Prediction Model for Transient NOx Emission of Diesel Engine Based on CNN-LSTM Network

Qianqiao Shen 1, Guiyong Wang 1, *, Yuhua Wang 1, *, Boshun Zeng 1, Xuan Yu 1 and Shuchao He 2

1 Yunnan Province Key Laboratory of Internal Combustion Engines, Kunming University of Science and Technology, Kunming 650500, China; shenqianqiao@stu.kust.edu.cn (Q.S.); zengboshun@stu.kust.edu.cn (B.Z.); yuxuan@stu.kust.edu.cn (X.Y.)
2 Kunming Yunnei Power Co., Ltd., Kunming 650500, China; heshuchao.130@163.com
* Correspondence: wangguiyong@kust.edu.cn (G.W.); wangyuhua@stu.kust.edu.cn (Y.W.);
Tel.: +86-130-3334-5298 (G.W.)

Abstract: In order to address the challenge of accurately predicting nitrogen oxide (NOx) emission from diesel engines in transient operation using traditional neural network models, this study proposes a NOx emission forecasting model based on a hybrid neural network architecture combining the convolutional neural network (CNN) and long short-term memory (LSTM) neural network. The objective is to enhance calibration efficiency and reduce diesel engine emissions. The proposed model utilizes data collected under the thermal cycle according to the world harmonized transient cycle (WHTC) emission test standard for training and verifying the prediction model. The CNN is employed to extract features from the training data, while LSTM networks are used to fit the data, resulting in the precise prediction of training NOx emissions from diesel engines. Experimental verification was conducted and the results demonstrate that the fitting coefficient ($R^2$) of the CNN-LSTM network model in predicting transient NOx emissions from diesel engines is 0.977 with a root mean square error of 33.495. Compared to predictions made by a single LSTM neural network, CNN neural network predictions, and back-propagation (BP) neural network predictions, the root mean square error (RMSE) decreases by 35.6%, 50.8%, and 62.9%, respectively, while the fitting degree $R^2$ increases by 2.5%, 4.4%, and 6.6%. These results demonstrate that the CNN-LSTM network prediction model has higher accuracy, good convergence, and robustness.

Keywords: diesel engine; NOx prediction; world harmonized transient cycle; convolutional neural network; long short-term memory networks; grid search

1. Introduction

The diesel engine has become a preferred choice in heavy transportation and automobile industries due to its high efficiency and power output. However, the emissions produced by diesel engines during operation contribute to global environmental pollution [1]. In recent years, increasingly stringent emission regulations have posed significant challenges to controlling diesel engine emissions. Simply relying on in-machine purification technology is no longer sufficient to meet regulatory requirements, necessitating the use of various post-treatment equipment such as diesel oxidation catalytic (DOC), diesel particulate filter (DPF), selective catalytic reduction (SCR), and more [2]. However, the accurate control strategy for injecting the reducing agent in the SCR system relies on the precise knowledge of the original NOx emissions from the diesel engine. Currently, due to the high cost of NOx sensors, the original NOx emission map is primarily obtained through extensive calibration tests, which are time-consuming and require significant investments. Therefore, there is a need to explore a more convenient method for predicting NOx emissions in diesel engines [3].

To accurately predict NOx emissions from diesel engines, researchers, both domestically and internationally, have proposed methods based on physical models [4,5], as well
as a combination of physical models with MAP mapping [6,7]. While these prediction methods can effectively estimate NOx emissions under steady-state conditions, they face challenges in accurately predicting transient NOx emissions due to the rapid changes in diesel engine speed, torque, and fuel injection during transient conditions. The deterioration of in-cylinder combustion during these transient conditions affects pollutant emissions, posing difficulties for precise prediction using these methods.

In recent years, there has been a remarkable increase in the utilization of machine learning techniques to address cutting-edge challenges in various fields, driven by the wave of interdisciplinary research [8]. Among these techniques, the LSTM network stands out for its robust capability to tackle both long-term and short-term problems. It has demonstrated exceptional performance in predicting nonlinear time series data [9], which is particularly relevant in the case of diesel engine transient emission data, as it is also represented in the form of time series data. Consequently, the LSTM network has found a wide application in the prediction of diesel engine transient emissions. For instance, Yang et al. [10] and Dai Jinch et al. [11] have employed the LSTM network to forecast NOx emissions under transient conditions. Seunghyup et al. [12] utilized a Bayesian hyperparametric optimization deep neural network model to predict NOx emissions under transient conditions. Yang Rong et al. [13] employed a genetic algorithm to optimize the LSTM network to predict transient NOx emissions in diesel engines. While all of the aforementioned models demonstrate some ability to predict transient emissions in diesel engines, they fail to fully capture the spatial correlation characteristics among various control parameters, such as the speed, torque, and fuel injection control. As a result, the prediction accuracy of these models under transient working conditions is compromised.

In light of this limitation, the present study proposes the utilization of a CNN. The CNN has achieved significant advancements in various domains such as image processing, data processing, air pollutant prediction, and power system load prediction [14–17]. It has also garnered considerable attention in the prediction of diesel engine emissions [18]. The CNN network structure possesses three key characteristics: local connection, weight sharing, and pooling [19]. These properties grant the network a certain level of invariance to translation, scaling, and rotation, enabling it to capture the spatial characteristics of data [20]. Consequently, when confronted with the spatial correlation among diesel engine control parameters, the CNN can effectively extract relevant feature information. However, relying solely on the spatial characteristics extracted by the CNN is insufficient to address the prediction challenges associated with diesel engine transient emissions, which necessitate the consideration of both temporal and spatial series.

