

## Article

# Research on Machine Learning-Based Method for Predicting Industrial Park Electric Vehicle Charging Load

Sijiang Ma <sup>1,2</sup>, Jin Ning <sup>1,2</sup>, Ning Mao <sup>1,3,\*</sup>, Jie Liu <sup>1,3</sup> and Ruirong Shi <sup>1,2</sup>

<sup>1</sup> Laboratory of Transport Pollution Control and Monitoring Technology, Beijing 100028, China; masijian233@163.com (S.M.); ning.jin@ncepu.edu.cn (J.N.); liujie@tpri.org.cn (J.L.); shi.ruirong@ncepu.edu.cn (R.S.)

<sup>2</sup> School of Control and Computer Engineering, North China Electric Power University, Beijing 102206, China

<sup>3</sup> Transport Planning and Research Institute, Ministry of Transport, Beijing 100028, China

\* Correspondence: maoning@tpri.org.cn

**Abstract:** To achieve global sustainability goals and meet the urgent demands of carbon neutrality, China is continuously transforming its energy structure. In this process, electric vehicles (EVs) are playing an increasingly important role in energy transition and have become one of the primary user groups in the electricity market. Traditional load prediction algorithms have difficulty in constructing mathematical models for predicting the charging load of electric vehicles, which is characterized by high randomness, high volatility, and high spatial heterogeneity. Moreover, the predicted results often exhibit a certain degree of lag. Therefore, this study approaches the analysis from two perspectives: the overall industrial park and individual charging stations. By analyzing specific load data, the overall framework for the training dataset was established. Additionally, based on the evaluation system proposed in this study and utilizing both Multilayer Perceptron (MLP) and Long Short-Term Memory (LSTM) algorithms, a framework for machine learning-based load prediction methods was constructed to forecast electric vehicle charging loads in industrial parks. Through a case analysis, it was found that the proposed solution for the short-term prediction of the charging load in industrial park electric vehicles can achieve accurate and stable forecasting results. Specifically, in terms of data prediction for normal working days and statutory holidays, the Long Short-Term Memory (LSTM) algorithm demonstrated high accuracy, with  $R^2$  coefficients of 0.9283 and 0.9154, respectively, indicating the good interpretability of the model. In terms of weekend holiday data prediction, the Multilayer Perceptron (MLP) algorithm achieved an  $R^2$  coefficient of as high as 0.9586, significantly surpassing the LSTM algorithm's value of 0.9415, demonstrating superior performance.



**Citation:** Ma, S.; Ning, J.; Mao, N.; Liu, J.; Shi, R. Research on Machine Learning-Based Method for Predicting Industrial Park Electric Vehicle Charging Load. *Sustainability* **2024**, *16*, 7258. <https://doi.org/10.3390/su16177258>

Academic Editor: Bin Ji

Received: 24 July 2024

Revised: 19 August 2024

Accepted: 20 August 2024

Published: 23 August 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 recent years, driven by the national “dual-carbon” policy, the electric vehicle (EV) industry in China has been flourishing. A large number of EVs have been integrated into the power grid, but this has also brought new challenges to the stable and secure operation of the grid. In industrial parks with high EV penetration rates, the charging behavior of EVs during peak electricity consumption periods has increased the peak load on the grid. This has led to unfavorable effects, such as increased peak-to-valley difference, load fluctuation rate, and maximum load ratio, which have further burdened the distribution grid and affected the security of grid operation. Accurately predicting EV charging loads is a necessary task in the early stages of EV development.

Currently, the methods for predicting EV loads can be mainly categorized into two types: statistical simulation-based methods and machine learning-based methods. Statistical simulation-based methods primarily employ Monte Carlo simulation for load

predictions. This research process has evolved from strict assumptions to releasing restrictions on EV types, charging power, and date types. Reference [1] focuses on the number of EVs in order to simulate EVs' unordered charging behavior by conducting Monte Carlo simulations on the battery characteristics and user travel habits of EVs. References [2–5] simulates the unordered charging behavior of EVs from four perspectives, namely seasonal characteristics, travel path conditions, multi-dimensional states of EVs, and scene correlations, and successfully predicts EV charging loads. Reference [6] proposes a new charging load prediction method based on the characteristics of EVs' charging behavior and nonlinear programming. Reference [7] predicts the charging load of privately owned EVs in Beijing by studying the characteristics of EV travel patterns, urban road networks, functional area distributions, and the distribution of charging infrastructure. Additionally, to further enhance the accuracy of load forecasting, researchers have also incorporated traffic conditions and user psychology as significant factors into considerations for model construction. Reference [8] integrates various methods, such as the Gaussian Mixture Model, Floyd's algorithm, and real-time energy consumption models, to construct a microscopic traffic simulation model. Based on this, it proposes a framework for predicting EV charging load, which is driven by electricity prices and involves real-time interaction with network information. References [9–11] take the user's psychology as the starting point, comprehensively considering multi-source information such as ambient temperature, real-time electricity prices, and traffic network conditions. Based on existing theories and methods, such as the trip chain and OD matrix, the Monte Carlo method is successfully applied to predict the spatiotemporal charging demands of electric vehicles under different scenarios. However, traditional load prediction algorithms face difficulties in constructing mathematical models for load variation patterns and influencing factors when dealing with the high randomness, volatility, and spatial differences in EV charging loads. Moreover, the prediction results often have a certain lag.

Machine learning-based EV load prediction methods mainly use neural networks for prediction. Neural networks, as powerful machine learning tools, have excellent versatility, scalability, and fault tolerance and exceptional nonlinear mapping capabilities. This makes them particularly advantageous in handling complex regression problems, especially in predicting electricity consumption loads in power systems. Reference [12] proposes a novel EV load prediction model, ISSA-CNN-GRU, by integrating the feature extraction advantages of a CNN, the time-series data processing capability of a GRU, and the ISSA optimization algorithm. Reference [13] significantly improves the prediction accuracy of the LSTM model by analyzing the weights of various influencing factors and making adjustments. References [14–17] all take Graph Neural Networks as the starting point and propose several new electric vehicle load forecasting models. Reference [18] uses a GAN to generate missing load data to mitigate the impact of low-quality EV load data on prediction results and improves the performance of the LSTM model using a new gating mechanism called Mogrifier. Reference [19] incorporates the gray theory into EV load prediction for control systems and analyzes the role of the EMGM (Enhanced Multivariate Gray Model) in considering multiple influencing factors and eliminating cumulative errors. Reference [20] proposes the MCCNN-TCN load prediction model by combining the fluctuation feature extraction capability of the MCCNN model in the time scale and the time series prediction capability of TCN. Reference [21] proposes an electric vehicle charging load forecasting model that integrates a Genetic Algorithm (GA) with a Back Propagation Neural Network (BPNN). This model optimizes the weights and thresholds of the BPNN through the GA, thereby enhancing the accuracy and efficiency of the predictions. Reference [22] employs variational mode decomposition (VMD) to separate charging load data into high-frequency and low-frequency components, which are then forecasted using LSTM neural networks and the Prophet model, respectively. The experimental results demonstrate that this method exhibits significant advantages in terms of prediction accuracy and effectiveness. Given the remarkable performance of Transformer models in various fields in recent years, an increasing number of researchers are attempting to integrate attention mechanisms with neural

networks, aiming to enhance the predictive capabilities of these models. References [23–25] integrate attention mechanisms with recurrent neural network models, such as Long Short-Term Memory (LSTM) and a Gated Recurrent Unit (GRU), significantly enhancing the algorithm's capability to capture dependencies and correlations in time series data. The experimental results indicate that the models with attention mechanisms have achieved a notable improvement in various performance metrics for prediction outcomes. Although machine learning-based EV load prediction methods have been widely applied, there is still a lack of research on the impact of date types on the accuracy of each machine learning method and the selection of machine learning methods suitable for different date types.

