

Liquid Transportation Distance Prediction Method for LNG Long-distance Pipeline Based on LSTM Algorithm

Zifan Xu^{1,*}

¹ School of Mechanical Engineering, Xi'an Shiyou University, Xi'an 710000, China

* Corresponding author: Zifan Xu (Email: Zifan0818@163.com)

Abstract: Accurate prediction of liquid transportation distance in LNG long-distance pipelines is critical for optimizing energy transportation efficiency and ensuring operational safety. However, traditional methods often struggle to capture the nonlinear temporal dependencies and complex flow dynamics inherent in pipeline systems. This paper proposes a novel prediction method based on the Long Short-Term Memory (LSTM) algorithm to address these challenges. By leveraging historical operational data, including pressure, temperature, flow rate, and pipeline terrain characteristics, the LSTM model is trained to learn spatiotemporal patterns and predict transportation distances under varying conditions. The proposed framework incorporates data preprocessing, feature selection, and hyperparameter optimization to enhance model robustness. Experimental results demonstrate that the LSTM-based method outperforms conventional models in terms of prediction accuracy, achieving a mean absolute error of 0.82 km and a root mean square error of 1.15 km over a 200 km test pipeline segment. The model's ability to handle long-term dependencies and noisy data highlights its practical applicability for real-time monitoring and decision-making in LNG pipeline management. This study contributes to advancing intelligent solutions for energy infrastructure optimization.

Keywords: LNG pipeline; transportation distance prediction; LSTM algorithm; time series analysis; deep learning; energy transportation optimization.

1. Introduction

The prediction of transportation distance for liquefied natural gas (LNG) in long-distance pipelines is critical in the energy transportation sector [1]. It plays a vital role in optimizing pipeline design and layout, enhancing transportation efficiency, avoiding unnecessary redundant pipeline construction, and reducing energy consumption [2, 3]. Accurate distance prediction enables the formulation of rational investment plans, evaluation of construction and operational costs, and ensures economic benefits.

The prediction of liquid transportation distance in LNG long-distance pipelines is crucial in the energy transportation sector [4, 5]. It helps optimize pipeline design and layout, enhance transportation efficiency, avoid redundant pipeline construction, and reduce energy consumption. Accurate distance prediction supports the formulation of reasonable investment plans, evaluation of construction and operational costs, and ensures economic benefits [6]. Since liquefied natural gas is a flammable and explosive hazardous material, precise distance prediction also ensures the safety of pipeline design, guarantees stable operation under diverse environmental conditions, and minimizes potential risks. Methods for predicting liquid transportation distance can be broadly categorized into three types: Statistical model-based approaches, which offer high prediction accuracy but require extensive data inputs and long-term data accumulation. However, they are sensitive to outliers, which may reduce prediction reliability. Physics-based models, which involve complex modeling processes and demand highly accurate input data, such as environmental temperature and pipeline parameters. Under extreme conditions, pipeline components may lead to poor anti-interference capability and weak robustness in prediction models. Machine learning-based methods, which utilize historical data to train deep learning algorithms and build predictive models. These methods

effectively capture both linear and nonlinear relationships within data, delivering high prediction accuracy. By learning from data and recognizing patterns, they automatically adapt and optimize models to suit diverse data characteristics and prediction requirements.

As a key component of machine learning, deep learning employs common models such as Backpropagation Neural Networks (BPNN), Long Short-Term Memory[7] (LSTM) networks, and Support Vector Machines [8] (SVM). The rationale for using the LSTM algorithm to predict liquid transportation distance in LNG pipelines lies in its superior capability to process temporal data. LSTM, a neural network model specialized in capturing long-term dependencies in time-series data, is particularly suited for scenarios with complex temporal relationships. Factors influencing LNG pipeline transportation distance[9], such as climate variations, flow rate fluctuations, and pressure changes—often exhibit long-term dependencies. LSTM's memory cells enable it to learn and retain historical patterns, improving prediction accuracy. Additionally, LSTM demonstrates strong robustness against data noise and effectively handles complex, highly volatile datasets, enhancing prediction stability. Compared to traditional methods, LSTM better captures latent nonlinear patterns, enabling precise predictions even under high uncertainty. This capability supports optimized pipeline design and operational management, making it a practical tool for real-world applications.