Based on the above problems, to enhance the accuracy of predicting NOx emissions from diesel engines in transient environments, a method that combines CNN with LSTM is proposed. This approach establishes a diesel engine NOx emission prediction model known as CNN-LSTM that is specifically designed for transient working conditions. By harnessing the spatial data extraction capabilities of the CNN, this model generates a plethora of valuable inputs that effectively complement the LSTM network model [21], enabling a comprehensive consideration of transient emission data from diesel engines.

2. Experimental Section and Method
2.1. Experimental Equipment

The test was conducted on a supercharged in-line 4-cylinder electronically controlled high-pressure common rail diesel engine, which complies with the national emission standards. Table 1 presents the key technical parameters of the engine. The test employed several essential instruments and equipment, including the AVL PUMA measurement and control system, AVL electric dynamometer, AVL AMA i60 exhaust measurement system, AVL FTIR i60 exhaust measurement system, 553 coolant temperature control system, and 735 fuel consumption meter. The layout and physical configuration of the test bench can be observed in Figures 1 and 2, respectively.
Table 1. Engine specifications.

<table>
<thead>
<tr>
<th>Parameter Description</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>D25TCIF</td>
</tr>
<tr>
<td>Engine type</td>
<td>Electronically controlled high-voltage common rail and In-line four</td>
</tr>
<tr>
<td>Bore × stroke</td>
<td>92 mm × 94 mm</td>
</tr>
<tr>
<td>Displacement</td>
<td>2.499 L</td>
</tr>
<tr>
<td>Rated speed</td>
<td>3000 rpm</td>
</tr>
<tr>
<td>Maximum torque</td>
<td>450 N·m</td>
</tr>
<tr>
<td>Maximum power</td>
<td>120 kw</td>
</tr>
<tr>
<td>Maximum injection pressure</td>
<td>1,600,000 hpa</td>
</tr>
<tr>
<td>Fuel injector flow rate</td>
<td>478 (cm³/30 s)</td>
</tr>
<tr>
<td>Nozzle diameter</td>
<td>0.135 mm</td>
</tr>
<tr>
<td>K coefficient</td>
<td>1.3</td>
</tr>
</tbody>
</table>

Figure 1. Layout of test bench.

Figure 2. The practicality picture of test bench.
2.2. Experimental Scheme

With the promulgation of nation VI emission regulations, it has become imperative to calibrate the hot and cold WHTC tailpipes of the diesel engine to comply with the emission limit. In order to enhance development efficiency, the calibration is primarily focused on the pure hot WHT. Therefore, this paper chooses the hot cycle within the WHTC test cycle as the testing condition for the proposed test system.

The WHTC test cycle is a test cycle proposed by Europe for Euro-VI emission standards. This test cycle takes into full consideration the road conditions worldwide and the driving characteristics of different vehicles. It consists of three main components: the cold start emission test, the hot dip emission test, and the hot start emission test. The cold start cycle and hot start cycle have a duration of 1800 s each, and their operating conditions are defined by a set of standard percentages of speed and torque that change every second. Figure 3 illustrates that, upon completion of the cold start test, a hot dip procedure lasting 10 ± 1 min should be immediately conducted as the engine’s hot start test pretreatment. The cold start test contributes to 14% of the final emission results, while the hot start test accounts for the remaining 86% [22].

According to the WHTC program’s cycle condition, which consists of the last 600 s with a tolerance of plus or minus 10 s, the hot cycle condition within this program is selected for testing purposes. Once the hot soak period of the diesel engine is completed, the bench WHTC cycle program is initiated to conduct the official hot cycle test. During this test, 11 relevant parameters are collected and recorded every second. Among these parameters, NOx emission is chosen as the output parameter for the prediction model. The input parameters for research and analysis include speed, torque, fuel pressure, fuel temperature, intake flow, pre-injection timing, pre-injection quantity, total fuel injection quantity, atmospheric temperature, and atmospheric humidity. These parameters serve as the basis for further investigation and analysis. Table 2 displays some of the thermal cycle data obtained from the testing process.
Table 2. Partial thermal state cycle data overview.

<table>
<thead>
<tr>
<th>Parameter Description</th>
<th>Title 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sampling time/s</td>
<td></td>
</tr>
<tr>
<td>Rotation speed/(tr min⁻¹)</td>
<td>804.6</td>
</tr>
<tr>
<td>Torque/(N·m)</td>
<td>4.690</td>
</tr>
<tr>
<td>Intake air flow/(kg·h⁻¹)</td>
<td>50.06</td>
</tr>
<tr>
<td>Pre-injection timing/°CA</td>
<td>4.329</td>
</tr>
<tr>
<td>Pre-injection quantity/(kg·h⁻¹)</td>
<td>1.600</td>
</tr>
<tr>
<td>Total fuel injection</td>
<td></td>
</tr>
<tr>
<td>Pre-injection</td>
<td></td>
</tr>
<tr>
<td>Fuel pressure/bar</td>
<td>4.590</td>
</tr>
<tr>
<td>Atmospheric</td>
<td></td>
</tr>
<tr>
<td>temperature/°C</td>
<td>24.630</td>
</tr>
<tr>
<td>Atmospheric humidity/%</td>
<td>51.582</td>
</tr>
<tr>
<td>Fuel temperature/°C</td>
<td>29.300</td>
</tr>
<tr>
<td>NOx value/ppm</td>
<td>45.7</td>
</tr>
</tbody>
</table>

3. Data Preprocessing

3.1. Data Correlation Analysis

Due to the excessive number of total sample input parameters recorded during the initial collection, and the lack of significant correlation between some parameters and the generation of NOx emissions, it becomes necessary to analyze each of the aforementioned parameters individually. By eliminating parameters with low correlation, we can effectively reduce the model’s dimensionality and enhance its accuracy.