The prediction of electric vehicle (EV) charging is a critically important topic. Scholars from both domestic and international backgrounds have proposed numerous constructive methods from various perspectives. These efforts have provided theoretical knowledge reserves and practical references for constructing a framework for predicting EV charging loads in industrial parks. However, there are still some limitations and deficiencies in the current study, which are summarized as follows:

(1) In the field of electric vehicle (EV) charging prediction and site selection, existing research predominantly focuses on urban-level regional data for charging load forecasting, with relatively limited attention being given to industrial parks, which are special areas characterized by industrial concentration.

(2) For traditional load forecasting methods, current research is quite stringent regarding the precision of mathematical modeling and the design of assumptions. Moreover, the forecasting results often exhibit a certain degree of lag.

(3) For machine learning methods, existing research indeed lacks in-depth theoretical exploration regarding the selection of prediction models for different types of dates in the context of electric vehicle charging.

In response to the aforementioned limitations, particularly the issue of model selection for different date types, this study constructs a framework for electric vehicle charging load prediction in industrial park scenarios, taking into account various date types. This framework is grounded in two machine learning algorithms: MLP and LSTM. Moreover, an evaluation index system and assessment algorithms are developed to evaluate the predictive performance of specific models under different date types.

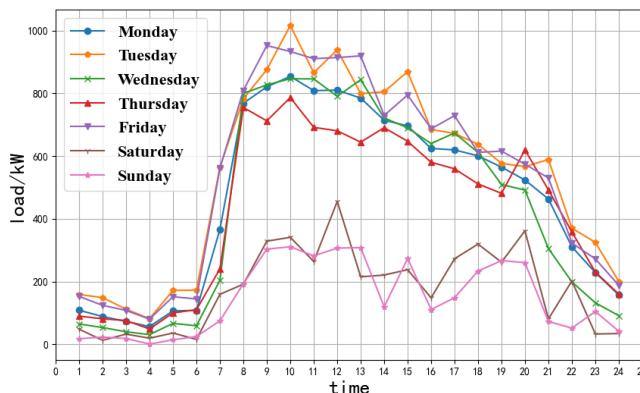
## 2. Factors Influencing the Electric Vehicle Charging Load in Industrial Parks

### 2.1. Factors Affecting the Operational Mode of Industrial Parks

#### 2.1.1. Impact of Commuting Patterns

The temporal characteristics variables that influence the electric vehicle (EV) charging load mainly include the start charging time and the daily mileage. Therefore, it is essential to fully understand the composition and daily usage patterns of EVs within the study area. According to an interview reported by the China Environment News, the development of new energy heavy-duty trucks is still in its initial stages, having only shown an accelerated trend in the past two years. These trucks are primarily used in closed environments and key industries involving bulk cargo transportation, such as steel, coal, and power. However, in most industrial parks, the main types of EVs are still buses, private cars, and official vehicles. Consequently, the time-related factors of the EV charging load in industrial parks are likely closely related to work schedules. Figure 1 shows the variation curve of the EV charging load in an industrial park over the course of a week.

From the figure, it can be observed that as the industrial park officially starts work at 8:30 AM, the EV charging load rapidly increases, reaching a peak at around 11:00 AM. Subsequently, as some private car batteries become fully charged and some official and public buses commence normal operations, the charging load begins to gradually decrease. By 6:00 PM, as most employees finish their workdays, the load further reduces. The load curve also indicates that a significant portion of load consumption remains outside of standard working hours, which is particularly noticeable between 6:00 PM and 11:00 PM.



**Figure 1.** Weekly charging load curve of charging vehicles in industrial parks.

Overall, unlike the nighttime single-peak load trend in urban-level EV loads, the charging load peak in industrial parks primarily occurs between 10:00 AM and 11:00 AM and exhibits clear periodicity. Additionally, specific load conditions are significantly influenced by holidays and company commuting policies.

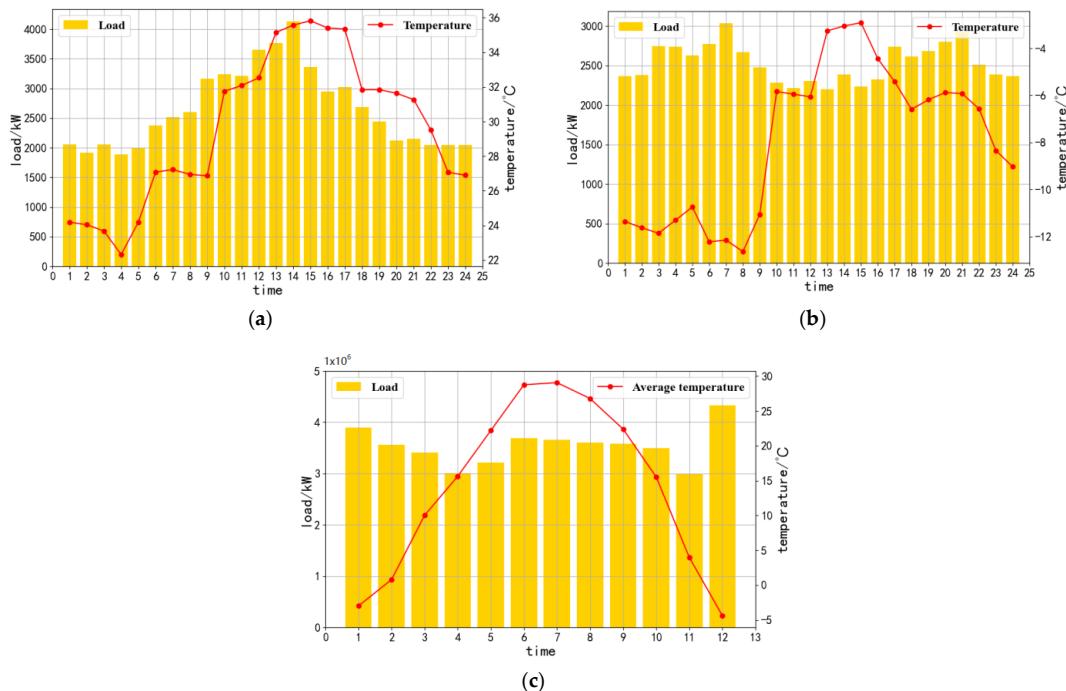
### 2.1.2. Impact of Load Regulation

In industrial parks, the peak EV charging load is primarily concentrated between 10:00 and 11:00 AM, which coincides with the peak load period for most electrical equipment in the park. The additional load from EV charging exacerbates fluctuations in the park's load curve, potentially adversely affecting the safe operation of the microgrid. Consequently, many industrial parks implement load regulation strategies to manage the EV charging load, adjusting the charging power during specific periods to ensure the safe operation of critical facilities.

