

Article

Energy Consumption Prediction for Drilling Pumps Based on a Long Short-Term Memory Attention Method

Chengcheng Wang ^{1,†}, Zhi Yan ^{2,*}, Qifeng Li ², Zhaopeng Zhu ² and Chengkai Zhang ²

¹ Karamay Drilling Company of CNPC Western Drilling Engineering Co., Ltd., Karamay 834000, China; zdwangcc@cnpc.com.cn

² State Key Laboratory of Petroleum Resources and Prospecting, China University of Petroleum, Beijing 102249, China; 2023210716@student.cup.edu.cn (Q.L.); zhuzp@cup.edu.cn (Z.Z.); zhangck@cup.edu.cn (C.Z.)

* Correspondence: 2023315103@student.cup.edu.cn

† These authors contributed equally to this work as co-first author.

Abstract: In the context of carbon neutrality and emission reduction goals, energy consumption optimization in the oil and gas industry is crucial for reducing carbon emissions and improving energy efficiency. As a key component in drilling operations, optimizing the energy consumption of drilling pumps has significant potential for energy savings. However, due to the complex and variable geological conditions, diverse operational parameters, and inherent nonlinear relationships in the drilling process, accurately predicting energy consumption presents considerable challenges. This study proposes a novel Long Short-Term Memory Attention model for precise prediction of drilling pump energy consumption. By integrating Long Short-Term Memory (LSTM) networks with the Attention mechanism, the model effectively captures complex nonlinear relationships and long-term dependencies in energy consumption data. Comparative experiments with traditional LSTM and Convolutional Neural Network (CNN) models demonstrate that the LSTM-Attention model outperforms these models across multiple evaluation metrics, significantly reducing prediction errors and enhancing robustness and adaptability. The proposed model achieved Mean Absolute Error (MAE) values ranging from 5.19 to 10.20 and R^2 values close to one (0.95 to 0.98) in four test scenarios, demonstrating excellent predictive performance under complex conditions. The high-precision prediction of drilling pump energy consumption based on this method can support energy optimization and provide guidance for field operations.

Keywords: energy consumption prediction; LSTM-Attention; drilling pump; time series data prediction



Citation: Wang, C.; Yan, Z.; Li, Q.; Zhu, Z.; Zhang, C. Energy Consumption Prediction for Drilling Pumps Based on a Long Short-Term Memory Attention Method. *Appl. Sci.* **2024**, *14*, 10750. <https://doi.org/10.3390/app142210750>

Academic Editor: Davide Astiaso Garcia

Received: 15 October 2024

Revised: 12 November 2024

Accepted: 19 November 2024

Published: 20 November 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In the context of the growing global energy demand, fossil fuels, particularly oil and natural gas, continue to dominate total energy consumption despite the historically rapid growth of renewable energy sources [1]. During the oil and gas extraction process, energy consumption at drilling sites accounts for a significant portion of the total energy expenditure associated with hydrocarbon extraction. Among the various components of this process [2], the drilling pump serves as one of the key pieces of equipment, ensuring the transportation of drilling mud into the wellbore, maintaining pressure balance within the well, removing cuttings from the bottom of the well, and providing cooling and lubrication to the drill bit. The energy consumption of drilling pumps constitutes a substantial share of overall drilling energy usage [3], making the study of their energy consumption prediction crucial for optimizing drilling energy efficiency.

By collecting operational data from drilling pumps [4], it becomes possible to accurately predict energy consumption in real time, identify any anomalous changes in energy usage, and implement appropriate control and improvements. This can involve adjusting drilling parameters and optimizing energy consumption, thereby achieving energy savings and emission reductions while maintaining drilling efficiency.

In recent years, numerous scholars have proposed various methods for predicting energy consumption [5], with common approaches for drilling pump energy consumption prediction categorized into empirical models, statistical methods, and machine learning techniques. Empirical models primarily rely on empirical data and traditional mathematical models. These methods [6] typically calculate energy consumption based on the operational parameters of the drilling pump (such as pressure, flow rate, and rotational speed) and the physical properties of the drilling mud. While these methods are straightforward and intuitive, the empirical formulas are often based on particular data from particular facilities under particular conditions. Consequently, the accuracy of predictions can be compromised in real-world operations due to factors such as complex geological conditions and equipment aging. Additionally, empirical models [7] struggle to meet the demands of energy consumption prediction under complex operating conditions, particularly when dealing with multivariable and highly nonlinear relationships, resulting in a significant decline in prediction accuracy.

Statistical methods involve analyzing historical data and building models to forecast future energy consumption [8], with common approaches including Autoregressive Integrated Moving Average (ARIMA) models and Gray Prediction Models (GMs). An ARIMA model employs a combination of autoregressive and moving average analyses on time series data to establish a mathematical framework for predicting future trends. In contrast [9], a Gray Prediction Model is suitable for small samples of and utilizes Gray System Theory to process time series data and forecast future energy consumption changes. While these methods perform well in capturing linear relationships and short-term predictions, their effectiveness diminishes in addressing the complexities and nonlinear characteristics of the oil drilling process. Furthermore, statistical methods are sensitive to outliers and noise, which can distort the prediction models and negatively impact the accuracy of results.

Machine learning prediction methods primarily include Random Forests (RFs), Support Vector Machines (SVMs) [10], and Gradient Boosting Trees (GBTs). Random Forests improve model generalization by constructing numerous decision trees based on the principles of ensemble learning, allowing for the processing of extensive feature data. Support Vector Machines find optimal separating hyperplanes to accomplish classification and regression tasks in high-dimensional spaces. Gradient Boosting Trees build a series of weak learners incrementally [11], combining their predictions to achieve efficient and accurate energy consumption forecasting. These machine learning methods are particularly effective in handling complex nonlinear relationships, accommodating more intricate patterns and interactions in drilling pump energy consumption predictions. However, these methods face certain challenges. First, machine learning models often require extensive parameter tuning, which can lead to lower computational efficiency and a complex, time-consuming training process, making it difficult to meet the real-time requirements of prediction. Second, these methods are prone to overfitting when processing large-scale data, resulting in models that perform well on training data but poorly on unseen data. Additionally, because machine learning techniques depend on large datasets for training, their predictive accuracy can be significantly affected when data are scarce or of low quality.

Deep Learning [12], a subfield of machine learning, utilizes multilayer neural networks to simulate the way the human brain processes information, automatically extracting features from data for prediction or classification tasks. Compared to traditional machine learning methods [13], deep learning is capable of handling more complex tasks such as image recognition, speech recognition, and natural language processing. The core concept involves using multiple layers of neural networks, where each layer progressively extracts abstract features from the data, thereby achieving higher accuracy on large datasets. Key technologies in deep learning include Artificial Neural Networks (ANNs), Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Generative Adversarial Networks (GANs), Autoencoders, and Transformers. These techniques have made significant advancements in various fields, driving the rapid development of artificial intelligence.

Artificial Neural Networks (ANNs) [14] form the foundation of deep learning, mimicking the functioning of biological neurons. An ANN consists of an input layer, one or more hidden layers, and an output layer, with full connections between layers. This structure allows ANNs to achieve complex nonlinear mappings. In the context of energy consumption prediction, ANNs can capture intricate relationships between operational parameters and energy usage. However, due to their high model complexity, ANNs, despite often performing well on training data, may exhibit poor generalization to unseen data. Furthermore, in the absence of substantial feature engineering, ANNs may struggle to extract useful information from raw data. Convolutional Neural Networks (CNNs) [15], initially designed for image processing, can also be applied to energy consumption prediction for time series data. By utilizing convolutional layers to extract local features, CNNs can capture spatial or temporal dependencies within the data. In energy consumption prediction, CNNs are effective in identifying local patterns in drilling pump operation data, such as energy consumption variations over specific time periods. Nevertheless, while CNNs excel at capturing short-term dependencies, their performance deteriorates when dealing with long-term dependencies, limiting their applicability to certain complex time series prediction tasks. Recurrent Neural Networks (RNNs) [16] are particularly well suited for handling sequential data. Unlike traditional feedforward neural networks, RNNs share parameters across the temporal dimension through their recurrent structure, enabling them to capture dependencies in time series data. This characteristic makes RNNs highly effective in tasks such as time series data analysis, speech recognition, and natural language processing. However, when dealing with long sequences, RNNs are prone to issues such as vanishing or exploding gradients, which can make the model difficult to train and impair its ability to effectively capture long-term dependencies.