The Spearman correlation coefficient and Pearson correlation coefficient are employed to analyze the pre-selected input parameters separately. The Spearman correlation coefficient is utilized to evaluate the degree of nonlinear correlation between parameters, while the Pearson correlation coefficient is utilized to assess the linear correlation between parameters [13]. The calculation equation for the Spearman correlation coefficient is presented in Equation (1) [23]:

\[
\rho = \frac{\sum_{i=1}^{n} x_i^2 - \frac{1}{2} \sum_{i=1}^{n} (x_i - y_i)^2}{\sqrt{\sum_{i=1}^{n} x_i^2 - \frac{1}{2} \sum_{i=1}^{n} x_i^2} \sqrt{\sum_{i=1}^{n} y_i^2 - \frac{1}{2} \sum_{i=1}^{n} y_i^2}} = 1 - \frac{6 \sum_{i=1}^{n} (x_i - y_i)^2}{n(n^2-1)} = 1 - \frac{6 \sum_{i=1}^{n} d_i^2}{n(n^2-1)}
\]

where \( n \) is the total sample number, \( d_i^2 \) is the rank difference of two variables after sorting, and \( d_i = (x_i - y_i) \), \( x_i \) is the corresponding input parameter, \( y_i \) is the emission value of NOx. The calculation equation of the correlation coefficient of Pearson was shown in Equation (2) [24]:

\[
R = \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{n \sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{\sqrt{n \sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2} \sqrt{n \sum_{i=1}^{n} y_i^2 - (\sum_{i=1}^{n} y_i)^2}}
\]

where \( x_i \) is the corresponding input parameter, \( y_i \) is the emission value of NOx, \( \text{cov}(X,Y) \) measures the covariance of two sets of data \( X \) and \( Y \), and \( \sigma_X \) and \( \sigma_Y \) are the standard deviations of \( X \) and \( Y \).

The results of the correlation coefficient analysis between each pre-selected parameter and NOx emissions are presented in Table 3. A positive correlation coefficient value indicates a positive correlation between the two parameters, suggesting that their changing trends are in the same direction. Conversely, a negative correlation coefficient value indicates a negative correlation between the two parameters, indicating that their changing
trends are the opposite. Moreover, when the absolute value of the correlation coefficient is closer to 1, it signifies a stronger correlation and a greater influence relationship between the parameters. Conversely, when the absolute value of the correlation coefficient is closer to 0, it indicates a weaker correlation and a smaller influence relationship [25–27].

Table 3. Correlation analysis results between input parameters and NOx emission values.

<table>
<thead>
<tr>
<th>Parameter Description</th>
<th>Spearman</th>
<th>Pearson</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rotation time/s</td>
<td>0.4</td>
<td>0.42</td>
</tr>
<tr>
<td>Torque/(N·m)</td>
<td>0.42</td>
<td>0.48</td>
</tr>
<tr>
<td>Intake air flow/(kg·h(^{-1}))</td>
<td>0.53</td>
<td>0.48</td>
</tr>
<tr>
<td>Pre-injection timing/°CA</td>
<td>0.43</td>
<td>0.45</td>
</tr>
<tr>
<td>Pre-injection quantity/(kg·h(^{-1}))</td>
<td>0.35</td>
<td>0.14</td>
</tr>
<tr>
<td>Total fuel injection quantity/(kg·h(^{-1}))</td>
<td>0.44</td>
<td>0.49</td>
</tr>
<tr>
<td>Fuel pressure/bar</td>
<td>−0.41</td>
<td>−0.42</td>
</tr>
<tr>
<td>Atmospheric temperature/°C</td>
<td>−0.22</td>
<td>−0.22</td>
</tr>
<tr>
<td>Atmospheric humidity/%</td>
<td>0.15</td>
<td>0.14</td>
</tr>
<tr>
<td>Fuel temperature/°C</td>
<td>−0.12</td>
<td>−0.15</td>
</tr>
</tbody>
</table>

Based on the analysis in Table 3, it is observed that the Pearson correlation coefficient and Spearman correlation coefficient of fuel temperature, atmospheric temperature, and atmospheric humidity with NOx emissions are small. Hence, the correlation between these three pre-selected parameters and NOx emissions is minimal and they can be eliminated. Additionally, the table reveals that the Pearson correlation coefficient for the pre-injection quantity is 0.14, but it increases to 0.35 in the Spearman correlation coefficient. This indicates that while the linear correlation between the pre-injection quantity and NOx emissions is relatively small, the degree of nonlinear correlation remains significant. Therefore, the pre-injection quantity is included as one of the input variables.

To summarize, this study excludes only three variables (fuel temperature, atmospheric temperature, and atmospheric humidity) from the pre-selected input parameters, while retaining the remaining input parameters.

3.2. Data Normalization Processing

Once the input and output parameters of the model have been determined, it is crucial to normalize the data to address potential issues arising from significant differences in data orders between the input and output parameters. By normalizing the data and mapping the values to the range of [0, 1], we can prevent excessive prediction errors and facilitate faster convergence of the model. The calculation equation for normalization is provided in Equation (3):

\[
x' = \frac{x - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}} \tag{3}
\]

where \( x \) is the value of each input parameter.