However, external meteorological conditions significantly impact the park's power load. Therefore, studying the relationship between meteorological factors and the park's power load is crucial for accurately predicting the EV charging load. According to the literature [26], among various meteorological conditions, temperature has the most critical impact on the park's power load. With the increasing precision and complexity of industrial equipment, these devices demand more stable working environment temperatures. Simultaneously, the demand for comfortable office environments within the park is also rising. To meet these requirements, it has become necessary to operate air conditioning systems continuously to regulate and maintain a constant temperature.

Figure 2 illustrates a comparison between the power load and temperature curves for an industrial park in North China.

An analysis of Figure 2 reveals that the unique temperate continental climate characteristics in Northern China have a significant impact on the power load of the industrial park. Consequently, during seasonal transitions, the power load in the park experiences drastic fluctuations. Additionally, the power load during summer shows significant variations throughout the day, which positively correlates with temperature changes. In contrast, the power load in winter is relatively stable, exhibiting a negative correlation with temperature fluctuations.



**Figure 2.** Electricity load and temperature curve of industrial parks. (a) Daily electricity load and temperature curve in industrial park during summer. (b) Daily electricity load and temperature curve in industrial park during summer. (c) Monthly electricity load and monthly average temperature curve in industrial park.

## 2.2. Factors Affecting the Load of Charging Stations

Charging stations, as key charging infrastructure for electric vehicles (EVs), are subject to a complex set of influencing factors that are crucial for accurately predicting EV charging loads. A literature review [27,28] suggested that the charging station load is primarily affected by temperature conditions, charging modes, the scale of EVs, and the types of vehicles. Given the high similarity in behavioral logic among the main vehicle types used in industrial parks, such as buses, private cars, and official vehicles, this study will focus on the impact of temperature conditions and charging modes on the charging load.

The influence of temperature on a charging station load manifests in two main aspects: first, it manifests in the thermal characteristics of EV battery packs. Currently, the power batteries used in EVs mainly include lead-acid batteries, nickel–cadmium batteries, nickel–metal hydride batteries, and lithium-ion batteries. As indicated in the literature [29], the optimal operating temperature for power batteries like lithium-ion is between 20 and 50 degrees Celsius. Operating at non-optimal temperatures affects battery efficiency, reduces EV range, and increases charging frequency. Charging at non-optimal temperatures also leads to increased battery internal resistance, further increasing the charging load. Secondly, temperature impacts driver behavior logic. At uncomfortable driving temperatures, drivers may use the vehicle's air conditioning to improve the in-car thermal environment, which increases energy consumption and thus the frequency and duration of charging.

The charging mode of EVs significantly affects the load on charging stations. The current mainstream charging modes include AC charging, DC charging, and battery swapping. To meet the commuting needs of industrial park employees and related official vehicles, most industrial parks are equipped with a large number of AC charging stations. AC charging mode uses a low-current constant voltage or constant current charging, with an output power typically not exceeding 5 kW and a charging time of about 3 to 8 h, which is sufficient for most charging needs and can fully utilize off-peak load periods, significantly improving the stability of the park's microgrid. However, industrial parks often face urgent commuting needs, requiring rapid charging for specific vehicles. For such

needs, DC charging mode or battery swapping mode becomes an alternative. Due to the longer charging time and higher charging current of DC charging, and the impact on the distribution network during charging, the current park trend is moving towards building a charging system primarily based on battery swapping, supplemented by whole-vehicle charging, to ensure the rapid battery replacement needs of vehicles.

Based on the analysis of the influencing factors discussed above, a collection of factors affecting the EV charging load in industrial parks is presented in Table 1.

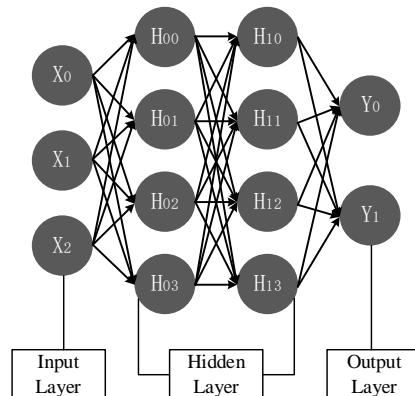
**Table 1.** Summary of influencing factors of electric vehicle charging load in industrial parks.

Influencing Factor	Feature	Description
Historical Data	load <sub>1t</sub> –load <sub>7t</sub>	Load data at time t from previous 7 days
Time Factors	is_weekend month day hour	Whether it is holiday or weekend Month of load Day of load Hour of load
Meteorological Factors	temp	Temperature at time t

### 3. Industrial Park Electric Vehicle Charging Load Forecasting Model

#### 3.1. Prediction Model Based on Multilayer Perceptron (MLP)

Multilayer Perceptron (MLP) is an advanced form of single-layer perceptron and belongs to the category of feedforward artificial neural networks. The architecture of this network consists of at least three layers: an input layer, one or more hidden layers, and an output layer. MLP transmits information through the connections between these layers and uses nonlinear activation functions to handle and solve more complex problems. Figure 3 shows the structure of a multilayer perceptron model with two hidden layers.



**Figure 3.** Multivariate perceptron model with hidden layer. In the figure, each X denotes an input feature, while every H signifies a neuron within the hidden layer. Similarly, each Y corresponds to a model output. The direction of the arrows illustrates the flow of data through the network.

The functions of the three layers in MLP are as follows:

1. Input Layer: This receives various raw data, where each input unit represents a data feature that needs attention;
2. Hidden Layers: This uses nonlinear activation functions to transform and process the input data, thereby extracting features from the input data;
3. Output Layer: This receives the features extracted by the hidden layers and, based on the weight parameters obtained from model training, makes the final decision or prediction.

During the training process, the first step is to calculate the output value of the nodes. For node  $j$  in the  $l$  layer of the multilayer perceptron, its output  $net_{l,j}$  is calculated using the following formula:

$$net_{l,j} = f_{act} \left( \sum_{k=1}^h w_{l,j,k} a_{l-1,k} + b_{l,j} \right) \quad (1)$$

In the formula,  $net_{l,j}$  represents the value of node  $j$  in the  $l$  layer,  $f_{act}(x)$  represents the activation function,  $w_{l,j,k}$  is the connection weight between node  $k$  in the  $l-1$  layer and the current node,  $a_{l-1,k}$  is the value of node  $k$  in the  $l-1$  layer, and  $b_{l,j}$  is the bias parameter of the current node.

Next, it is necessary to calculate the loss value between the output value and the true value, which is typically carried out using either the Mean Squared Error (MSE) or Cross-Entropy methods (CE). The calculation formulas are as follows:

*MSE Loss:*

$$MSE\ Loss = \frac{1}{n} \sum_{i=1}^n (y_{true} - y_{pre})^2 \quad (2)$$

*CE Loss:*

$$CE\ Loss = -\frac{1}{n} \sum_{i=1}^n y_{true} \log(y_{pre}) \quad (3)$$

In the formula,  $n$  is the number of samples,  $y_{true}$  is the true value of samples, and  $y_{pre}$  is the predicted value of samples.