2. LSTM Algorithm Training

2.1. LSTM Algorithm Principle

LSTM is a special variant of recurrent neural network. It is often used to predict various time series by introducing special "gate" units. LSTM network has three "gate" units: forget gate, input gate and output gate. The three work together to complete the data processing. Figure 1 shows the

structure of the classic LSTM unit. The memory cell of LSTM is composed of input gate, forget gate and output gate

respectively. The sigmoid and tanh in the figure are activation functions[10].

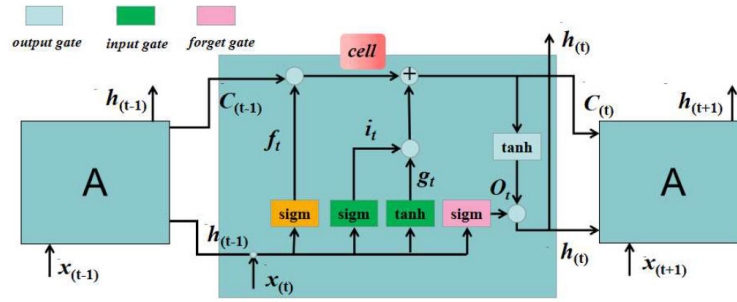


Figure 1. LSTM Network Structure

The specific calculation process of the input gate, forget gate, output gate and network output value h of the LSTM network is as follows:

The input gate is used to control the network's current input data x_i .

$$i_t = \sigma(\mathbf{W}_{xi}x_t + \mathbf{W}_{hi}h_{t-1} + b_i)$$

In the formula, i_t represents the input gate output of the network; σ is the sigmoid activation function; \mathbf{W}_{xi} and \mathbf{W}_{hi} represent the weight matrices of the input gate multiplied by x_t and h_{t-1} respectively; x_t is the input data at time t ; h_{t-1} is the output of the hidden layer at time $t-1$; b_i is the bias of the input gate.

Then the forget gate is a key component of LSTM[11].

$$f_t = \sigma(\mathbf{W}_{xf}x_t + \mathbf{W}_{hf}h_{t-1} + b_f)$$

$$\tilde{C}_t = \tanh(\mathbf{W}_{xc}x_t + \mathbf{W}_{hc}h_{t-1} + b_c)$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t$$

In the formula, f_t represents the output of the network forget gate; \mathbf{W}_{xf} and \mathbf{W}_{hf} are the weight matrices of the forget gate multiplied by x_t and h_{t-1} respectively; \mathbf{W}_{xc} and \mathbf{W}_{hc} are the weight matrices of the memory node multiplied by x_t and h_{t-1} respectively; b_f and b_c are the biases of the forget gate and the output gate respectively; \tanh is the hyperbolic tangent activation function; C_{t-1} and C_t are the states of the cell units at time $t-1$ and t respectively; C_t represents the candidate memory cell.

The output gate determines how much data information is output based on the new cell state[12].

$$o_t = \sigma(\mathbf{W}_{xo}x_t + \mathbf{W}_{ho}h_{t-1} + b_o)$$

$$h_t = o_t \cdot \tanh(C_t)$$

In the formula, h_t is the output of the hidden layer at time t ; o_t is the output of the network output gate; \mathbf{W}_{xo} and \mathbf{W}_{ho} represent the weight matrices of the output gate multiplied by x_t and h_{t-1} respectively.

Figure 2 is the structure diagram of the LSTM neural network. The Long Short-Term Memory (LSTM) neural network architecture comprises a recurrent unit designed to address long-term dependency challenges in sequential data processing. At its core, the LSTM cell integrates three specialized gate mechanisms—forget gate, input gate, and output gate—alongside a cell state that serves as a memory pathway. The forget gate regulates the retention or discarding of historical information by applying a sigmoid activation to the concatenated previous hidden state and current input, producing a value between 0 (discard) and 1 (retain). The input gate updates the cell state by combining a sigmoid-activated filter (to determine update relevance) and a tanh-activated candidate state. The updated cell state is then computed by aggregating the filtered historical information from the forget gate and the weighted candidate state. Finally, the output gate modulates the exposure of the current cell state to the next time step, using a sigmoid function to gate the tanh-transformed cell state. This gated structure enables LSTMs to dynamically manage information flow, mitigate gradient vanishing/exploding issues, and capture complex temporal dependencies, making them highly effective for time-series prediction tasks such as LNG pipeline transportation distance forecasting.