4. Construction of Emission Prediction Model Based on CNN-LSTM

4.1. CNN Neural Network

The CNN is a type of deep neural network that incorporates convolutional structures. It is primarily composed of a convolution layer, pooling layer, and fully connected layer [28]. The convolution layer is responsible for feature extraction, followed by the pooling layer, which reduces the parameter dimension and improves the training efficiency by transmitting data information to the next layer in the network. Finally, the results are output through linear transformation in the fully connected layer.

Different convolutional dimensions are utilized in CNNs for various processing domains. One-dimensional convolutional neural networks (1D-CNN) are employed for processing one-dimensional and two-dimensional data or images. Two-dimensional convolutional neural networks (2D-CNN) are mainly used for image classification tasks, while
three-dimensional convolutional neural networks (3D-CNN) are predominantly applied in video processing and the detection of actions and behaviors of individuals [29]. The structure of a CNN is illustrated in Figure 4.

![Convolutional neural network structure diagram.](image)

**Figure 4.** Convolutional neural network structure diagram.

The 1D-CNN utilizes matrix multiplication to perform convolution calculations on time series data and it maps data variables to a high-dimensional space and extracts local features based on spatial and time series correlations. During the data processing, the convolution kernel of the 1D-CNN can only move in the horizontal or vertical direction of the data. In the case of time series data, the convolution kernel slides along the time series direction, making it particularly suitable for processing time series data recorded by sensors. It is also well-suited for analyzing various types of signal data within a fixed length of time.

Since transient NOx emission data from diesel engines involves time series emissions recorded by sensors, the CNN can effectively extract characteristics from the emission data and enhance the prediction accuracy of the model. The calculation equation for a one-dimensional convolution is shown in Equation (4) [30]:

$$k^n_m = f \left( \sum_{l=1}^{N} k^{n-1}_l * w^n_{lm} + b^n_m \right)$$  \hspace{1cm} (4)

where $k^n_m$ is the $m$th feature map of layer $n$, $f(.)$ is an activation function, $N$ is the input feature size, * is the convolution operation between the $l$th feature map of the former layer $[l-1]$th layer] and the convolution kernel $w^n_{lm}$ and $b^n_m$ is the corresponding bias.

To enhance the fitting ability and sparsity of the CNN, the ReLU function has been chosen as the activation function. When compared to the Sigmoid and tanh function, the ReLU function effectively addresses the issues of gradient disappearance and slow convergence. Its calculation equation for the ReLU function is shown in Equation (5):

$$R_{ReLU}(a) = \begin{cases} a, & a > 0 \\ 0, & a \leq 0 \end{cases}$$ \hspace{1cm} (5)

where $a$ is the value obtained after the convolution operation.

The pooling layer serves the purpose of data and parameter compression, bit reduction, and addressing overfitting issues. It performs downsampling operations, which enhance computation speed and the resilience of extracted features. Additionally, it diminishes redundant features while preserving the key characteristics of NOx emissions from diesel engines. The pooling operation consists of two types: maximum pooling and average pooling. The calculation equation for the pooling operation is shown in Equation (6):

$$p(i, j) = \frac{1}{s^2} \sum_{u=(i-1)s+1}^{is} \sum_{v=(j-1)s+1}^{js} a(u, v)$$ \hspace{1cm} (6)

where $p(i, j)$ is the value of the $i$th row in the $j$th column of the pooling layer output matrix, $a(u, v)$ is the value of the $u$th row in the $v$th column of the pooling layer input matrix, and $s$ is the boundary value of the region participating in the set.
The data, post-convolution and pooling, is fed into the fully connected layer. Depending on whether the task is regression or classification, different activation functions are employed to produce the final output. Its calculation equation for the fully connected layer is shown in Equation (7):

\[ y = Wx_f + B \]  

(7)

where \( y \) is the output value of the fully connected layer, \( x_f \) is the input value of the fully connected layer, \( W \) is the weigh matrix, and \( B \) is the bias vector.

4.2. LSTM Neural Network

The LSTM network is a specialized type of recurrent neural network (RNN) commonly employed to address the issues of the gradient vanishing or exploding during prolonged information transmission [31]. Unlike the RNN, the LSTM network incorporates a more intricate neuron structure within the hidden layer. It introduces a cell state to retain long-term information and utilizes three control mechanisms: the input gate, forgetting gate, and output gate to regulate the state. Each LSTM module consists of a storage unit and three control gates, as illustrated in Figure 5, representing the fundamental building block of the neural network [32,33].

![Long-short-term memory neural network structure diagram.](image)

The red dotted box in the figure shows the distinctive structure of the forget gate, which plays a crucial role in determining the portion of the cell case that should be forgotten from the previous time step. The calculation formula for the forget gate is shown in Equation (8):

\[ f_t = \delta \left( W_f[h_{t-1}, x_t] + b_f \right) \]  

(8)

where \( f_t \) is the value of forget gate, \( \delta \) is the Sigmoid function, \( W_f \) is the weight of the forget gate, \( h_{t-1} \) is the implied unit of the \((t-1)\text{th}\) moment, \( x_t \) is the input data of the \(t\text{th}\) moment, and \( b_f \) is the bias of the forget gate.

