "Forecasting of WTI crude oil using combined ANN-Whale optimization algorithm," *Energy Sources, Part B: Economics, Planning and Policy*, vol. 17, no. 1, 2022, doi: 10.1080/15567249.2022.2083728.
- [5] M. E. A. Ben Seghier, D. Höche, and M. Zheludkevich, "Prediction of the internal corrosion rate for oil and gas pipeline: Implementation of ensemble learning techniques," *J Nat Gas Sci Eng*, vol. 99, p. 104425, 2022, doi: <https://doi.org/10.1016/j.jngse.2022.104425>.
- [6] M. Farsi et al., "Predicting Formation Pore-Pressure from Well-Log Data with Hybrid Machine-Learning Optimization Algorithms," *Natural Resources Research*, vol. 30, no. 5, pp. 3455–3481, 2021, doi: 10.1007/s11053-021-09852-2.
- [7] K. Tissaoui, T. Zaghdoudi, A. Hakimi, and M. Nsaibi, "Do Gas Price and Uncertainty Indices Forecast Crude Oil Prices? Fresh Evidence Through XGBoost Modeling," *Comput Econ*, vol. 62, no. 2, pp. 663–687, 2023, doi: 10.1007/s10614-022-10305-y.
- [8] X.-L. Pan, L.-J. Zhang, G. He, and P. Zhang, "Prediction model for residual strength of warship seawater pipelines with internal corrosion and test verification," *Chuan Bo Li Xue/Journal of Ship Mechanics*, vol. 25, no. 2, pp. 202–209, 2021, doi: 10.3969/j.issn.1007-7294.2021.02.008.
- [9] R. Yang, X. Liu, R. Yu, Z. Hu, and X. Duan, "Long short-term memory suggests a model for predicting shale gas production," *Appl Energy*, vol. 322, 2022, doi: 10.1016/j.apenergy.2022.119415.
- [10] V. Shahbazbegian, H. Hosseini-saz, M. Shafie-Khah, and M. Elmusrati, "Forecasting Crude Oil Prices using a Hybrid Model Combining Long Short-Term Memory Neural Networks and Markov Switching Model," in *2023 International Conference on Future Energy Solutions, FES 2023*, 2023. doi: 10.1109/FES57669.2023.10182444.
- [11] S. Song et al., "An intelligent data-driven model for virtual flow meters in oil and gas development," *Chemical Engineering Research and Design*, vol. 186, pp. 398–406, 2022, doi: 10.1016/j.cherd.2022.08.016.
- [12] A. S. Dyer et al., "Applied machine learning model comparison: Predicting offshore platform integrity with gradient boosting algorithms and neural networks," *Marine Structures*, vol. 83, 2022, doi: 10.1016/j.marstruc.2021.103152.
- [13] H. He, M. Sun, X. Li, and I. A. Mensah, "A novel crude oil price trend prediction method: Machine learning classification algorithm based on multi-modal data features," *Energy*, vol. 244, p. 122706, 2022, doi: 10.1016/J.ENERGY.2021.122706.
- [14] A. Altan and S. Karasu, "Crude oil time series prediction model based on LSTM network with chaotic Henry gas solubility optimization," *Energy*, vol. 242, p. 122964, 2022, doi: 10.1016/J.ENERGY.2021.122964.
- [15] H. Guliyev and E. Mustafayev, "Predicting the changes in the WTI crude oil price dynamics using machine learning models," *Resources Policy*, vol. 77, 2022, doi: 10.1016/j.resourpol.2022.102664.
- [16] H. He, M. Sun, X. Li, and I. A. Mensah, "A novel crude oil price trend prediction method: Machine learning classification algorithm based on multi-modal data features," *Energy*, vol. 244, 2022, doi: 10.1016/j.energy.2021.122706.
- [17] M. Yusoff, J. Ariffin, and A. Mohamed, "DPSO based on a min-max approach and clamping strategy for the evacuation vehicle assignment problem," *Neurocomputing*, vol. 148, pp. 30–38, 2015, doi: <https://doi.org/10.1016/j.neucom.2012.12.083>.
- [18] X. Hu, Y. Shi, and R. Eberhart, "Recent advances in particle swarm," in *Proceedings of the 2004 Congress on Evolutionary Computation, CEC2004*, 2004, pp. 90–97. [Online]. Available: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-4344682502&partnerID=40&md5=60528ec5f19a0f07655319c9c06b8e28>.