Finally, the weights are updated from the output layer to the input layer by calculating the gradient of the loss function with respect to the weights of each layer. The specific calculation formulas are as follows:

*Weight Gradient Calculation Formula:*

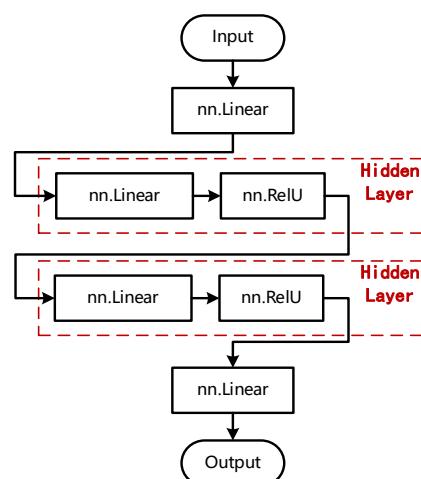
$$\frac{\partial L}{\partial w_{l,j,k}} = \frac{\partial L}{\partial net_{l,j}} \frac{\partial net_{l,j}}{\partial w_{l,j,k}} \quad (4)$$

*Weight Update Formula:*

$$w_{l,j,k} \leftarrow w_{l,j,k} - \alpha \frac{\partial L}{\partial w_{l,j,k}} \quad (5)$$

In the formula,  $L$  is the value loss, and  $\alpha$  is the learning rate of the model.

The industrial park electric vehicle charging load MLP prediction model is shown in Figure 4.

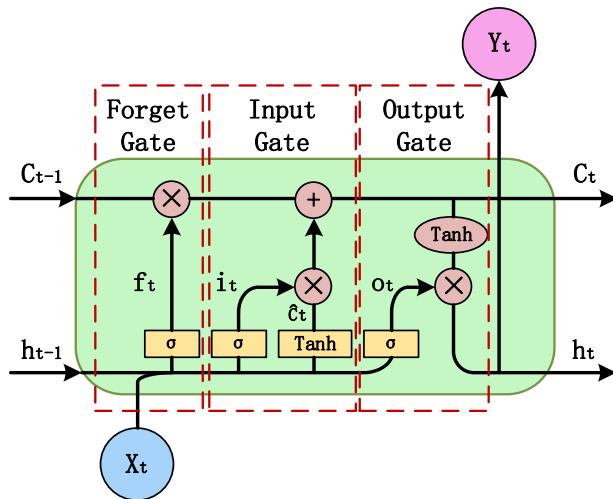


**Figure 4.** MLP prediction model of electric vehicle charging load in industrial parks. Arrow direction represents the flow of data.

### 3.2. Prediction Model Based on Long Short-Term Memory Network (LSTM)

In the early stages of neural network-based power load forecasting, researchers typically used Recurrent Neural Networks (RNNs) to predict power load conditions [30]. However, during training, RNNs faced issues such as vanishing and exploding gradients caused by the backpropagation algorithm. These issues were particularly pronounced when dealing with long sequences of data, making it difficult for RNNs to effectively learn long-distance temporal dependencies in time series data.

To address this problem, Shi X et al. [31] proposed a neural network structure composed of three gate structures, forget gate, input gate, and output gate, known as Long Short-Term Memory (LSTM). The architecture of an LSTM neuron is shown in Figure 5.



**Figure 5.** LSTM neuron architecture diagram. In the diagram, the multiplication symbol “ $\times$ ” represents matrix multiplication, and the addition symbol “ $+$ ” denotes matrix addition.

The functions of the three gates in LSTM are as follows:

1. Forget gate: This determines which information should be forgotten from the previous cell state.
2. Input gate: This determines which input information at this time will be stored in the current cell state.
3. Output gate: This determines which information from the current cell state will be output to the next hidden state.

During the training process, first, the forget gate selectively erases some content from the previous memory cell based on the previous hidden state  $h_{t-1}$  and the current input  $x_t$ . The calculation formula is as follows:

$$f_t = \sigma[W_f(h_{t-1} + x_t)] + b_f \quad (6)$$

In the formula,  $h_{t-1}$  represents the hidden layer state from the previous time step,  $x_t$  represents the input state at the current time step,  $\sigma(\cdot)$  represents the Sigmoid activation function,  $W_f$  is the weight parameter of the forget gate, and  $b_f$  is the bias parameter of the forget gate.

Next, the input gate integrates new information into the memory cell processed by the forget gate based on the previous hidden state  $h_{t-1}$  and the current input  $x_t$ . The calculation formulas are as follows:

$$i_t = \sigma[W_i(h_{t-1} + x_t)] + b_i \quad (7)$$

$$\hat{C}_t = \text{Tanh}[W_c(h_{t-1} + x_t)] + b_c \quad (8)$$

In the formula,  $\text{Tanh}(\cdot)$  is the Tanh activation function,  $\hat{C}_t$  is the candidate memory cell for the new state,  $i_t$  is used to decide which memory cells need to be updated,  $W_i$  and  $W_c$  are the weight parameters of the input gate, and  $b_i$  and  $b_c$  are the bias parameters of the input gate.

Finally, the output gate calculates the current hidden state  $h_t$  and the current output  $y_t$  based on the previous hidden state  $h_{t-1}$ , the current input  $x_t$ , and the results of the forget gate and input gate. The calculation formulas are as follows:

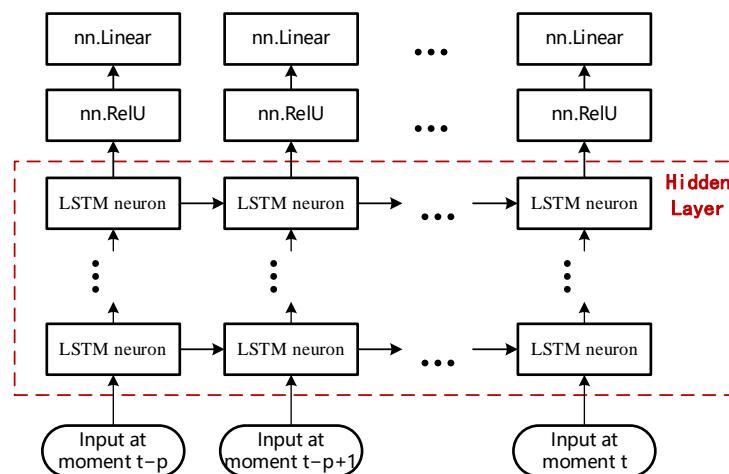
$$C_t = C_{t-1} \cdot f_t + \hat{C}_t \cdot i_t \quad (9)$$

$$o_t = \sigma[W_o(h_{t-1} + x_t)] + b_o \quad (10)$$

$$y_t = h_t = o_t \cdot \text{Tanh}(C_t) \quad (11)$$

In the formula,  $W_o$  is the weight parameter of the output gate,  $b_o$  is the bias parameter of the output gate, and  $\cdot$  represents matrix multiplication.

The industrial park electric vehicle charging load LSTM prediction model is shown in Figure 6.



**Figure 6.** LSTM prediction model of electric vehicle charging load in industrial parks. Arrow direction represents the flow of data.