The blue-dashed box in the figure depicts the precise structure of the input gate, which is responsible for determining the portion of the network input that should be preserved at the current time step. The calculation formula for the input gate is shown in Equations (9) and (10):

\[ i_t = \delta(W_i[h_{t-1}, x_t] + R_i) \]  

(9)
\[ g_t = \tanh(W_g[h_{t-1}, x_t] + b_g) \]  

where \( i_t \) is the value of input gate, \( W_i \) is the weight of the input gate, \( R_i \) is the offset term of the input gate, \( g_t \) is the input node, \( W_g \) is the weight of the input node, and \( b_g \) is the bias of the input node.

The green-dashed box in the figure represents the specific structure for updating the cell case. It operates based on the combined influence of the forget gate and input gate, enabling it to retain relevant information from the distant past while discarding irrelevant or invalid information that should not be propagated through the network. The calculation formula for updating the cell case is shown in Equation (11):

\[ c_t = f_t \times c_{t-1} + i_t \times g_t \]  

where \( c_t \) is the cell case of the \( t \)th moment, and \( c_{t-1} \) is the cell case of the \((t - 1)\)th moment.

The purple-dotted frame in the figure shows the specific structure of the output gate, which plays a crucial role in determining the impact of long-term memory on the current output and updating the hidden unit. The calculation formula for the output gate is shown in Equations (12) and (13):

\[ o_t = \delta(W_o[h_{t-1}, x_t] + b_o) \]  

\[ h_t = o_t \times \tanh(c_t) \]  

where \( o_t \) is the value of output gate, \( W_o \) is the weight of the output gate, \( b_o \) is the bias of the output gate, and \( h_t \) is the output value of the \( t \)th moment.

The LSTM network excels at preserving the distinctive traits found in long time series data and possesses the capacity for long-term memory. Leveraging its capabilities through sequence learning and feature training, it proves advantageous in enhancing the accuracy of predicting transient NOx emissions in diesel engines.

### 4.3. CNN-LSTM Neural Network Prediction Model

The LSTM network prediction model is employed to effectively model time series data, incorporating high-dimensional feature information extracted by the CNN. By capturing the temporal patterns within these features, the LSTM network model enables the accurate representation of the nonlinear dynamics associated with transient emission in diesel engines. Consequently, this approach enhances the prediction accuracy of NOx emissions in diesel engine transient environments.

The CNN-LSTM network prediction model is typically divided into two components: the CNN’s feature extraction module and the LSTM network’s time series prediction module. The first part focuses on extracting spatial feature information from preprocessed data related to diesel engine parameters and emissions. This extracted feature information serves as the input for the LSTM network model. The second part employs the LSTM network for its ability to maintain long-term memory, enabling the accurate extraction of time series characteristics from the data. Consequently, the model can effectively predict transient NOx emissions in diesel engines.

#### 4.3.1. Determination of Structural Parameters of CNN-LSTM Neural Network

In the process of debugging the CNN structure, it has been observed that when the number of convolution layers is too small, the model may suffer from underfitting due to insufficient feature extraction capabilities. Conversely, an excessive number of convolution layers can lead to overfitting. While the pool layer can mitigate overfitting, employing too many pool layers results in a reduced number of feature dimensions being fed into the LSTM network. This reduction can adversely affect the extraction of time series features by LSTM, consequently diminishing the effectiveness of network fitting. After numerous rounds of debugging, a network structure comprising three convolution layers, one pool
layer, and one flat layer has been ultimately selected to successfully predict the transient NOx emission of diesel engines.

The Adam optimizer is utilized to automate the updating of the weight matrix and bias of the LSTM network model, as well as adaptively adjust the learning rate throughout the training process. A grid search is employed to swiftly optimize parameters, including model depth \(N_l\), the number of neurons in hidden layers \(N_u\), and the batch size \(B_c\), for the LSTM network prediction. The optimized parameters have proven to significantly enhance model performance and improve prediction accuracy.

4.3.2. Optimization of Super-Parameter of Prediction Model by Grid Search Method

The optimization of neural network hyperparameters through the grid search method involves an exhaustive exploration of the hyperparameter space subset of the algorithm [34]. This approach divides the search range into a grid and systematically examines all intersections within it. By evaluating the feedback results from these intersections, the best combination of hyperparameters can be determined. This process provides relatively optimal modeling parameters for the prediction module in the CNN-LSTM network model. The main steps for optimizing the CNN-LSTM network prediction model using the grid search method are outlined below:

1. Determine the search range for the hyperparameters in the LSTM network prediction module of this model. Pass this range to the grid search function, which will organize all possible combinations within the specified range. Define the search range for the hyperparameters as follows: model depth \(N_l = (1:10)\); The number of neurons in the hidden layer \(N_u = (10:200)\); Batch size \(B_c = (10:100)\);

2. Different CNN-LSTM neural network prediction models are constructed based on each parameter combination;

3. The loss function is defined to evaluate the performance of model parameters, and the mean square error (MSE) is adopted as the chosen loss function. The calculation formula for updating the MSE is shown in Equation (14):

\[
MSE = \frac{1}{n} \sum_{i=1}^{n} (y^i_a - y^i_p)^2
\]  

(14)

where \(n\) is the total sample number, \(y^i_a\) is the true value, and \(y^i_p\) is the predicted value.

4. Set the number of network epochs iterations to 50, and obtain the final value of the loss function for each prediction model after the network training reaches the maximum learning iteration;

5. The optimal solution with the minimum loss function value is selected to determine the optimal hyperparameter combination for the CNN-LSTM network prediction model.