The primary advantage of Long Short-Term Memory (LSTM) [17] networks lies in their ability to effectively handle long-term dependencies in time series data. The energy consumption of drilling pumps is a typical time series problem, where the current energy state is influenced not only by the present operational parameters but also by past operating conditions and process parameters over a period of time. Although traditional Recurrent Neural Networks (RNNs) are capable of processing time series data, they often suffer from vanishing or exploding gradient problems when dealing with long sequences, resulting in poor performance in capturing long-term dependencies. As a deep learning model, LSTM can automatically extract features from data without relying on manual feature engineering. Traditional empirical models and statistical methods typically depend on handcrafted features and are based on linear assumptions. However, the energy consumption of drilling pumps is affected by various nonlinear factors, such as pressure fluctuations, flow rate variations, and mechanical wear, which are difficult to capture accurately using conventional methods. LSTMs, with their multi-layered network structure, can automatically learn and extract complex patterns and nonlinear relationships from data without human intervention. This capability enables LSTMs to better understand and model the complex interactions between energy consumption and various operational parameters in drilling pump energy prediction, thereby improving predictive accuracy. Moreover, drilling pump energy consumption data may not always be collected at regular time intervals, or different time periods may involve different critical variables. Traditional predictive models often assume uniformly distributed data and can only handle a limited number of input variables, which constrains their applicability in real-world scenarios.

LSTM models are capable of flexibly handling irregular time intervals and multivariate inputs [18]. By dynamically adjusting the time steps, LSTM can adapt to data collected at varying time intervals and simultaneously process multidimensional input data, capturing complex interactions between these variables. This feature enhances LSTM's adaptability and predictive capability in dealing with the complexity of real-world data. To address these issues, this paper proposes an LSTM-based approach for predicting the energy consumption of drilling pumps, with an attention mechanism integrated to enhance the model's ability to capture key features, thereby improving prediction accuracy and interpretability. By

incorporating the attention mechanism [19] into the LSTM model, the proposed approach can automatically focus on the most critical time steps and features for energy consumption prediction. This not only enhances predictive performance but also provides deeper insights into the impact of different time points and conditions on drilling pump energy consumption. This study offers a novel technical solution to achieve accurate energy consumption prediction for drilling pumps and provides valuable references for energy optimization and efficiency improvement in production.

This paper proposes an LSTM algorithm based on the attention mechanism to handle complex temporal relationships for predicting drilling pump energy consumption. The method first introduces the Random Sample Consensus (RANSAC) algorithm [20] to select valid data based on the correlations between pump pressure, flow rate, and pump power. Then, it combines LSTM and the attention mechanism for intelligent prediction of drilling pump energy consumption. Leveraging the model's capability to capture long-term sequence data features, the proposed approach demonstrates excellent performance in predicting pump power, thereby extending the application of artificial intelligence in the performance prediction of drilling equipment.

2. Materials and Methods

The Long Short-Term Memory (LSTM) network, as illustrated in Figure 1, is designed to address the limitations of traditional Recurrent Neural Networks (RNNs) in handling long-term dependencies and mitigating the issues of vanishing and exploding gradients. The LSTM cell is the fundamental building block of the LSTM network, featuring a unique internal structure that incorporates three key gating mechanisms: the input gate, the forget gate, and the output gate. These gates regulate the flow of information, enabling the model to effectively capture long-term temporal dependencies in sequential data. The following section provides a detailed explanation of the components and functioning of the LSTM network, as depicted in the diagram.

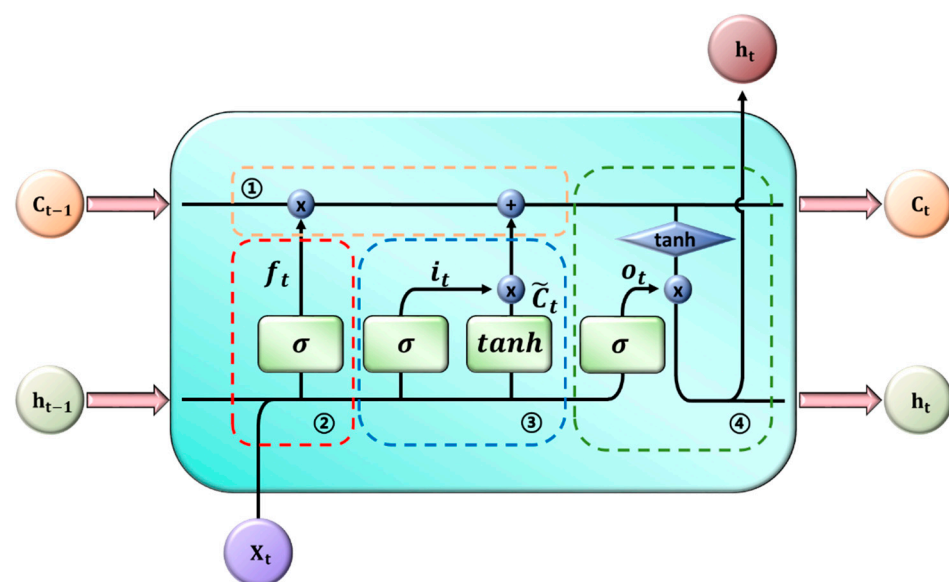


Figure 1. The structure of the LSTM.

The forget gate f_t determines the extent to which information from the previous cell state C_{t-1} should be retained or discarded. It takes the hidden state from the previous time step h_{t-1} and the current input x_t as inputs and computes a forget factor, which ranges between 0 and 1, where 0 indicates complete forgetting and 1 indicates complete

retention of the information. This operation is represented in section 1 of the Figure 1. The computation is defined by the following equation:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (1)$$

The input gate i_t controls how much of the new information x_t should be updated in the cell state C_t . It consists of two parts: the input gate and the candidate cell state. The input gate, regulated by a sigmoid function, determines the significance of the current input, while the candidate cell state, generated by a tanh function, creates new information \tilde{C}_t to be potentially added to the cell state. This process is illustrated in section 2 of the Figure 1. The equations for the input gate are as follows:

$$\begin{aligned} i_t &= \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \\ \tilde{C}_t &= \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \end{aligned} \quad (2)$$

The cell state C_t is the core component of the LSTM cell, maintaining the memory of the network. The cell state is updated by combining the previous cell state C_{t-1} and the current candidate cell state \tilde{C}_t , weighted by the forget gate and input gate, respectively. This step is depicted in section 3 of the Figure 1. The cell state update is governed by the following equation:

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \quad (3)$$

The output gate o_t determines the value of the hidden state h_t for the current time step, which is subsequently used as input for the next time step. The output gate applies a sigmoid function to control the amount of information from the cell state that is passed to the hidden state. This operation is shown in section 4 of the Figure 1. The output from the hidden state is modulated by the tanh of the updated cell state C_t . The equations for the output gate are as follows:

$$\begin{aligned} o_t &= \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \\ h_t &= o_t \cdot \tanh(C_t) \end{aligned} \quad (4)$$

LSTM networks are highly effective for predicting drilling pump energy consumption due to their ability to capture the temporal dependencies inherent in such data. By leveraging its gated architecture, an LSTM can accurately model the sequential patterns and complex relationships between various operational parameters and energy usage.

As illustrated in Figure 2, the attention mechanism is a neural network architecture designed to effectively capture important information within data. It has been widely applied in various fields such as natural language processing and computer vision. The core idea of the attention mechanism is to dynamically adjust the importance of different elements in the input sequence by calculating the similarity between them. The figure depicts the structure of the Scaled Dot-Product Attention mechanism and its computation process.

In Figure 2, the attention mechanism is composed of three main components: Query (Q), Key (K), and Value (V). These components are used to compute the contribution of each input element to the output. By calculating the similarity between the query vector Q and the key vector K, the weights for the value vector V are determined, resulting in a weighted sum of the values that contributes to the output.