#### 4. Industrial Park Electric Vehicle Charging Load Forecasting Method

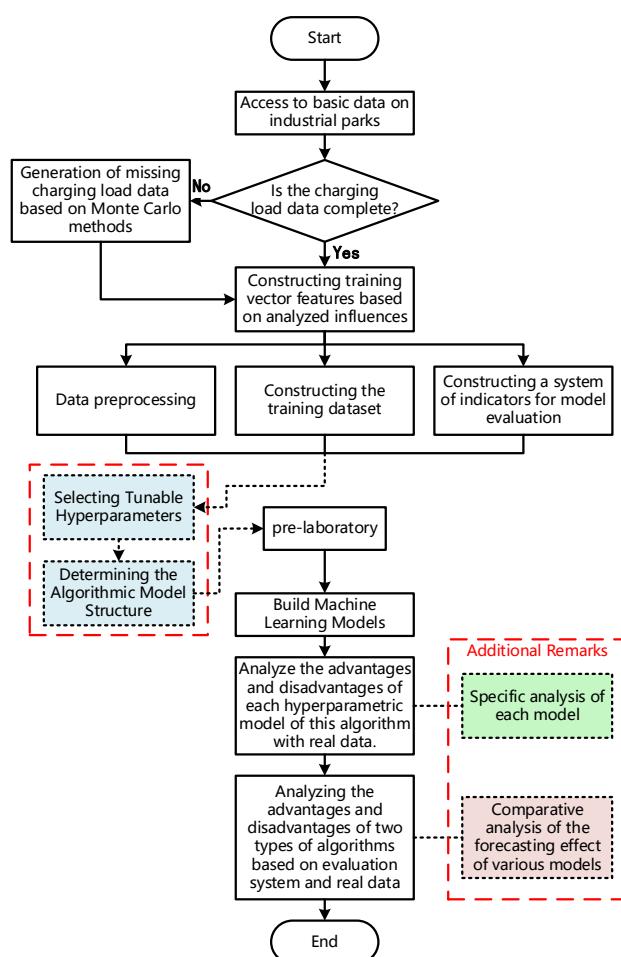
##### 4.1. Solution Methodology Workflow

Given that the electric vehicle charging scenarios within the industrial park mainly serve commuting needs and the scenarios are relatively simple, the research problem of electric vehicle charging load prediction in the industrial park is essentially a time series regression analysis problem. The research objectives are to use different machine learning methods to predict the electric vehicle charging load in the industrial park under the same application scenario and to evaluate the prediction results based on a certain evaluation system to determine the most suitable machine learning method for different types of dates. The specific calculation process is as follows:

1. Investigate the basic data of the target industrial park, including the area of the park, the historical data of the park's electric vehicle charging load, the number of electric vehicles in the park, the meteorological data of the park, and the park's normal commuting periods.
2. Perform basic data processing on the survey data, and based on the processed data, construct the feature structure of the machine learning training vector, and build the training dataset and the model evaluation indicator system accordingly.
3. Determine the hyperparameter settings for various machine learning algorithms through a large number of preliminary experiments, and based on these designs, train the corresponding hyperparameter models.
4. Apply the trained hyperparameter models for the prediction of various types of date data and evaluate the model algorithms based on the constructed evaluation system. At the same time, compare the predicted values of each group of hyperparameter mod-

- els with the real values, and specifically analyze the advantages and disadvantages of each group of hyperparameters.
5. Summarize the calculation results of the evaluation system and prediction results of the optimal hyperparameter models of various types in order to comprehensively assess the advantages and disadvantages of the three machine learning algorithms.

To implement machine learning algorithms and validate the accuracy of electric vehicle charging load prediction in industrial parks, this study utilized the PyTorch package for model construction. The detailed solution workflow is shown in Figure 7.



**Figure 7.** The process of solving the problem of electric vehicle charging load prediction in industrial parks.

#### 4.2. Data Processing

When using machine learning methods to predict electric vehicle charging loads, data preprocessing is a crucial step.

##### 4.2.1. Data Classification

Holidays and weekends typically lead to changes in people's travel patterns, which significantly affect the charging demand for electric vehicles. When training machine learning models, the weight of this feature is particularly emphasized, thereby influencing the model's perception of other features. Consequently, the dataset is divided into three categories: weekend holidays, statutory holidays, and normal working days, based on whether it is a holiday or a weekend. The information on the number of datasets in each category is presented in Table 2.

**Table 2.** Number of dataset groups by category.

Influencing Factor	Feature
weekend holidays	2064
statutory holidays	648
normal working days	5881

#### 4.2.2. Outlier Treatment

This study employs the mean value filling method to fill in the outliers. The relevant filling formula is as follows:

$$y_{load\_t} = \frac{y_{load\_t-1} + y_{load\_t+1}}{2} \quad (12)$$

In the formula,  $y_{load\_t}$  is the charging load value at the time of the outlier,  $y_{load\_t-1}$  is the charging load value at the previous time step, and  $y_{load\_t+1}$  is the charging load value at the next time step. In practical applications, there are often missing values for a period of time. If the missing day is a working day (holiday), the charging load at the same time on the previous working day (holiday) can be used as  $y_{load\_t-1}$ , and the charging load at the same time on the previous working day (holiday) can be used as  $y_{load\_t+1}$ , and these values are then brought into the formula for calculation.

#### 4.2.3. Data Standardization

Different features in the dataset have different dimensions and scales, especially the numerical scale of the historical load features load1–load7, which is much larger than that of other features. This leads to a shift in the focus of the model, thereby neglecting the impact of other features. To address this issue, this study uses Z-Zero standardization to process the raw data. The relevant calculation formula is as follows:

$$x_{ij} = \frac{x_{ij} - \mu(x_i)}{\sigma(x_i)} \quad (13)$$

In the formula,  $x_{ij}$  represents the standardized value of the  $j$  value of the  $i$  feature parameter,  $\mu(x_i)$  is the mean of the  $i$  feature parameter, and  $\sigma(x_i)$  is the variance in the  $i$  feature parameter.

#### 4.3. Dataset Structure

When constructing the load data part of the dataset, this study treats the charging load data from the same moment of the previous seven days as a class of feature vectors to be considered. The purpose of this design is to enhance the model's ability to understand and process time series data. By analyzing the charging load patterns of the past seven days, the model can identify potential periodicity and trends, thereby making more accurate predictions of future charging needs. The specific dataset structure and some data are shown in Tables 3 and 4.

**Table 3.** Historical load data structure and some data on electric vehicles.

Load1/kW	Load2/kW	Load3/kW	Load4/kW	Load5/kW	Load6/kW	Load7/kW
394.63	184.42	250.78	264.73	312.37	214.44	73.02
270.22	182.93	172.85	228.49	217.95	185.71	79.18
204.28	134.82	144.38	202.64	174.33	178.58	69.13
...	...	...	...	...	...	...

**Table 4.** Time and meteorological data structure and some data.

Load Month/Month	Load Date/Day	Load Hour/h	Temp/°C	Is_Weekend	Label/kW
1	8	0	-4.50	1	75.93
1	8	1	-4.25	1	67.17
1	8	2	-4.63	1	47.28
...	...	...	...	...	...

#### 4.4. Evaluation Metrics System

In the prediction of the electric vehicle charging load in industrial parks, as a typical regression analysis problem, the construction of a model evaluation metric system is crucial for both selecting models and addressing the problem itself. After considering the specific needs, this study constructs an evaluation system based on the Mean Absolute Error (*MAE*), Root Mean Squared Error (*RMSE*), and R-square value (*R*<sup>2</sup>).