By continuously iterating and optimizing the hyperparameter combinations of the prediction model using the grid search method, and performing training under different hyperparameter combinations, the final optimal hyperparameter combination is obtained as follows: \(N_l = 2\); \(N_u = 20\); \(B_c = 60\). The framework of the transient NOx emission prediction model for diesel engines, based on the CNN-LSTM network optimized using the grid search method, is depicted in Figure 6.
5. Training, Verification, and Comparison of Forecasting Model

5.1. Training and Verification of Prediction Model

After data preprocessing, the WHTC thermal cycle dataset consisting of 1800 data points is divided into a training set and a validation set using an 8:2 ratio. The training set includes seven input features and one output label, which are fed into a CNN and convolved three times. Following the convolutional operations, a ReLU activation function is applied to map the features to high-dimensional nonlinear intervals, preventing overfitting. Subsequently, a one-layer maximum pooling layer is used to reduce the output dimension. The number of convolution kernels is set to 32, 64, and 128 sequentially, with convolution and pooling kernel sizes set to $1 \times 3$ and a stride of 1 for both the convolutional and pooling layers. After three consecutive convolutions and a maximum pooling operation, a feature matrix of size $128 \times 16$ is obtained. This matrix is then flattened into a one-dimensional vector of length 2048, serving as the global feature extraction for the LSTM network. The feature extraction process of the 1D-CNN is illustrated in Figure 7.
To ensure the accuracy of the model, the optimal hyperparameter combination obtained through the grid search is used as an input for the LSTM network. The final network structure consists of one input layer, two hidden layers (each with 20 neurons), one output layer, and one fully connected layer. The mean square error (MSE) is chosen as the loss function for fitting and predicting the transient NOx emission data of diesel engines. The LSTM network iteratively trains the input gate, forgetting gate, and output gate to adjust their respective parameters. The feature vectors extracted by the CNN are trained, and the weights of the neural network are updated iteratively using the Adam algorithm. The initial learning rate is set to 0.001, and the weights and biases of each neuron are continually updated using the momentum and adaptive learning rate, resulting in an optimized output from the loss function [35].

After 50 iterations of training and validation with the first group of data, the optimal model is obtained. Finally, the prediction dataset is inputted into the optimal model to predict a new transient NOx emission value for the diesel engine. The loss trend of the training set and validation set of the model is depicted in Figure 8. From the curve in the figure, it can be observed that the loss values of the training set and validation set generally decrease with oscillations as the number of iterations increases. The training results demonstrate that the model converges well and achieves good training performance without overfitting.

5.2. Model Prediction Evaluation Index

To evaluate the performance of the prediction model, four evaluation metrics will be utilized: the mean absolute error (MAE), root mean square error (RMSE), mean absolute percentage error (MAPE), and fitting coefficient ($R^2$). These metrics provide a comprehensive assessment of the model’s performance. The calculation formulas for these four evaluation metrics are shown in Equations (15)–(18):

\[
MAE = \frac{1}{n} \sum_{i=1}^{n} |y_a^i - y_p^i| \tag{15}
\]

\[
RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_a^i - y_p^i)^2} \tag{16}
\]

\[
MAPE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{y_a^i - y_p^i}{y_a^i} \right| \times 100\% \tag{17}
\]
\[ R^2 = 1 - \frac{\sum_{i=1}^{n} (y_a^i - y_p^i)^2}{\sum_{i=1}^{n} (y_a^i - \bar{y})^2} \]

where \( n \) is the total sample number, \( y_a^i \) is the true value, \( y_p^i \) is the predicted value, and \( \bar{y} \) is the average of the actual responses.

---

**Figure 8.** Trend diagram of loss value of CNN-LSTM model.

**5.3. Comparison of Model Prediction**

To compare the advantages of the CNN-LSTM network prediction model in predicting NOx emissions under transient working conditions of a diesel engine, new NOx emission data is collected during the WHTC thermal cycle test as the prediction dataset. This dataset is then compared with the predictions from the LSTM network prediction model, CNN prediction model, and the grid search-optimized BP neural network prediction model. The network structure design used for this comparison is as follows:

1. The structure of the CNN prediction model optimized by grid search consists of three convolutional layers, one maximum pooling layer, and two fully connected layers. The number of convolutional kernels is set to 16, 32, and 64, with a kernel size and pooled kernel size of \( 1 \times 3 \). The pooled characteristic data is then fitted through the fully connected layers. To prevent overfitting, the ReLU activation function is employed. This configuration enables the prediction of transient NOx emissions in diesel engines;

2. The LSTM network prediction model optimized by grid search is a network structure composed of one input layer, two hidden layers (each containing 64 neurons), one output layer, and one fully connected layer. The MSE is utilized as the loss function in order to predict the transient NOx emissions of diesel engines;

3. The structure of the BP neural network prediction model optimized by the grid search consists of one input layer, eleven hidden layers, and one output layer. Each hidden layer is comprised of 35 neurons. This network configuration, which utilizes the mean square error as the loss function, enables the prediction of transient NOx emissions in diesel engines.

The training curves of each neural network prediction model are depicted in Figure 9. Based on the four different colored curves, it can be observed that the training loss of each prediction model generally exhibits a downward trend with oscillations as the number of iterations increases. After approximately 20 iterations, the CNN-LSTM, CNN, and...
LSTM network prediction models tend to stabilize, while the BP neural network prediction model tends to stabilize after around 40 iterations. This indicates that all four neural network prediction model structures converge without overfitting. Furthermore, it is noteworthy that the CNN-LSTM network model demonstrates a relatively fast convergence speed compared to the other three prediction models, second only to the CNN model. Additionally, once reaching a stable state, the model exhibits relatively low loss values, suggesting that the prediction model is more robust compared to the LSTM and CNN network prediction models.