The first step involves performing a matrix multiplication between the query vector Q and the transpose of the key vector K to obtain a score matrix representing the similarity between the query and the keys. This is computed as follows:

$$\text{Score} = Q \times K^T \quad (5)$$

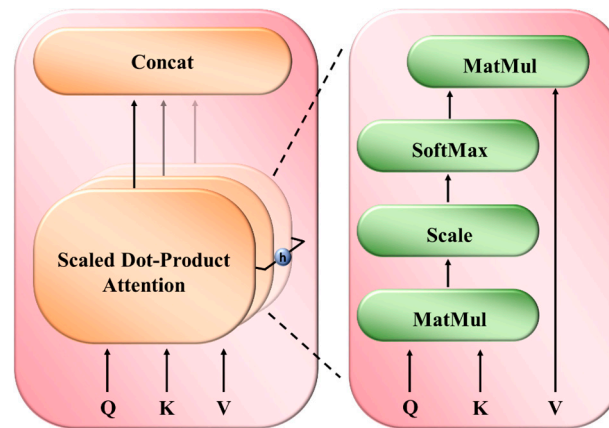


Figure 2. Attention mechanism.

To avoid excessively large dot product values, which could lead to vanishing gradients, the score matrix is scaled by dividing it by the square root of the dimensionality of the key vector, $\sqrt{d_k}$. The formula is given by the following equation:

$$\text{Scaled_Score} = \frac{Q \times K^T}{\sqrt{d_k}} \quad (6)$$

Next, the scaled score matrix is passed through the SoftMax function to convert it into a weight matrix, where the elements of each row sum to 1. This process represents the relative importance of each key in relation to the query:

$$\text{Attention_Weights} = \text{SoftMax}(\text{Scaled_Score}) \quad (7)$$

Once the weight matrix is obtained, it is multiplied by the value vector V to generate the final output. The formula for this operation is as follows:

$$\text{Attention_Output} = \text{Attention_Weights} \times V \quad (8)$$

In the context of drilling pump energy consumption prediction, the attention mechanism can identify the impact of different time points or features on energy consumption variations in time series data. Its advantage lies in the dynamic allocation of importance weights to different features, overcoming the limitations of fixed weights in traditional methods and improving the accuracy of the model's predictions.

3. Results and Discussion

3.1. Data Processing

The data processing procedure in this study includes the following steps: outlier detection, data filtering, data normalization, and data segmentation. Each step is described in detail below.

Outlier detection and removal were conducted using the RANSAC (Random Sample Consensus) algorithm. RANSAC is a robust model fitting method particularly suitable for datasets containing a significant number of outliers. It iteratively selects a random subset of the data and fits a model, classifying data points as inliers or outliers based on their distance from the model. In this study, pump pressure (MPa), inlet flow rate (L/min), and pump power (kW) were selected as input features for outlier detection using the RANSAC algorithm.

As shown in Figure 3, during each iteration, a random set of data points is selected to establish a linear regression model relating pump pressure, inlet flow rate, and pump power. The deviation of all data points from this model is then calculated, and those with deviations below a certain threshold are classified as inliers, while the rest are classified as

outliers. This process is repeated multiple times to ensure that the selected set of inliers is as large as possible. In this study, the main parameters of the RANSAC algorithm include the maximum number of iterations and the deviation threshold. The maximum number of iterations was set to 1000 to ensure the algorithm's convergence, and the deviation threshold was set to a small value based on the data distribution to strictly select the inliers. After processing with the RANSAC algorithm, a total of 1034 data points were identified as outliers due to their large deviations from the majority of the data points and were removed. Finally 8715 valid data points were retained for subsequent model training and analysis.

Outlier Detection Based on the RANSAC Algorithm

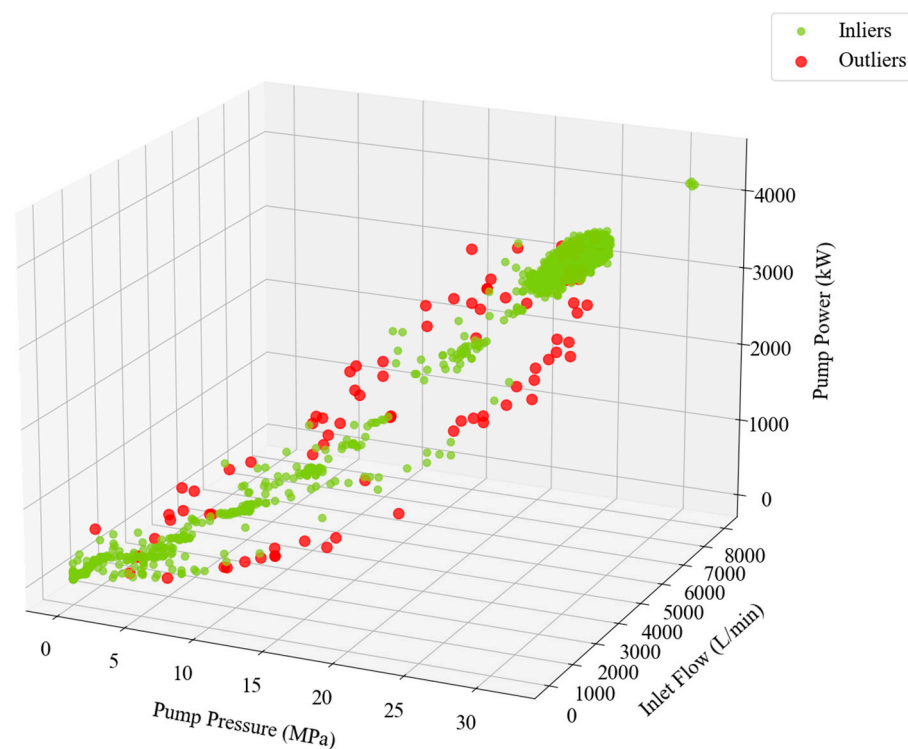


Figure 3. Application of the RANSAC mechanism.

Figure 4 shows the utilization of a sliding window method to collect real-time pumping power data. Each sliding window corresponds to a distinct data stream (e.g., X_0 , X_1 , and X_2), with varying time intervals for each window. This design ensures the independence of data during training and testing processes by preventing overlaps across windows. The training and testing of each data stream are conducted independently, and within a single data stream, there is no overlap between the test data windows and training data windows. To ensure data independence and the validity of the test results, we train and test each data stream separately, rather than combining all data streams into a single model for training. Therefore, each data stream's model is trained in isolation, without any interconnection or influence from other streams. This approach allows the model to capture patterns across different time periods while avoiding potential issues arising from high correlation among neighboring samples. By ensuring the separation of training and testing data, the model can be trained without exposure to test data, enhancing its reliability in predicting future unseen data. This design reduces the risk of data leakage and ensures the validity of validation and testing processes. Moreover, using samples from different sliding windows provides a more comprehensive assessment of the model's generalization capability, preventing bias in test results.

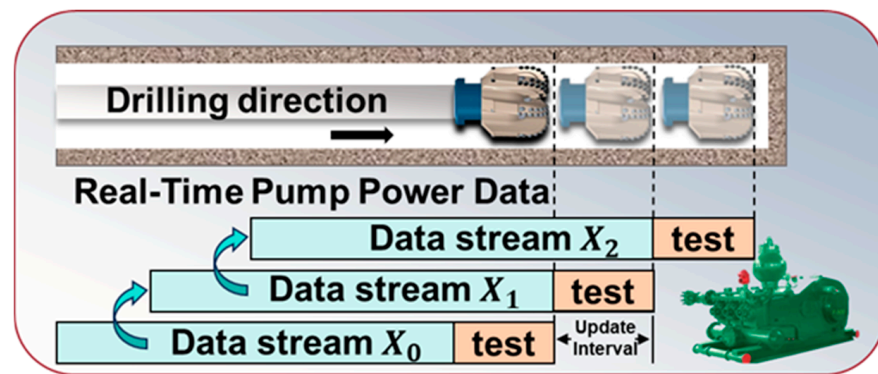


Figure 4. Independent sliding window approach for real-time pump power data.