##### 4.4.1. Mean Absolute Error (*MAE*)

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (14)$$

In the formula, *n* is the number of test set samples, *y<sub>i</sub>* is the true value of the electric vehicle charging load, and *ŷ<sub>i</sub>* is the predicted value of the electric vehicle charging load. By calculating the *MAE*, the average deviation between the model's predicted load and the actual load can be assessed, thereby judging the accuracy of the model's prediction. A smaller value of *MAE* indicates a more accurate prediction.

##### 4.4.2. Root Mean Squared Error (*RMSE*)

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (15)$$

In the formula, *n* is the number of test set samples, *y<sub>i</sub>* is the true value of the electric vehicle charging load, and *ŷ<sub>i</sub>* is the predicted value of the electric vehicle charging load. Compared to the *MAE*, the *RMSE* imposes a greater penalty on larger errors. In cases where the prediction errors are significant, the *RMSE* can more stringently evaluate the performance of the model's predictions.

##### 4.4.3. R-Square (*R*<sup>2</sup>)

$$R^2 = 1 - \frac{SSE}{SST} = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (16)$$

In the formula, *n* is the number of test set samples, *y<sub>i</sub>* is the true value of the electric vehicle charging load, *ŷ<sub>i</sub>* is the predicted value of the electric vehicle charging load, and *ȳ* is the average of the true electric vehicle charging load. *R*<sup>2</sup> is used to assess the interpretability of the predictions. The better the interpretability, the closer the coefficient *R*<sup>2</sup> is to 1. If the coefficient *R*<sup>2</sup> is less than 0, it indicates that the predictability of the predictions is very poor.

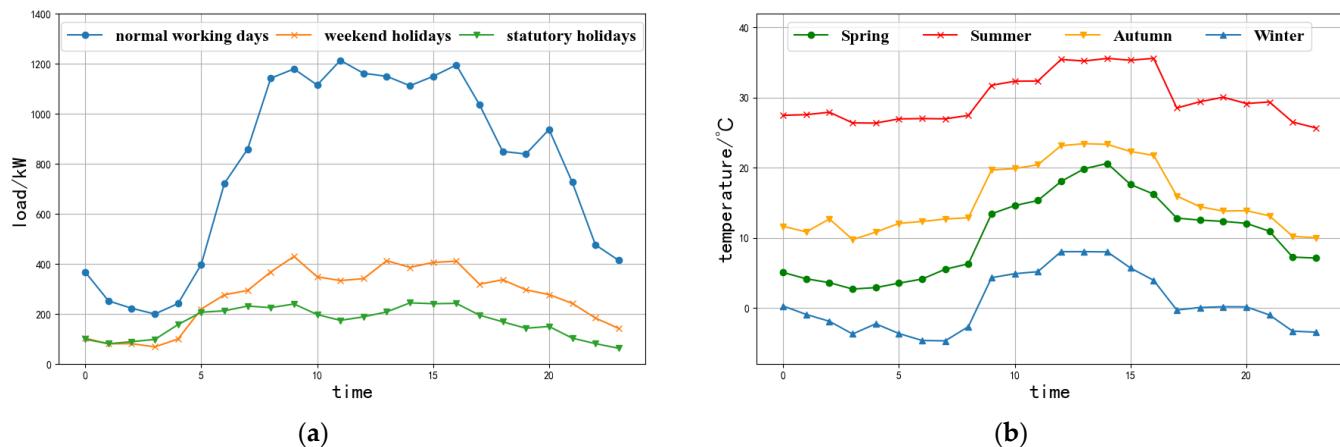
## 5. Case Study Analysis

### 5.1. Source of Case Study

To validate the constructed dataset structure and the effectiveness of the solution process, and to clarify which machine learning method is more effective under different types of dates, this study takes an industrial park in the North China region as a case

study. Based on the initial charging distribution and initial charging capacity distribution provided in the literature [32], combined with existing partial data, the Monte Carlo method was used to simulate a year's worth of electric vehicle charging load data. Subsequently, a 24 h predictive analysis of the charging load for electric vehicles in the park was conducted.

The case study in this study covers a full year of time data, temperature data, and load data for the year 2023 with a data time resolution of 1 h. Curves of various typical day-related data are shown in Figure 8.



**Figure 8.** Typical daily correlation curves. (a) Electric vehicle charging load profiles for various typical days. (b) Typical daily temperature curves for each season.

### 5.2. Experimental Scheme and Model Parameter Design

For the MLP model, the case study involves dividing the data of the 365 days into a training set and a testing set in a ratio of 4:1. These data are then utilized for model training. The prediction day is divided into 24 points, with data statistics conducted every hour. These data are subsequently applied to the prediction model to obtain the charging load curve for the typical day in the industrial park. The initial hyperparameter design for this analysis is presented in Table 5.

**Table 5.** Initial hyperparameters of Multilayer Perceptron (MLP).

Hyperparameter Type	Hyperparameter Value
Activation Function	Tanh(), Relu()
Number of Hidden Layers	2
Epoch	40,000
learning_rate	0.00001–0.001
Training Set Proportion	80%
Test Set Proportion	20%
Normalization Range	(−1, 1)

For the LSTM model, the case study involves grouping the data of the 365 days into two categories: 5-day granularity and 7-day granularity. Each group is then divided into a training set and a testing set in a ratio of 4:1. These data are then utilized for model training. The prediction day is divided into 24 points, with data statistics conducted every hour. These data are subsequently applied to the prediction model to obtain the charging load curve for a typical day in the industrial park. The initial hyperparameter design for this analysis is presented in Table 6.