![Train_loss change of different model.](image)

The final prediction results of each neural network prediction model are presented in Figure 10. Figure 10a–d demonstrate that all four neural network models effectively predict the trend of NOx emission values during the thermal cycle of a diesel engine. However, when it comes to extreme values where NOx emissions experience significant changes, the prediction values of the CNN-LSTM network prediction model are closer to the actual values compared to the other three neural network prediction models. This indicates that the CNN-LSTM network model performs better in predicting NOx emissions under transient and changing conditions, showcasing its robustness and adaptability. On the other hand, the fitting effects of the CNN, LSTM network, and BP neural network models are slightly inferior. This can be attributed to the fact that these three neural networks are only sensitive to either spatial characteristics or time series characteristics individually. Consequently, the degree of fitting for NOx emissions under transient thermal cycles is insufficient for these three neural network prediction models.

Table 4 and Figure 11 display the prediction errors of the four models. In comparison to the CNN, LSTM network, and BP neural network, the CNN-LSTM network exhibits significantly smaller prediction errors and higher accuracy. When comparing the prediction results of the CNN-LSTM network model to the other three neural network prediction models, the MAE, RMSE, and $R^2$ values are 23.981, 33.495, and 0.977, respectively. Furthermore, when compared to the LSTM network prediction model, the CNN-LSTM network model showcases a 18.9% reduction in MAE, a 35.6% decrease in RMSE, and a 2.5% increase in $R^2$. In comparison to the CNN prediction model, the CNN-LSTM network model exhibits a 43.7% decrease in MAE, a 50.8% decrease in RMSE, and a 4.4% increase in $R^2$. Similarly, when compared to the BP neural network prediction model, the CNN-LSTM network model shows a 43.1% reduction in MAE, a 62.9% decrease in RMSE, and a 6.6% increase in $R^2$. These results highlight the improved prediction accuracy and the model’s ability to fit transient NOx emissions. It indicates that the CNN-LSTM network can effectively explore the relationships among variables in transient NOx emissions of a diesel engine.
and extract crucial time-series characteristic information from historical NOx emission data, thus exhibiting robust learning capabilities.

Table 4. Models prediction accuracy.

<table>
<thead>
<tr>
<th>Neural Network Model</th>
<th>MAE</th>
<th>RMSE</th>
<th>MAPE</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNN-LSTM</td>
<td>23.981</td>
<td>33.495</td>
<td>18.4</td>
<td>0.977</td>
</tr>
<tr>
<td>LSTM</td>
<td>28.982</td>
<td>48.023</td>
<td>24.0</td>
<td>0.953</td>
</tr>
<tr>
<td>CNN</td>
<td>37.380</td>
<td>56.287</td>
<td>41.5</td>
<td>0.935</td>
</tr>
<tr>
<td>BP</td>
<td>37.156</td>
<td>64.208</td>
<td>28.2</td>
<td>0.915</td>
</tr>
</tbody>
</table>

Figure 10. Cont.
Prediction model of CNN-LSTM network. (b) Prediction model of CNN. (c) Prediction model of LSTM network. (d) Prediction model of BP neural network.

Figure 10. Prediction comparison chart of different models. (a) Prediction model of CNN-LSTM network. (b) Prediction model of CNN. (c) Prediction model of LSTM network. (d) Prediction model of BP neural network.

Figure 12 illustrates the regression accuracy of each model in the prediction set. It can be observed that the deviations of the prediction results for all four models are randomly distributed on both sides of the regression line, which aligns with the random distribution of experimental errors. In terms of the LSTM network prediction model, a few points with significant deviations can be observed at low NOx emissions, indicating average prediction accuracy of the model in that range. Both the CNN prediction model and the BP neural network prediction model exhibit large deviation points throughout the entire emission range, with a high dispersion degree. This suggests that the prediction precision of these models for the overall emission cycle is low. However, in the CNN-LSTM network prediction model, the deviation distribution for each segment of the emission prediction is relatively uniform, without any points displaying a wide range of deviations. Additionally, the dispersion degree is low. This indicates that the model addresses the problem of poor prediction accuracy encountered by the aforementioned three neural network models at
specific points. Moreover, it demonstrates that the CNN-LSTM network prediction model possesses higher nonlinear fitting and prediction accuracy compared to the CNN, LSTM network, and BP neural network prediction models.

![Figure 11. Prediction accuracy chart of different models.](image1)

![Figure 12. Regression prediction chart of different models.](image2)

(a) Model regression verification of CNN-LSTM. (b) Model regression verification of LSTM. (c) Model regression verification of CNN. (d) Model regression verification of BP.
6. Conclusions

In order to address the current issue of low accuracy in predicting transient NOx emissions of diesel engines, this paper proposes a prediction model, namely the CNN-LSTM network, which combines the CNN with LSTM network. Based on the validation of experimental data, the following conclusions have been drawn:

1. The CNN-LSTM network prediction model combines the powerful memory capability of the LSTM network in time series prediction with the CNN’s ability to extract deep features from the transient NOx emission data of diesel engines. This integration enables the model to uncover the intricate relationship between the characteristics of transient NOx emission data and enhance the accuracy of predicting diesel engine transient NOx emissions;