Data normalization was conducted to unify the value ranges of different features and eliminate the impact of scale differences on model training. In this study, we chose the Max normalization method based on the following considerations: during data preprocessing, we had already removed outliers from the dataset; therefore, we believed that Max normalization would not be significantly affected by any remaining extreme values. This method effectively scales all features to a similar range, ensuring that each feature contributes more evenly to the loss function during model training and accelerates the model's convergence. While we also experimented with other normalization methods, such as standard deviation-based normalization, our results showed that Max normalization provided better training efficiency and model performance. Therefore, given the characteristics of our data, Max normalization is a reasonable and effective choice for this study. The normalization formula is as follows:

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (9)$$

This formula scales all feature values to the range [0, 1]. The process involves calculating the maximum and minimum values for each feature (such as pump pressure, inlet flow rate, and pump power) and then transforming all data points for that feature using the normalization formula. After normalization, all feature values are compressed to the range [0, 1], eliminating the differences in scale between different features. This standardization ensures that the influence of each feature on the loss function is balanced during model training and accelerates the model's convergence process, thereby improving training efficiency.

In time series prediction tasks, the sliding window method is commonly used to segment the data to capture temporal dependencies. In this study, the sliding window length was set to 100 data points, and the sliding step was set to 30 data points, generating multiple overlapping time segments. Each sliding window contains 100 consecutive data points, serving as input features for the model. These 100 data points encapsulate the complete temporal information, enabling the model to capture temporal dependencies within the data. With a step size of 30 data points, there is an overlap of 70 data points between successive windows. This setting allows the generation of more training samples without losing historical information, enhancing the model's generalization capability and predictive accuracy. The target variable for each window is the pump power value at the end of the window, indicating the influence of the current time segment's input features on the pump power. Through this configuration, the model can learn patterns in the time series data and make accurate energy consumption predictions. Starting from the first data point, a new window is generated every 30 data points until the end of the data is reached. Each window's 100 data points are defined as input features, with the corresponding pump power value as the output label. This process resulted in multiple sample pairs for subsequent model training and evaluation. The sliding window setup effectively captures complex relationships in the time series, improving the model's predictive performance.

After data segmentation, all generated sample pairs were randomly divided into training and test sets for model training and performance evaluation. The proportions of data assigned to the training and test sets were set at 70% and 30%, respectively. The training set, containing 70% of the data samples, was used to train the model. The model learns the patterns and rules within the training data by continuously optimizing its internal parameters to fit the data. The test set, containing 30% of the data samples, was used to evaluate the model's performance.

3.2. Construction of LSTM-Attention Model

As illustrated in Figure 5, the LSTM-Attention model comprises three main components: the LSTM layer, the attention layer, and the dense layer. This model integrates the capabilities of temporal feature extraction and dynamic weight allocation to better address the complex data patterns associated with predicting drilling pump energy consumption. The following sections elaborate on the design principles and functionalities of each layer in the context of practical applications.

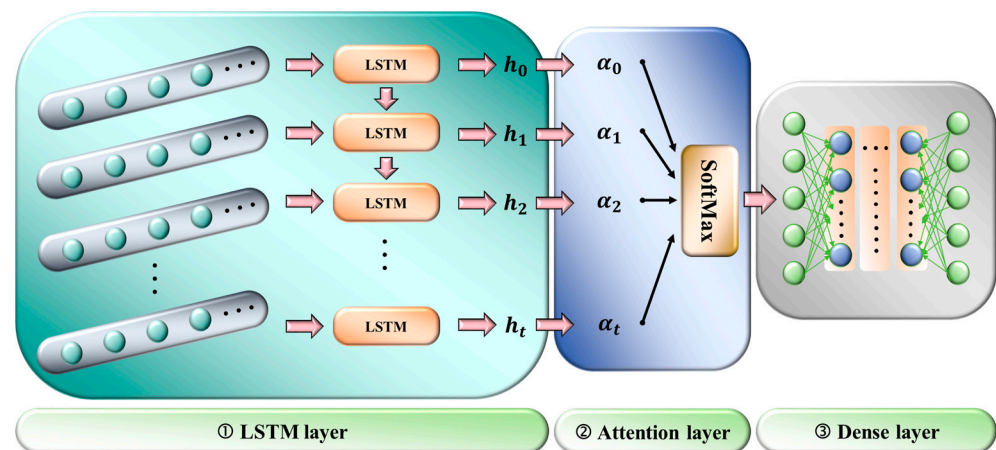


Figure 5. LSTM-Attention model.

The LSTM layer is designed to handle the complex temporal dependencies within the sequential data and extract dynamic features of energy consumption changes during drilling operations. The energy consumption of drilling pumps is influenced by multiple factors such as pump pressure, flow rate, drilling depth, mud density, and mechanical wear. These factors exhibit intricate dependencies over different time steps. In the input stage, drilling data are fed into the LSTM layer in the form of time sequences. Each LSTM unit utilizes its internal memory and gating mechanisms to effectively capture the dynamic changes and long-term dependencies among various parameters in the time series. The LSTM layer maps the input sequence to a series of hidden states, which encapsulate historical information across different time steps, providing a detailed feature representation for subsequent layers. For example, as drilling progresses through various geological formations, the pump's operating conditions may change significantly. The LSTM layer can identify and retain these temporal variation features, enabling accurate prediction of future energy consumption trends.

The attention layer is introduced to address the varying contributions of different time steps and parameters to energy consumption. During drilling, operational conditions such as drilling depth and formation hardness can have differential impacts on pump energy consumption. The attention mechanism automatically allocates weights to each time step, allowing the model to focus on the key time periods that most significantly affect energy consumption. In the figure, each hidden-state output by the LSTM layer is assigned an attention weight through the attention mechanism. These weights are normalized using the SoftMax function, reflecting the relative importance of different time steps. This enables the attention layer to focus on time periods or specific operational parameters that have

a greater impact on energy consumption variations, thereby extracting more meaningful contextual information. The context vector is generated by a weighted sum of the hidden states, representing the most important information throughout the sequence. For drilling pump energy consumption prediction, this mechanism helps the model capture critical features during operational changes, such as increased wear on the drill bit or variations in formation properties, and accurately predict their impact on energy consumption.

The dense layer, also known as the fully connected layer, is responsible for mapping the contextual information extracted by the attention layer to specific energy consumption predictions. The context vector, which summarizes the key time steps of the sequence, serves as the core basis for energy consumption prediction. In the context of drilling pump energy consumption prediction, the model needs to consider the combined influence of multiple operational parameters on energy consumption. Therefore, the dense layer not only captures the direct impact of each parameter but also models the interactions between different parameters. For example, combined variations in pump pressure and flow rate can lead to nonlinear fluctuations in energy consumption. The dense layer can model and predict these complex relationships through nonlinear mappings. By transforming the high-dimensional features in the context vector into specific energy consumption values, the dense layer provides precise prediction outputs.

The structural design of the LSTM-Attention model specifically addresses the business needs of drilling pump energy consumption prediction, focusing on resolving several key challenges:

Capturing Complex Temporal Dependencies: The LSTM layer effectively captures the temporal dependencies in energy consumption variations under different operational conditions, such as the dynamic changes in pump pressure, flow rate, and drilling depth.

Focus on Key Features: The attention mechanism identifies the most impactful time steps and parameters for energy consumption prediction, allowing the model to focus on features that are critical to the prediction outcome, thus enhancing prediction accuracy.

Improvement in Prediction Accuracy: The dense layer integrates all critical contextual information, providing more accurate prediction outputs. This is crucial for real-time decision support and energy optimization in drilling operations.

3.3. Experiments and Analysis

Based on this model, we predicted the drilling pump energy consumption under four different drilling conditions, and the test results are shown in Figure 6. The four experimental results illustrate the model's performance in predicting drilling pump power consumption under various time periods and working conditions. By comparing the actual data (blue dots) with the predicted values (orange line) across different experiments, we can conduct an in-depth analysis from several perspectives, including overall trend prediction, short-term fluctuation capture, high-frequency response, and performance under specific operating conditions. Each aspect is discussed in detail below.

In time series forecasting, the ability to capture long-term trends and short-term fluctuations is a critical criterion for evaluating model performance. Thus, Figures 7–10 are not merely enhancements for visual presentation; they serve as validation of the model's capabilities in these two key aspects.