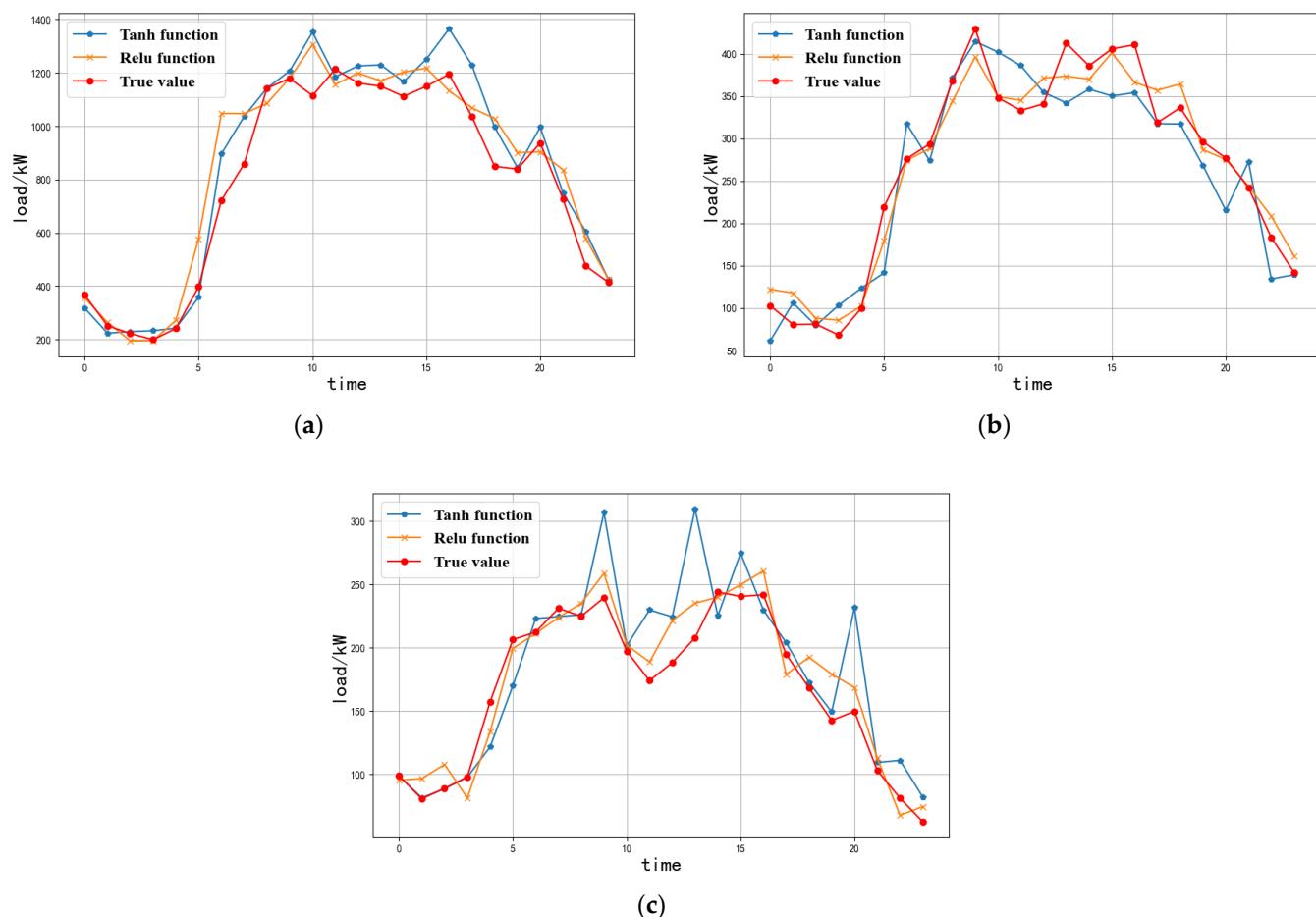
**Table 6.** Initial hyperparameters of Long Short-Term Memory Network (LSTM).

Hyperparameter Type	Hyperparameter Value
seq_length	$24 \times 5, 24 \times 7$
Number of Hidden Layers	1
epoch	150
learning_rate	0.00001~0.001
batch_size	32
hidden_dim	16
Training Set Proportion	80%
Test Set Proportion	20%
Normalization Range	(0, 1)

### 5.3. Analysis of Prediction Results

#### 5.3.1. An Analysis of the MLP Load Forecasting Structure

To investigate the effectiveness of the MLP model in predicting the electric vehicle load in industrial parks, and to examine the impact of different activation functions on the forecasting results, this study employs two MLP models: one with the Tanh activation function and the other with the Relu activation function. These models are used to predict the electric vehicle charging load in the same industrial park. The comparison between the predicted and actual load curves is illustrated in Figure 9.

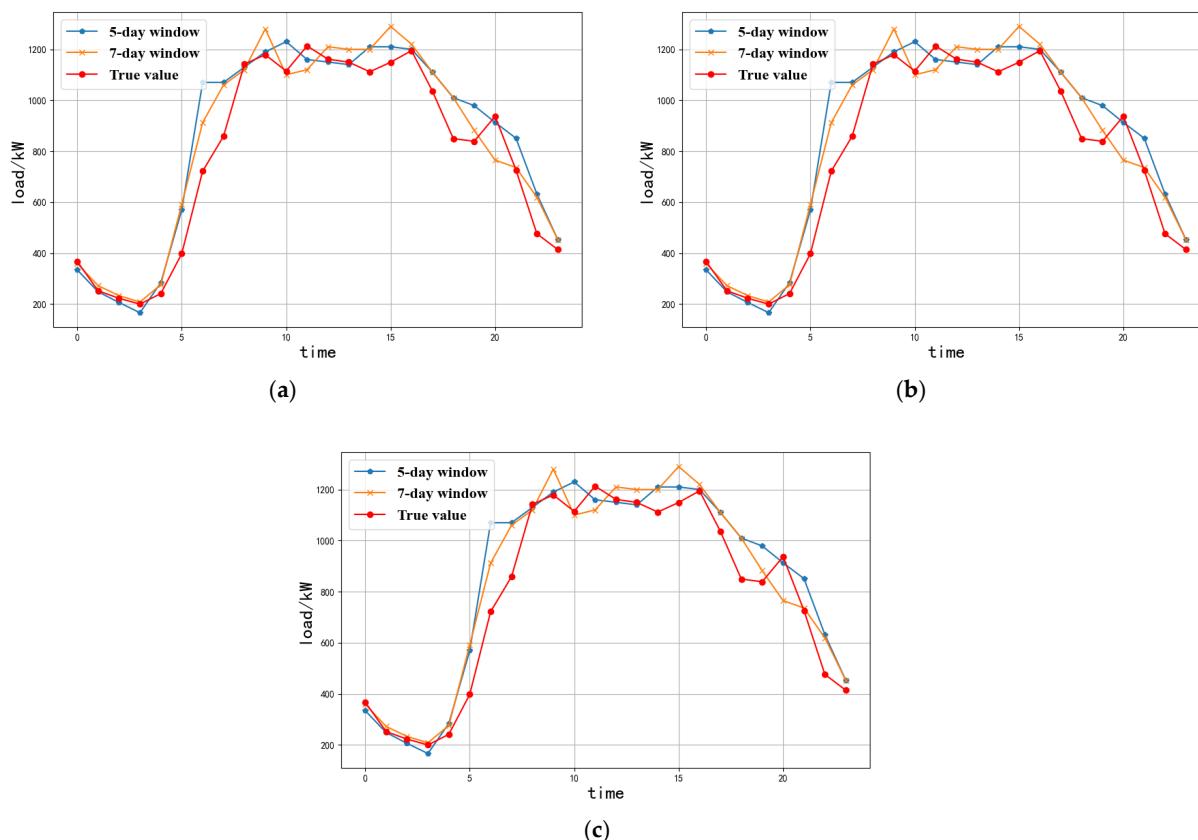


**Figure 9.** Load forecasting curves of the MLP model based on various date data. (a) Load forecast curve of normal working day. (b) Load forecast curve of weekend holidays. (c) Load forecast curve of statutory holidays.

An analysis of Figure 9 reveals that, in predicting normal working day data, both MLP models with the Tanh and Relu activation functions demonstrate good forecasting performance. However, in predicting weekend and holiday data, the model with the Tanh activation function shows a poorer trend prediction, whereas the model with the Relu activation function exhibits superior performance in capturing the trend details at peak points. Moreover, in forecasting data for statutory holidays, the model with the Relu activation function displays a high degree of curve fitting, effectively capturing and accurately simulating the micro-bimodal load characteristics in the peak regions.

### 5.3.2. An Analysis of the LSTM Load Forecasting Structure

To assess the effectiveness of the LSTM model in predicting the electric vehicle load in industrial parks, and to examine the impact of different historical window sizes on the forecasting results, this study employs two LSTM models: one with a 5-day historical window and the other with a 7-day historical window. These models are used to predict the electric vehicle charging load in the same industrial park. The comparison between the predicted and actual load curves is illustrated in Figure 10.