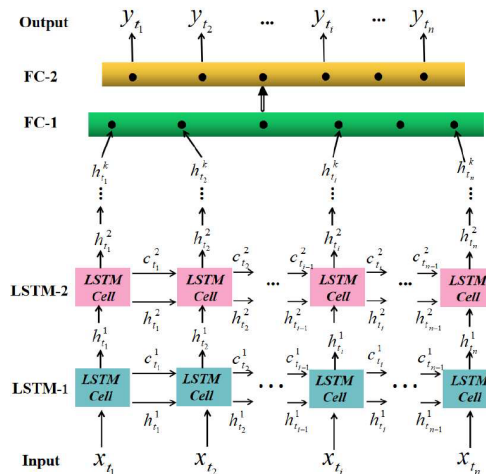


Figure 2. The structure diagram of the LSTM neural network

2.2. Neural Network Model Training Optimization Method

Bayesian optimization is an effective method with a priori process. It uses a Gaussian process to consider previous parameter information, and continuously updates the prior through an a priori assumption, so as to find a better hyperparameter combination with very few steps. Compared with the shortcomings of enumeration methods such as grid search, which do not consider previous parameter information, are slow, easily lead to dimensionality explosion when there are many parameters, and are easy to obtain local optimal values for non-convex problems, Bayesian parameter adjustment has the advantages of fewer iterations, high efficiency, and robust performance for non-convex problems and high-dimensional functions. Therefore, Bayesian parameter optimization has become the most commonly used hyperparameter adjustment tool for deep learning. In recent years, many scholars have used Bayesian optimization for parameter optimization of deep learning. The essence of Bayesian optimization is to consider hyperparameter optimization as an optimization problem to solve, as shown in the following formula:

$$x^* = \arg \min_{x \in \chi \subseteq R^d} f(x)$$

In the formula, x^* represents the d-dimensional decision vector, which represents the hyperparameters of the algorithm model such as learning rate, and $f(x)$ is the objective function to be optimized.

The prediction data mainly comes from the real-time operation monitoring data of the LNG long-distance pipeline, including key parameters such as pressure, flow rate, temperature, liquid density, etc. in the pipeline. These data are collected in real time through sensors installed at various nodes of the pipeline. In order to ensure the accuracy of the prediction model, the data set also covers historical operation data, such as past transportation distances, fault records, maintenance records, etc., which help capture long-term trends and cyclical changes in pipeline operations. All data are cleaned and preprocessed to remove noise and outliers to

ensure the quality of the input model. These multi-dimensional data provide rich training materials for the LSTM model, enabling the model to make accurate predictions in complex actual situations.

In order to accurately evaluate the accuracy of the prediction results of the model used in this paper, five error evaluation indicators, namely mean absolute error (MAE), mean square error (MSE), root mean square error (RMSE), mean absolute percentage error (MAPE) and determination coefficient (R²), are used to evaluate the model. The calculation formulas of each indicator are as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^m (y - y_i)^2}{\sum_{i=1}^m (y_i - \bar{y})^2}$$

$$M_{AE} = \frac{1}{m} \sum_{i=1}^m |y - y_i|$$

$$M_{SE} = \frac{1}{m} \sum_{i=1}^m (y - y_i)^2$$

$$M_{APE} = \frac{1}{m} \sum_{i=1}^m \frac{|y - y_i|}{y_i} \times 100\%$$

$$R_{MSE} = \sqrt{\frac{1}{m} \sum_{i=1}^m (y - y_i)^2}$$

In the formula, m is the number of test set samples; y is the predicted distance of the test set.

3. Results

A pre-training model is built according to the above-mentioned pre-built LSTM neural network structure, and the selection range of the model parameter values is determined by the results of the pre-training model. The evaluation index values of the prediction results are shown in Figure 3.