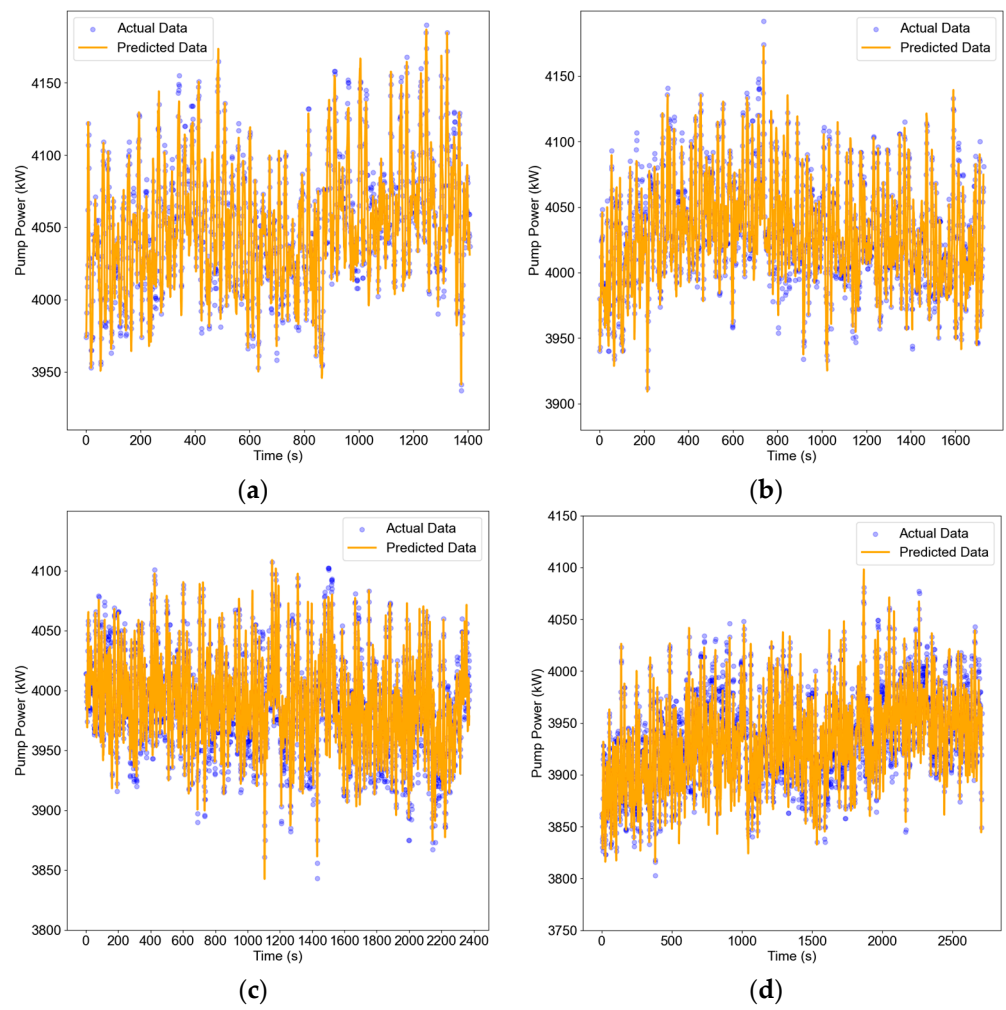


Figure 6. Pump power prediction experiments: (a) Test 1; (b) Test 2; (c) Test 3; (d) Test 4.

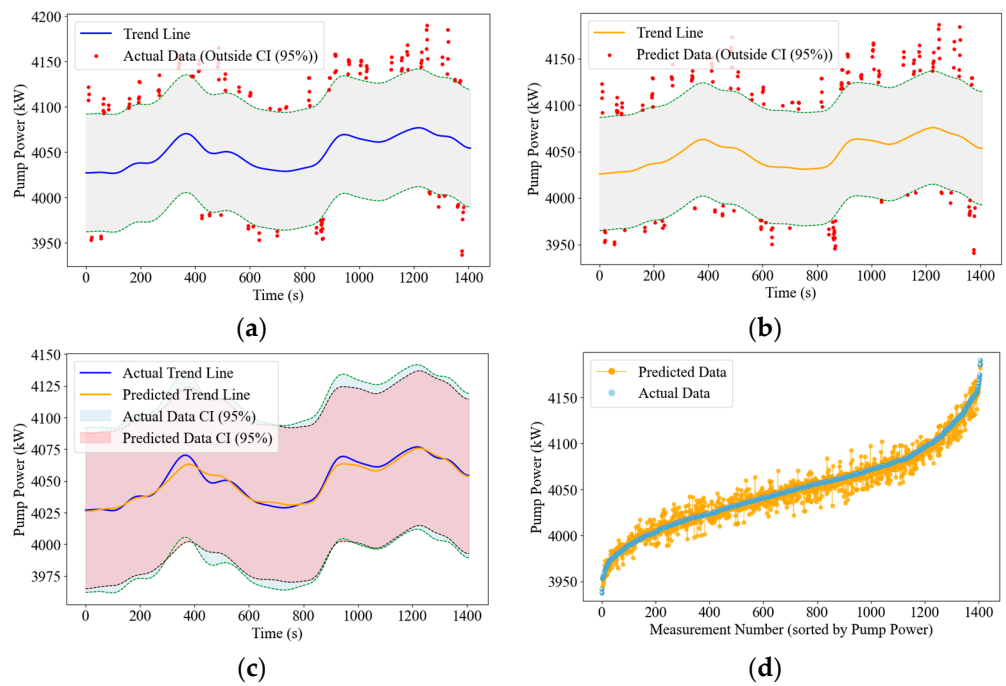


Figure 7. Trends and fluctuations in Test 1.

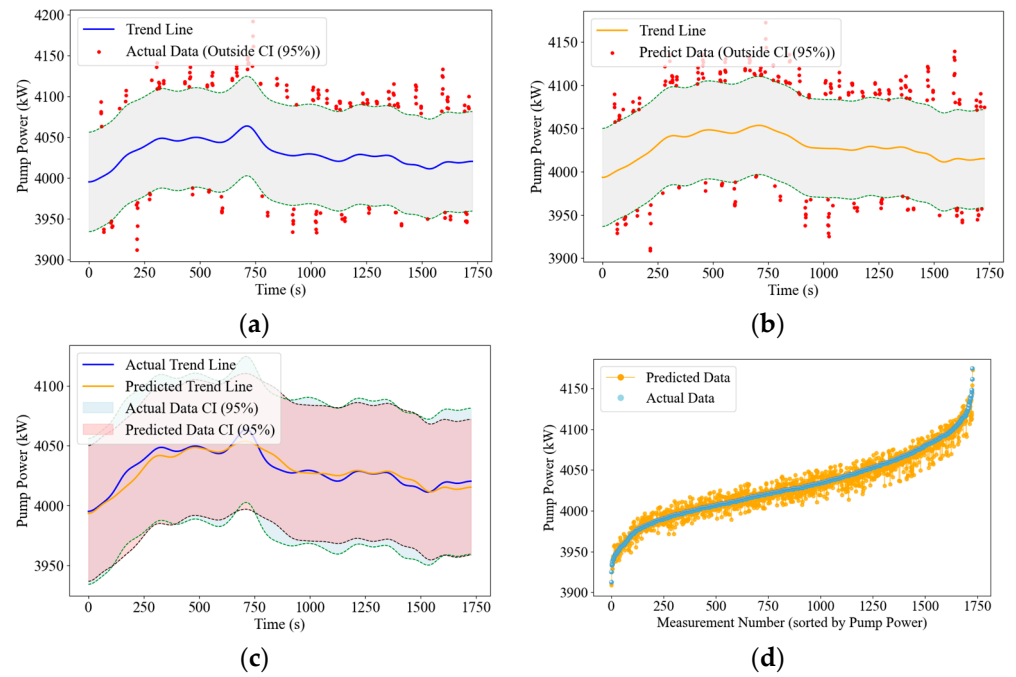


Figure 8. Trends and fluctuations in Test 2.