- [19] M. Riaz, A. Hanif, S. J. Hussain, M. I. Memon, M. U. Ali, and A. Zafar, "An optimization-based strategy for solving optimal power flow problems in a power system integrated with stochastic solar and wind power energy," *Applied Sciences (Switzerland)*, vol. 11, no. 15, 2021, doi: 10.3390/app11156883.
- [20] Y. Shi and R. Eberhart, "Monitoring of Particle Swarm Optimization," *Front Comput Sci China*, vol. 3, no. 1, pp. 31–37, 2009, doi: 10.1007/s11704-009-0008-4.
- [21] M. Yusoff, A. N. M. Basir, N. A. Kadir, and S. A. Bahari, "Evaluation of Particle Swarm Optimization for strength determination of tropical wood polymer composite," *IAES International Journal of Artificial Intelligence*, vol. 9, no. 2, pp. 364–370, 2020, doi: 10.11591/ijai.v9.i2.pp364-370.
- [22] Y. Shi and R. C. Eberhart, "Population diversity of particle swarms," in *2008 IEEE Congress on Evolutionary Computation, CEC 2008*, 2008, pp. 1063–1067. doi: 10.1109/CEC.2008.4630928.
- [23] X. Lei, "Stock Market Forecasting Method Based on LSTM Neural Network," in *2023 IEEE 3rd International Conference on Power, Electronics and Computer Applications, ICPECA 2023*, 2023, pp. 1534–1537. doi: 10.1109/ICPECA56706.2023.10076100.
- [24] K. Gajamannage, Y. Park, and D. I. Jayatilake, "Real-time forecasting of time series in financial markets using sequentially trained dual-LSTMs," *Expert Syst Appl*, vol. 223, 2023, doi: 10.1016/j.eswa.2023.119879.
- [25] M. S. Sawah, S. A. Taie, M. H. Ibrahim, and S. A. Hussein, "An accurate traffic flow prediction using long-short term memory and gated recurrent unit networks," *Bulletin of Electrical Engineering and Informatics*, vol. 12, no. 3, pp. 1806–1816, 2023, doi: 10.11591/eei.v12i3.5080.



- [26] "Crude Oil Price | WTI Price Chart - Investing.com." 2023. [Online]. Available: <https://www.investing.com/commodities/crude-oil>.
- [27] "Brent Crude Oil Price - Investing.com." 2023. [Online]. Available: <https://www.investing.com/commodities/brent-oil>.
- [28] A. Altan and S. Karasu, "Crude oil time series prediction model based on LSTM network with chaotic Henry gas solubility optimization," *Energy*, vol. 242, p. 122964, 2022, doi: 10.1016/J.ENERGY.2021.122964.
- [29] K. Fujioka and M. Deng, "Nonlinear temperature control of heat exchanger process using particle filter with interquartile range particle weights," *International Conference on Advanced Mechatronic Systems, ICAMechS*, vol. 2021-December, pp. 130–134, 2021, doi: 10.1109/ICAMECHS54019.2021.9661547.
- [30] A. A. Ismail and M. Yusoff, "An Efficient Hybrid LSTM-CNN and CNN-LSTM with GloVe for Text Multi-class Sentiment Classification in Gender Violence," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 9, 2022, doi: 10.14569/IJACSA.2022.0130999.
- [31] H. Jeon, J. Ryu, K. M. Kim, and J. An, "The Development of a Low-Cost Particulate Matter 2.5 Sensor Calibration Model in Daycare Centers Using Long Short-Term Memory Algorithms," *Atmosphere (Basel)*, vol. 14, no. 8, 2023, doi: 10.3390/atmos14081228.
- [32] C. Yu, X. Qi, H. Ma, X. He, C. Wang, and Y. Zhao, "LLR: Learning learning rates by LSTM for training neural networks," *Neurocomputing*, vol. 394, pp. 41–50, 2020, doi: <https://doi.org/10.1016/j.neucom.2020.01.106>.
- [33] N. Lynn, M. Z. Ali, and P. N. Suganthan, "Population topologies for particle swarm optimization and differential evolution," *Swarm Evol Comput*, vol. 39, pp. 24–35, 2018, doi: <https://doi.org/10.1016/j.swevo.2017.11.002>.
- [34] V. Mann, A. Sivaram, L. Das, and V. Venkatasubramanian, "Robust and efficient swarm communication topologies for hostile environments," *Swarm Evol Comput*, vol. 62, p. 100848, 2021, doi: <https://doi.org/10.1016/j.swevo.2021.100848>.