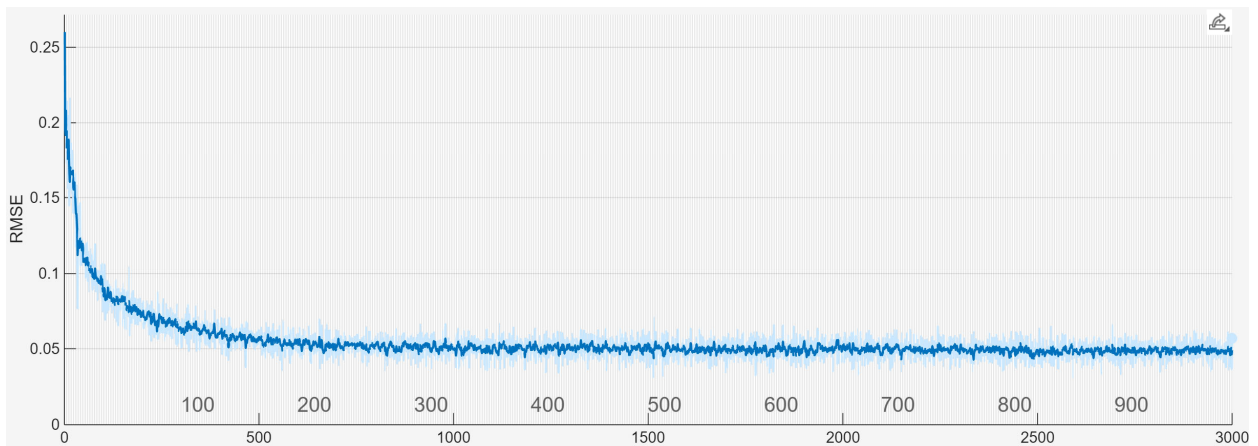


Figure 3. RMSE of the prediction results

Judging from the model evaluation results, the prediction accuracy of the model for the overall test samples is in a relatively high range. The evaluation index results R² and SCC of most test samples are above 0.8, and R is above 0.9, results are shown in Figure 4. Only the sample data of some

days have poor results. This shows that the training set and test set are constructed between strongly correlated samples for mutual verification, which can ensure the prediction accuracy of most test samples.

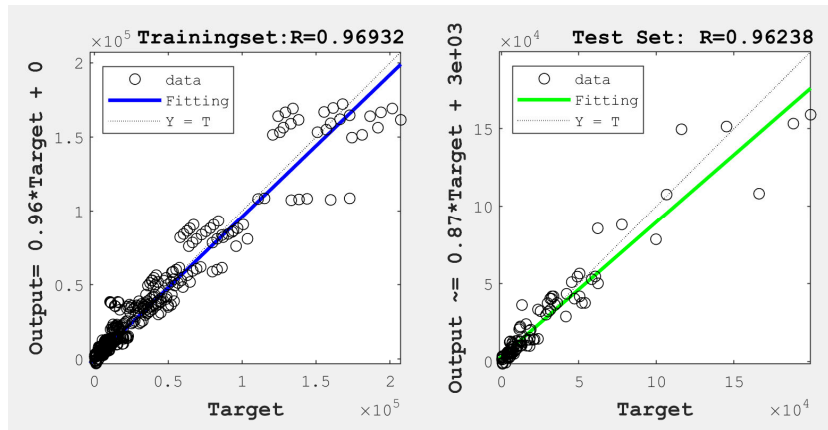


Figure 4. R of the prediction results

Table 1 compares the evaluation indicators of several machine learning algorithms for the model. Among the above error indicators, the LSTM model reduces the error repetition the most. The R values of the other three models are significantly smaller than the LSTM model. The determination coefficient of LSTM is the largest and closest

to 1. From the above data, it can be concluded that the LSTM error index is significantly better than the other three models, indicating that the LSTM model proposed in this paper has the highest prediction accuracy for the liquid transportation distance of LNG long-distance pipelines and has the most obvious advantages.

Table 1. Comparison of evaluation indexes for four models

modal	RMSE	MAE	MSE	MAPE	R
SVM	0.41	2.0123	5.2133	24.5234	0.876
BP	0.33	0.3214	0.8874	12.3782	0.915
RBF	0.12	0.8793	0.4586	14.3213	0.923
LSTM	0.05	0.2213	0.1323	6.2317	0.969

After obtaining the LSTM prediction model, the model is used to predict new data and to construct the relationship between different variables and the LNG liquid transportation distance.

Under the same other conditions, the maximum liquid transportation distance under different inlet pressures was calculated for LNG pipeline transportation, and the results are

shown in Figure 5. When the inlet pressure changes from 1MPa to 2MPa, the maximum liquid transportation distance increases linearly with the change of inlet pressure. The greater the inlet pressure, the longer the transportation distance. This is because the greater the inlet pressure, the higher the subcooling degree of the inlet state, which makes the transportation distance longer.