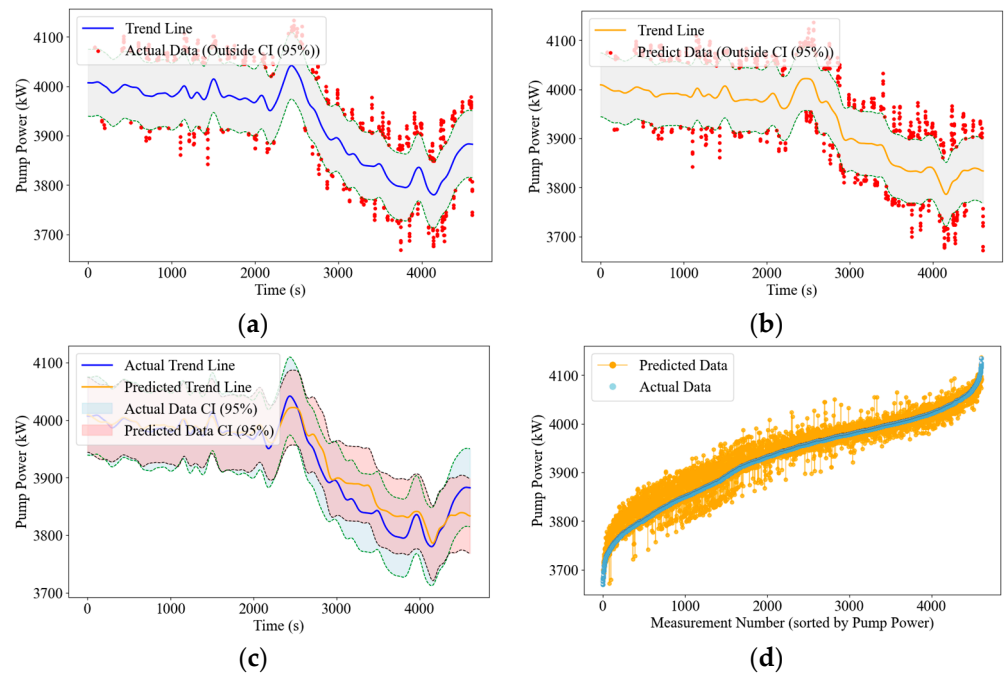


Figure 9. Trends and fluctuations in Test 3.

As shown in Figures 7–10, the actual and predicted trend lines in the image indicate that the model can capture the major changes in electricity at different times, where predicted values closely follow the fluctuations of actual data. This predictive capability demonstrates the model’s ability to capture long-term trends effectively. The shaded area represents the 95% confidence interval. The model’s predicted trend mostly falls within the confidence interval of the actual trend, indicating a certain level of reliability and accuracy. In the figure, the measurement numbers (instead of time) are sorted and plotted, providing a clearer perspective for observing the subtle differences between the predicted and actual

data. The predicted data and actual data show close consistency in trends, especially where significant fluctuations occur.

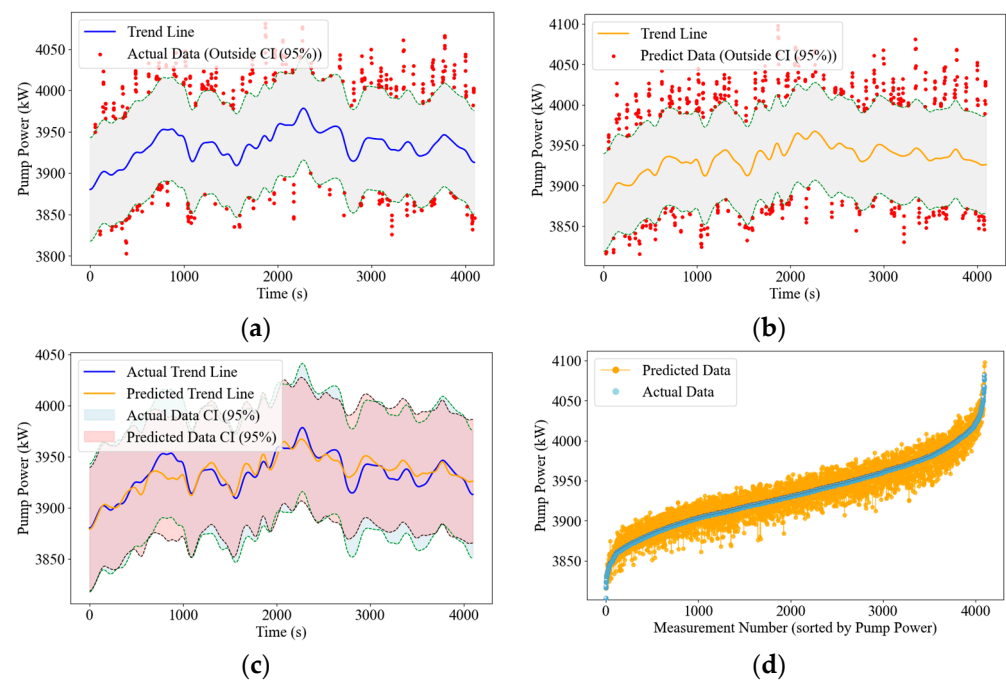


Figure 10. Trends and fluctuations in Test 4.

Firstly, as shown in Figures 7a–c, 8a–c, 9a–c and 10a–c, the confidence intervals for the actual and predicted data in these figures cover the 95% stable range of each dataset, and the trend lines plotted based on these intervals illustrate the long-term trends in the data. By comparing the positions and shapes of the actual and predicted trend lines, we can assess the model’s accuracy and consistency in capturing long-term trends. These trends provide insights into the overall trajectory of energy consumption, which is essential for long-term energy planning and optimization. Secondly, as shown in Figures 7d, 8d, 9d and 10d, to better evaluate the model’s performance in capturing short-term fluctuations, we sorted the actual power data as a baseline. By comparing the relative positions of the predicted power values against this baseline, we can directly analyze the model’s responsiveness to minor fluctuations and rapid changes. This capability is crucial for real-time monitoring and anomaly detection in practical applications, as it enables us to identify short-term variations in the system and adjust operational strategies accordingly.

Overall Trend Prediction. The results from all four experiments indicate that the model can accurately capture the overall trend of actual pump power fluctuations. In each experiment, the overall fluctuation trend of the predicted values (orange line) closely aligns with the actual data (blue dots). The model effectively follows the changes in actual data, whether during periods of increasing or decreasing power. This demonstrates the model’s strong performance in recognizing the long-term trends and patterns of pump power variations. Figure 7a: In the 0–1400 s time range, pump power exhibits frequent upward and downward fluctuations, ranging between 3950 kW and 4150 kW. The predicted values closely follow the actual data, especially during the 600–1000 s interval, where the model accurately predicts both the rise and fall in power. This demonstrates the model’s ability to adapt to complex fluctuation patterns during this period, capturing the peaks and troughs with high precision. Figure 8b: The time range is extended to 1600 s. Compared to Experiment 1, the power fluctuations are relatively moderate, ranging between 3950 kW and 4100 kW. Despite the smoother fluctuations, the model still accurately tracks the changes, particularly during the 600–1200 s stable period, where the predicted values align almost perfectly with the actual data, highlighting the model’s high prediction accuracy under

stable operating conditions. Figure 9c: This experiment spans a longer duration (0–2400 s) and includes several distinct phases of power changes. The model accurately captures these changes, especially during the power reduction phase between 1400–1800 s. The model also successfully predicts the rebound trend around 2000 s, showcasing its reliability in predicting long-term trends. Figure 10d: Covering a 2500-second interval, the pump power fluctuates significantly between 3750 kW and 4150 kW. The model provides precise predictions throughout this extended period, particularly during the intervals of 0–1000 s and 2000–2500 s, where significant power rises and falls are observed. This indicates the model's strong adaptability when dealing with long-term and wide-range fluctuations.