2. The CNN-LSTM network prediction model exhibits faster convergence speed and lower training loss compared to the LSTM network and BP neural network prediction models. However, it falls slightly short when compared to the CNN prediction model. Nonetheless, this prediction model demonstrates greater robustness when compared to the LSTM network and BP neural network prediction models;

3. Compared to the LSTM network prediction model, the CNN-LSTM network prediction model exhibits significant improvements in accuracy metrics. Specifically, it demonstrates a reduction in MAE and RMSE by 18.9% and 35.6%, respectively, while $R^2$ increases by 2.5%. When compared to the CNN neural network prediction model, the MAE and RMSE decrease by 43.7% and 50.8% respectively, and $R^2$ increases by 4.4%. In comparison to the BP neural network prediction model, the MAE and RMSE decrease by 43.1% and 62.9%, respectively, while $R^2$ increases by 6.6%. These results unequivocally highlight the superior accuracy of the CNN-LSTM network prediction model in forecasting transient NOx emissions from diesel engines;

4. Compared to CNN, LSTM network, and BP neural network prediction models, the CNN-LSTM network prediction model exhibits a notable absence of significant deviation points in the prediction of the NOx emission for each concentration under transient thermal cycle conditions of diesel engines. It demonstrates low dispersion, high prediction precision, and improved fitting effect.

Author Contributions: Conceptualization, G.W., Y.W. and Q.S.; methodology, Q.S.; software, Q.S.; validation, Q.S., G.W. and Y.W.; formal analysis, Q.S.; investigation, Q.S., B.Z. and X.Y.; resources, S.H.; data curation, S.H.; writing—original draft preparation, Q.S.; writing—review and editing, Q.S., G.W. and Y.W.; visualization, B.Z. and X.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Major Science and Technology Special Plan of Yunnan Provincial Science and Technology Department, grant number (202102AC080004), funder: G.W.; Key R&D projects of Yunnan Provincial Department of Science and Technology, grant number (202103AA080002), funder: G.W.

Data Availability Statement: The study did not report any data.

Conflicts of Interest: The authors declare no conflict of interest.

Nomenclature

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNN</td>
<td>Convolutional neural network</td>
</tr>
<tr>
<td>LSTM</td>
<td>Long short-term neural network</td>
</tr>
<tr>
<td>CNN-LSTM</td>
<td>Convolutional neural network-long short-term memory networks</td>
</tr>
<tr>
<td>BP</td>
<td>Back propagation</td>
</tr>
<tr>
<td>NOx</td>
<td>Nitrogen oxides</td>
</tr>
<tr>
<td>WHTC</td>
<td>World harmonized transient cycle</td>
</tr>
<tr>
<td>$R^2$</td>
<td>R-Square</td>
</tr>
<tr>
<td>RMSE</td>
<td>Root mean square error</td>
</tr>
<tr>
<td>MAE</td>
<td>Mean absolute error</td>
</tr>
</tbody>
</table>
MSE  Mean square error  
MAPE Mean absolute percentage error  
DOC Diesel oxidation catalytic  
DPF Diesel particulate filter  
SCR Selective catalytic reduction  
1D-CNN One-dimensional convolutional neural network  
2D-CNN Two-dimensional convolutional neural network  
3D-CNN Three-dimensional convolutional neural network  
RNN Recurrent neural network  
ReLU Rectified Linear Units  
tanh Hyperbolic tangent  
n Total sample number  
d_i Rank difference of two variables after sorting  
x_i Corresponding input parameter  
y_i Emission value of NOx  
ρ Spearman correlation coefficient  
R Pearson correlation coefficient  
cov Covariance  
σ Standard deviation  
f(·) Activation function  
N Input feature size  
b Correspond bias  
* Convolution operation  
k Feature map  
a Value obtained after convolution operation  
p Value of the pooling layer output matrix  
α Value of the pooling layer input matrix  
s Boundary value of the region participating in the set  
y Output value of the fully connected layer  
x_f Input value of the fully connected layer  
W Weigh matrix  
B Bias vector  
f_t Value of forget gate  
δ Sigmoid function  
h Implied unit  
i_i Value of input gate  
R_i Offset term of the input gate  
c_i Cell case  
o_i Value of output gate  
y_i^a True value  
y_i^p Predicted value  
y_i^b Average of the actual responses

References
Energies 2023, 16, 5347


10. Yu, Y.; Wang, Y.; Li, J.; Fu, M.; Shah, A.N.; He, C. A Novel Deep Learning Approach to Predict the Instantaneous NOx Emissions from Diesel Engine. IEEE Access 2021, 9, 11002–11013. [CrossRef]


29. Sindhi, H.; Nour, M.; Rawla, M.; Öztürk, Ş.; Polat, K. Random fully connected layered 1D CNN for solving the Z-bus loss allocation problem. Measurement 2021, 171, 108794. [CrossRef]


33. Oran, A.; Tocoglu, M.A. A Term Weighted Neural Language Model and Stacked Bidirectional LSTM Based Framework for Sarcasm Identification. IEEE Access 2021, 9, 7701–7722. [CrossRef]
34. Barbero Jiménez, Á.; López Lázaro, J.; Dorronsoro, J.R. Finding optimal model parameters by deterministic and annealed focused
grid search. *Neurocomputing* **2009**, *72*, 2824–2832. [CrossRef]

35. Yang, W.; Pu, C.X.; Yang, K.; Zhang, A.A.; Qu, G.L. Short-term fault prediction method for a transformer based on a CNN-GRU

**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.