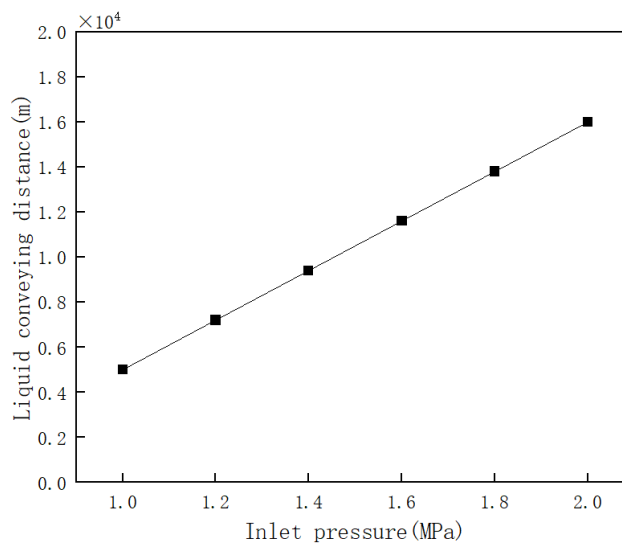


Figure 5. The longest distance changes with inlet pressure

As can be seen from Figure 6, when the heat flux increases, the amount of heat transferred to the low-temperature working fluid per unit time increases, accelerating the liquid phase temperature to approach the saturation temperature, causing the working fluid to undergo a phase change to form

a gas-liquid two-phase flow. The heat exchange between the pipeline and the environment causes the working fluid to absorb heat and rise in temperature, and its temperature rise rate is positively correlated with the heat flux.

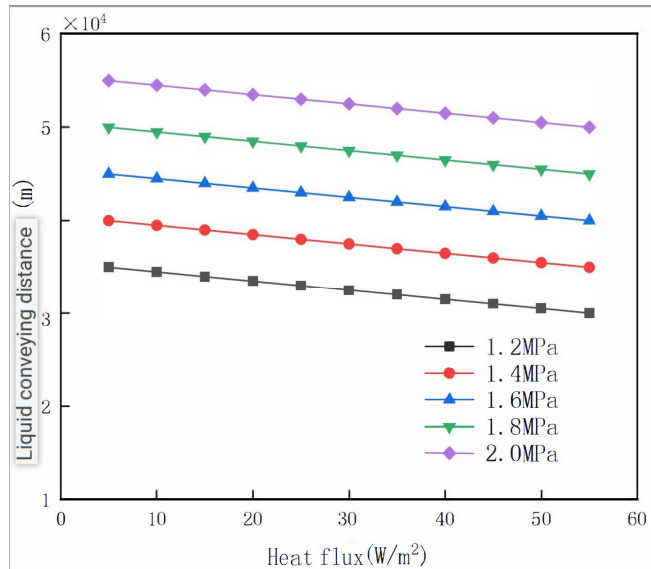


Figure 6. The longest distance changes with heat flux

As shown in Figure 7, the increase in the inner diameter of the LNG long-distance pipeline significantly affects the flow and heat transfer characteristics of the cryogenic medium. The maximum liquid transportation distance of LNG pipeline transportation tends to increase significantly with the increase of the pipe diameter. When the pipe diameter is small, the

transportation distance is short. The greater the inlet pressure, the greater the influence of the pipe diameter on the maximum liquid transportation distance. The trend of increasing the transportation distance with the increase of the pipe diameter is more obvious.