Effective Short-term Fluctuation Capture. In all experiments, the actual data exhibit frequent short-term fluctuations, reflecting the dynamic changes in operational parameters such as pump pressure, flow rate, and drilling depth during the drilling process. The model is able to accurately predict the amplitude and frequency of these short-term fluctuations in most cases, indicating its excellent performance in short-term forecasting. Figure 7a: During the periods of 200–400 s and 800–1000 s, the pump power shows frequent rapid fluctuations. The predicted line almost perfectly matches these peaks and troughs, demonstrating the model's capacity to accurately capture rapid changes in pump power. This capability is crucial for identifying and anticipating sudden events or operational adjustments that impact energy consumption. Figure 8b: In the intervals of 600–800 s and 1200–1400 s, the pump power experiences relatively stable short-term fluctuations. The predicted values closely follow the actual data, indicating the model's ability to handle minor fluctuations in stable conditions effectively. This ensures stable prediction performance under calm operating conditions. Figure 9c: The model precisely predicts the short-term fluctuations during the periods of 1200–1400 s and 2000–2200 s. It accurately tracks the peaks and troughs, reflecting the model's quick response to rapid changes in power, which is beneficial for real-time monitoring and control. Figure 10d: During the periods of 500–700 s and 2000–2200 s, frequent high-frequency fluctuations are observed in the actual data. The model captures most of these fluctuations but shows slight lag during some abrupt changes around 2000 s. This suggests that while the model performs well in most high-frequency scenarios, its response speed and precision can still be improved in handling extremely rapid changes.

High-frequency Response and Analysis of Specific Operating Conditions. While the model shows good robustness in handling high-frequency fluctuations and special operating conditions, it still exhibits some deviations during sudden severe fluctuations or abnormal conditions. Figure 7a: Around 1100 s, the actual data shows a sudden surge in pump power. The model's prediction lags slightly in this instance, indicating a delayed response to sudden power increases. This suggests that the model may require more detailed operational data, such as bit wear status and formation hardness, to improve its accuracy in such scenarios. Figure 8b: During the period of 1200–1600 s, the model successfully predicts multiple short-term increases and decreases in power, almost mirroring the actual values. This demonstrates the model's high prediction accuracy under steady and slow fluctuation conditions. Figure 9c: Around 2100 s, the actual pump power drops rapidly to 3850 kW. While the model shows a slight lag in prediction, it still captures the downward trend well, indicating good performance even in the face of sudden large fluctuations. Figure 10d: Around 2000 s, there is a rapid increase in pump power. The model underestimates this surge, suggesting that its training may lack sufficient representation of similar extreme cases. Enhancing the dataset with more samples of such events or refining the model architecture could improve performance in these instances.

Overall, the model demonstrates strong performance in predicting drilling pump energy consumption, accurately capturing both overall trends and short-term fluctuations in actual energy consumption. It provides reliable predictions under most operating conditions. Additionally, the model exhibits good robustness in handling high-frequency fluctuations and complex operating scenarios, effectively responding to common variations encountered during the drilling process.

3.4. Model Performance Comparison

Based on the above experiments, we conducted a comparative analysis of the performance of different models in predicting drilling pump energy consumption. We employed independent testing datasets collected from pumps operating in different oilfields, with no overlap with the training data. These datasets were gathered from diverse geographic locations and under varying operational conditions. Consequently, the model's performance on the test data reflects not only its ability to fit the training data but also its predictive robustness across different data sources and changing conditions. To evaluate the models, we introduced four evaluation metrics: Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Square Error (RMSE), and the Coefficient of Determination (R^2). The formulas for these metrics are as follows:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (10)$$

MAE measures the average absolute difference between predicted and actual values, reflecting the overall prediction error of the model.

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (11)$$

MSE represents the average squared difference between predicted and actual values and is sensitive to large errors. A lower MSE indicates better control over the overall prediction error.

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (12)$$

RMSE is the square root of MSE, providing a measure of the average deviation between predicted and actual values. RMSE is also sensitive to large errors and offers a comprehensive evaluation of model performance.

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (13)$$

R^2 reflects the proportion of variance in the dependent variable that is predictable from the independent variables. An R^2 value closer to one indicates a better fit of the model to the data.

These four metrics provide a comprehensive assessment of the model's performance in predicting drilling pump energy consumption. MAE and MSE primarily evaluate the magnitude of prediction errors, while RMSE offers an overall evaluation by accounting for both small and large errors. R^2 measures how well the model captures the overall trend in the data. By comparing these metrics, we can better understand the strengths and weaknesses of different models, providing valuable insights for further optimization and improvement of energy consumption prediction models. The performance comparison is shown in Figure 11.

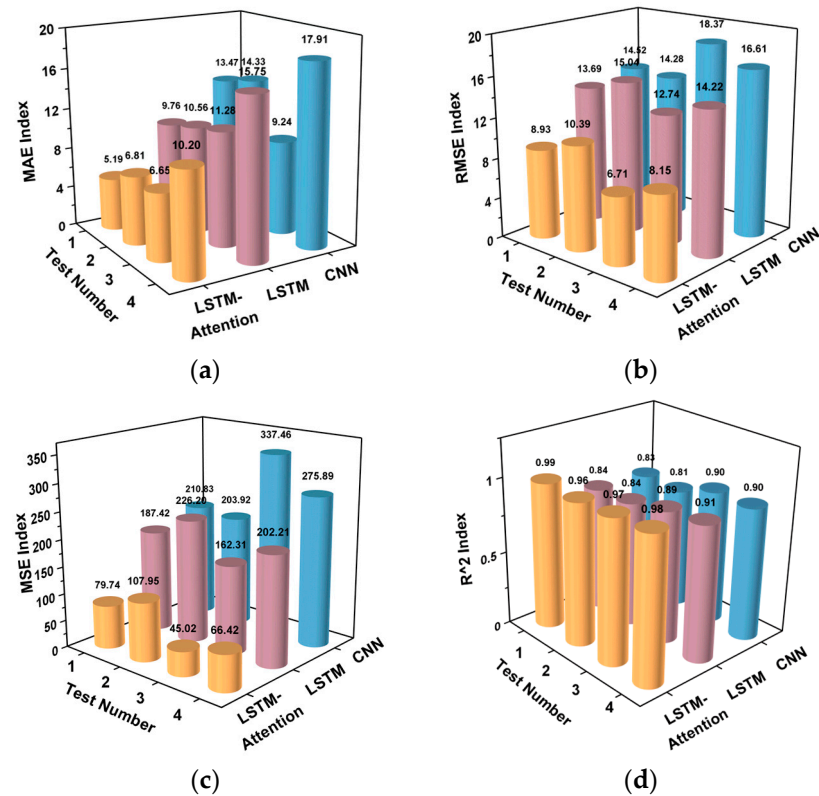


Figure 11. Performance comparison with different indices: (a) MAE; (b) RMSE; (c) MSE; (d) R².

As illustrated in the figures, the LSTM-Attention model outperforms the traditional LSTM and CNN models across all evaluation metrics (MAE, MSE, RMSE, and R²), demonstrating superior predictive capabilities and robustness. The MAE and MSE values indicate that this model effectively reduces prediction errors, especially when handling complex nonlinear relationships and long-term dependencies, with significantly lower error rates than other models, showcasing its high accuracy and reliability. The RMSE metric further validates its excellent performance in controlling overall errors and mitigating large discrepancies. In contrast, the LSTM and CNN models show higher error values and lower goodness of fit when dealing with extreme conditions and high-frequency fluctuations, making them less effective in accurately predicting fluctuations in drilling pump energy consumption. Moreover, the LSTM-Attention model achieves an R² value close to 1, indicating its strong capability to capture the fluctuation patterns in actual data and providing a much better explanation of energy consumption variations than traditional models. Overall, the LSTM-Attention model not only excels in prediction accuracy but also exhibits high robustness and adaptability, offering effective technical support for energy management and optimization in drilling operations.

4. Conclusions

This study proposed a novel LSTM-Attention model for accurately predicting drilling pump energy consumption. Through comprehensive experiments and comparisons with traditional LSTM and CNN models, the results demonstrated that the proposed model outperformed existing methods across multiple evaluation metrics. In the four test cases, the Mean Absolute Error (MAE) of the LSTM-Attention model ranged from 5.19 to 10.20, significantly lower than that of the traditional LSTM models (12.43 to 18.76) and CNN models (13.27 to 19.82). Additionally, the Mean Squared Error (MSE) values for the LSTM-Attention model, ranging from 45.02 to 66.42, were notably lower than those of the traditional LSTM models (130.67 to 180.91) and CNN models (140.85 to 190.74). The Root Mean Square Error (RMSE) ranged from 6.71 to 8.15, outperforming the traditional LSTM (11.43 to 13.45)

and CNN models (12.78 to 14.52). The R^2 values were close to one (ranging from 0.95 to 0.98), indicating the model's exceptional performance in capturing the trends of energy consumption variations.

The model's superior performance lies in its ability to handle both short-term fluctuations and long-term trends in energy consumption, which is crucial for optimizing drilling operations. Its robustness and adaptability under various operating conditions make it an ideal tool for real-time energy management and decision-making in drilling processes. The incorporation of the attention mechanism enhances the model's focus on critical features and time steps, further improving prediction accuracy and interpretability.

Although the LSTM-Attention model shows significant improvements in prediction performance compared to traditional models, there is still room for further enhancement. Future research could consider incorporating additional operational parameters and external factors to refine the model's predictions under more diverse and challenging conditions. Moreover, exploring more advanced deep learning architectures, such as Transformer models or hybrid approaches, may offer even greater predictive power and applicability.

Overall, the LSTM-Attention model provides a valuable framework for improving the accuracy and efficiency of drilling energy consumption forecasting. Its application can contribute to enhanced energy utilization, cost reduction, and improved operational safety, laying a solid foundation for intelligent and sustainable drilling practices in the oil and gas industry.

Author Contributions: Conceptualization, C.W.; Methodology, Z.Y.; Software, Z.Y.; Validation, C.W. and C.Z.; Formal analysis, C.W. and C.Z.; Investigation, Q.L.; Resources, Q.L. and Z.Z.; Data curation, Z.Y.; Writing—original draft, Z.Y.; Writing—review & editing, Z.Z.; Visualization, Q.L.; Supervision, C.Z.; Project administration, Z.Z. All authors have read and agreed to the published version of the manuscript.

Funding: The authors gratefully acknowledge the support of the National Science Fund for Distinguished Young Scholars (No. 52125401).

Data Availability Statement: The datasets presented in this article are not readily available because of company confidentiality requirements. Requests to access the datasets should be directed to corresponding author.

Conflicts of Interest: Author Chengcheng Wang was employed by the Karamay Drilling Company of CNPC Western Drilling Engineering Co., Ltd. The authors declare that this study did not receive any funding from Karamay Drilling Company of CNPC Western Drilling Engineering Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Gan, C.; Wang, X.; Wang, L.; Cao, W.; Liu, K.; Gao, H.; Wu, M. Multi-source information fusion-based dynamic model for online prediction of rate of penetration (ROP) in drilling process. *Geoenergy Sci. Eng.* **2023**, *230*, 212187. [[CrossRef](#)]
2. Guo, J.; Yang, Y.; Li, H.; Wang, J.; Tang, A.; Shan, D.; Huang, B. A hybrid deep learning model towards fault diagnosis of drilling pump. *Appl. Energy* **2024**, *372*, 123773. [[CrossRef](#)]
3. Li, G.; Hu, J.; Ding, Y.; Tang, A.; Ao, J.; Hu, D.; Liu, Y. A novel method for fault diagnosis of fluid end of drilling pump under complex working conditions. *Reliab. Eng. Syst. Saf.* **2024**, *248*, 110145. [[CrossRef](#)]
4. Guo, J.; Yang, Y.; Li, H.; Dai, L.; Huang, B. A parallel deep neural network for intelligent fault diagnosis of drilling pumps. *Eng. Appl. Artif. Intell.* **2024**, *133*, 108071. [[CrossRef](#)]
5. Yesilyurt, H.; Dokuz, Y.; Dokuz, A.S. Data-Driven Energy Consumption Prediction of a University Office Building using Machine Learning Algorithms. *Energy* **2024**, *310*, 133242. [[CrossRef](#)]
6. Zhang, Z.; Xiang, S.; Huo, H.; Xie, T.; Zhang, X.; Xu, J. Effect of pump start-stop operation in drilling of offshore Extreme-HPHT directional wells upon bottomhole transient fluctuating pressure. *Geoenergy Sci. Eng.* **2024**, *235*, 212708. [[CrossRef](#)]
7. Jiang, Y.; Guo, J.; Zhao, D.; Li, Y. Intelligent energy consumption prediction for battery electric vehicles: A hybrid approach integrating driving behavior and environmental factors. *Energy* **2024**, *308*, 132774. [[CrossRef](#)]
8. Yu, Y.; Zhang, C.; Wang, G.; Wang, X.; Yang, L. Short-term energy consumption prediction for university dormitory buildings based on energy consumption grouping by considering meteorological factors and teaching schedules. *Energy Build.* **2024**, *312*, 114219. [[CrossRef](#)]

9. Oh, J.; Min, D. Prediction of energy consumption for manufacturing small and medium-sized enterprises (SMEs) considering industry characteristics. *Energy* **2024**, *300*, 131621. [[CrossRef](#)]
10. Shaohu, L.; Yuandeng, W.; Rui, H. Prediction of drilling plug operation parameters based on incremental learning and CNN-LSTM. *Geoenergy Sci. Eng.* **2024**, *234*, 212631. [[CrossRef](#)]
11. Yehia, T.; Gasser, M.; Ebaid, H.; Meehan, N.; Okoroafor, E.R. Comparative analysis of machine learning techniques for predicting drilling rate of penetration (ROP) in geothermal wells: A case study of FORGE site. *Geothermics* **2024**, *121*, 103028. [[CrossRef](#)]
12. Sharifinasab, M.H.; Emami Niri, M.; Masroor, M. Developing GAN-boosted Artificial Neural Networks to model the rate of drilling bit penetration. *Appl. Soft Comput.* **2023**, *136*, 110067. [[CrossRef](#)]
13. Moraga, J.; Duzgun, H.S.; Cavur, M.; Soydan, H. The Geothermal Artificial Intelligence for geothermal exploration. *Renew. Energy* **2022**, *192*, 134–149. [[CrossRef](#)]
14. Liu, Z.; Yang, X. Insight of low flammability limit on sustainable aviation fuel blend and prediction by ANN model. *Energy AI* **2024**, *18*, 100423. [[CrossRef](#)]
15. Du, J.; Zeng, J.; Wang, H.; Ding, H.; Wang, H.; Bi, Y. Using acoustic emission technique for structural health monitoring of laminate composite: A novel CNN-LSTM framework. *Eng. Fract. Mech.* **2024**, *309*, 110447. [[CrossRef](#)]
16. Chee Min Tan, J.; Cao, Q.; Quek, C. FE-RNN: A fuzzy embedded recurrent neural network for improving interpretability of underlying neural network. *Inf. Sci.* **2024**, *663*, 120276. [[CrossRef](#)]
17. Long, J.; Ye, L.; Peng, H.; Tian, Z. Efficient prediction framework for large-scale nonlinear petrochemical process based on feature selection and temporal-attention LSTM: Applied to fluid catalytic cracking. *Chem. Eng. Sci.* **2025**, *301*, 120733. [[CrossRef](#)]
18. Zaidi, F.S.; Dai, H.; Imran, M.; Tran, K.P. Enhanced Hybrid LSTM and SLAR Modeling for In-Depth Analysis of Temporal and Spatial Patterns in Compositional Data for Environmental Monitoring. *Process Saf. Environ. Protect.* **2024**, *191*, 2013–2025. [[CrossRef](#)]
19. Zou, Z.; Yan, X.; Yuan, Y.; You, Z.; Chen, L. Attention mechanism enhanced LSTM networks for latency prediction in deterministic MEC networks. *Intell. Syst. Appl.* **2024**, *23*, 200425. [[CrossRef](#)]
20. Zhou, G.; Hu, G.; Zhang, D.; Zhang, Y. A novel algorithm system for wind power prediction based on RANSAC data screening and Seq2Seq-Attention-BiGRU model. *Energy* **2023**, *283*, 128986. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.