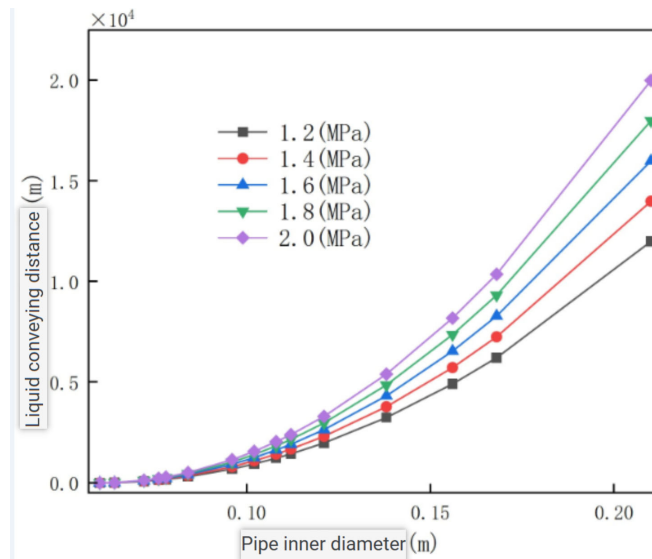


Figure 7. The longest distance changes with pipe inner diameter

4. Conclusion

In this study, we explored the application of LSTM (Long Short-Term Memory) algorithm to predict the liquid transportation distance of LNG long-distance pipelines. As a deep learning algorithm, LSTM can effectively process time series data, especially in scenarios with complex nonlinear relationships and long-term dependencies, showing superior prediction performance. Compared with traditional prediction methods, LSTM can adaptively learn and capture the rules in historical data without manual feature extraction, and has strong robustness to noise and fluctuations in input data. Therefore, the LSTM algorithm not only improves the accuracy of liquid transportation distance prediction, but also provides a more reliable tool for pipeline design,

transportation optimization and safety assurance. The results of this study show that the LSTM algorithm has important application value in the prediction of LNG long-distance pipeline transportation distance, which can help improve transportation efficiency, reduce costs, ensure safety, and provide strong support for decision-making in the energy industry. Future research can further combine more external factors (such as climate change, geographical conditions, etc.) and real-time data to improve the adaptability and generalization ability of the model.

References

[1] Howarth R W. The greenhouse gas footprint of liquefied natural gas (LNG) exported from the United States[J]. Energy Science & Engineering, 2024, 12(11): 4843-4859.

- [2] Nubli H, Jung D, Kim S J, et al. Structural impact under accidental LNG release on the LNG bunkering ship: Implementation of advanced cryogenic risk analysis[J]. *Process Safety and Environmental Protection*, 2024, 186: 329-347.
- [3] Zhu Y, Allen D T, Ravikumar A P. Geospatial life cycle analysis of greenhouse gas emissions from US liquefied natural gas supply chains[J]. *ACS Sustainable Chemistry & Engineering*, 2024, 12(49): 17843-17854.
- [4] Ghiyami M, Bahramian A, Sedae B. Development and performance assessment of a novel multigeneration clean energy system based on biomass-LNG using a dynamic model and supervised learning algorithms[J]. *International Journal of Hydrogen Energy*, 2024, 66: 10-23.
- [5] Balali A, Moghimi M. Energy system design based on municipal solid waste gasification and SOFC for multi-demand applications through multi-heat recovery approach: 4E assessment and machine learning optimization[J]. *Process Safety and Environmental Protection*, 2025.
- [6] Al-Selwi S M, Hassan M F, Abdulkadir S J, et al. RNN-LSTM: From applications to modeling techniques and beyond—Systematic review[J]. *Journal of King Saud University-Computer and Information Sciences*, 2024: 102068.
- [7] Xiang P, Zhang P, Zhao H, et al. Seismic response prediction of a train-bridge coupled system based on a LSTM neural network[J]. *Mechanics Based Design of Structures and Machines*, 2024, 52(8): 5673-5695.
- [8] Kurani A, Doshi P, Vakharia A, et al. A comprehensive comparative study of artificial neural network (ANN) and support vector machines (SVM) on stock forecasting[J]. *Annals of Data Science*, 2023, 10(1): 183-208.
- [9] Zhang B, Xu N, Wang B, et al. A mathematical framework for modeling and designing a long distance cryogenic liquefied natural gas pipeline: A practical study and analysis in Shandong, China[J]. *Energy Conversion and Management*, 2024, 312: 118516.
- [10] Xu K, Hu J, Zhang L, et al. A risk factor tracing method for LNG receiving terminals based on GAT and a bidirectional LSTM network[J]. *Process safety and environmental protection*, 2023, 170: 694-708.
- [11] Wu Y, Yang G, Sun J, et al. Digital twin modeling and leak diagnosis of temperature and stress fields in LNG storage tanks[J]. *Measurement*, 2024, 228: 114374.
- [12] Oyewola D O, Akinwunmi S A, Omotehinwa T O. Deep LSTM and LSTM-Attention Q-learning based reinforcement learning in oil and gas sector prediction[J]. *Knowledge-Based Systems*, 2024, 284: 111290.