**Figure 10.** Load forecasting curves of the LSTM model based on various date data. (a) Load forecast curve of normal working day. (b) Load forecast curve of weekend holidays. (c) Load forecast curve of statutory holidays.

An analysis of Figure 10 indicates that, in predicting normal working day data, both LSTM models with 5-day and 7-day historical windows demonstrate good forecasting performance. However, in predicting weekend and holiday data, both models exhibit lower prediction accuracy for peak values, with the predicted values generally being slightly higher than the actual load values. This discrepancy is particularly significant during the 10:00–12:00 time period. In forecasting data for statutory holidays, the LSTM model with a 7-day historical window performs similarly to the MLP model with the

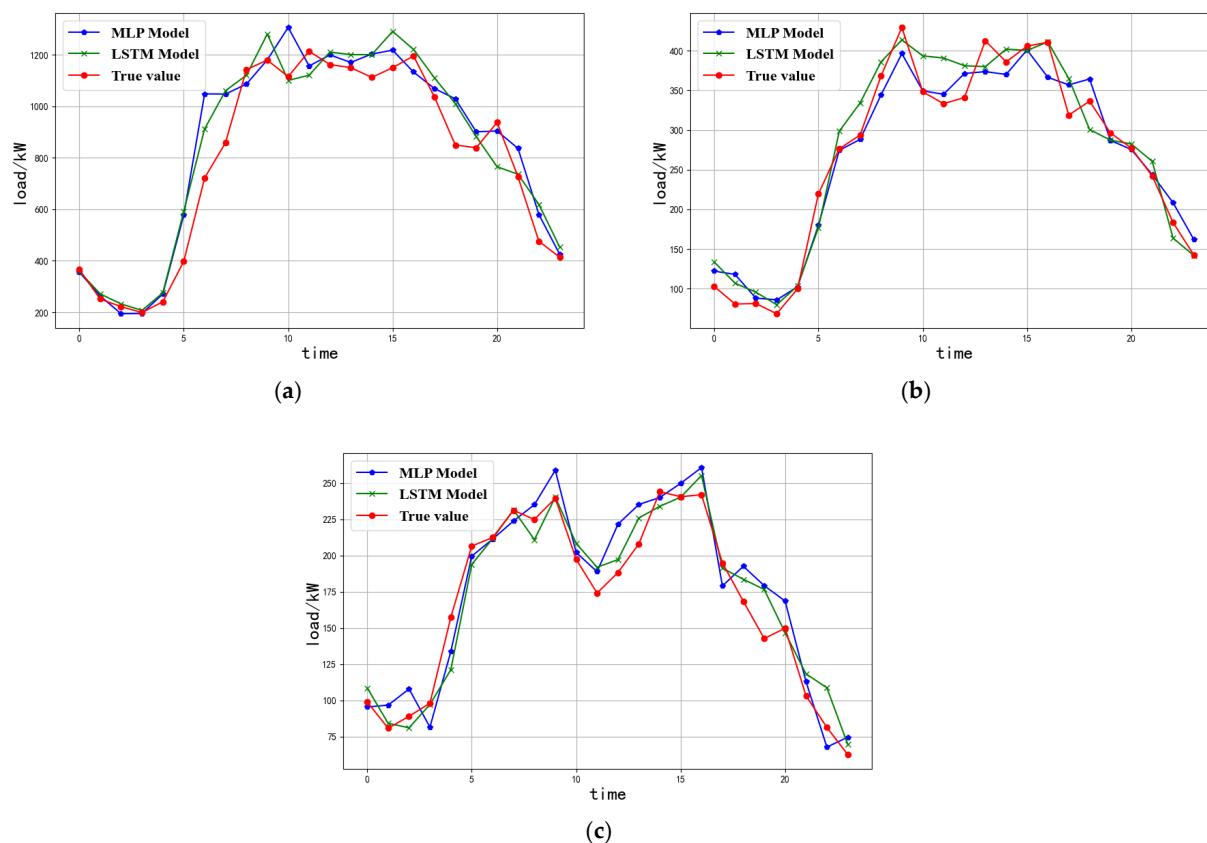
Relu activation function, effectively capturing and accurately simulating the micro-bimodal load characteristics in the peak regions.

#### 5.4. Comparative Analysis of Different Model Methods

To comprehensively compare the forecasting results of the two machine learning methods, based on the prediction results in Figures 9 and 10, and using the evaluation system proposed in this study, the relevant evaluation metrics under different schemes are calculated, as shown in Table 7. The comparison between the predicted and actual load curves is illustrated in Figure 11.

**Table 7.** The prediction results of the two machine learning methods.

Date Type	Evaluation Metrics	Algorithm	
		MLP	LSTM
normal working days	MAE	78.9347	74.1825
	RMSE	111.5846	97.0972
	$R^2$	0.9054	0.9283
weekend holidays	MAE	19.0759	23.2543
	RMSE	23.7417	28.2115
	$R^2$	0.9586	0.9415
statutory holidays	MAE	15.2526	11.3021
	RMSE	17.7142	14.9829
	$R^2$	0.9098	0.9154



**Figure 11.** Comparison of load forecast curves for different types of date data for two algorithmic models. (a) Load forecast curve of normal working day. (b) Load forecast curve of weekend holidays. (c) Load forecast curve of statutory holidays.

Based on the analysis of Table 7 and Figure 11, it can be observed that, in terms of predictive accuracy, both the MLP and LSTM models perform relatively well. Specifically, during normal working days and statutory holidays, where the duration is longer and the temporal logic of the load data is more continuous, LSTM, which is specifically designed for sequence data processing, demonstrates superior performance in these scenarios. For weekend and holiday data, the continuous data within a week typically accounts for only a small proportion, which may not be sufficient for the LSTM network to fully utilize its capabilities. In such cases, the MLP model, which primarily predicts the charging load based on feature vectors, exhibits better performance.

## 6. Conclusions

By addressing the prediction challenges of high randomness, high volatility, and significant spatial variability in electric vehicle charging loads in industrial parks, this paper constructs predictive models based on the MLP and LSTM machine learning algorithms. Through case studies, machine learning models are established and successfully predict the electric vehicle charging load in industrial parks within the next 24 h. The main conclusions are as follows:

(1). For industrial parks with relatively uniform types and uses of electric vehicles, the dataset structure and solution process proposed in this study can achieve an accurate prediction of electric vehicle charging loads.

(2). In predicting electric vehicle charging loads during normal working days and statutory holidays, the Long Short-Term Memory (LSTM) algorithm stands out as the most effective. However, during weekend holidays, when continuous data are limited, the Multilayer Perceptron (MLP) algorithm demonstrates superior performance.

The findings of this research can provide a reference for predicting electric vehicle charging loads in similar industrial parks and offer decision-making support for the construction and selection of electric vehicle infrastructure planning within these parks.

**Author Contributions:** R.S.: Writing—Original Draft, Writing—Review and Editing, Methodology, Investigation, Conceptualization, and Funding Acquisition. S.M.: Writing—Original Draft, Writing—Review and Editing, Software, Methodology, and Formal Analysis. J.N.: Software, Data Curation, and Validation. J.L.: Writing—Review and Editing, Methodology, and Conceptualization. N.M.: Writing—Review and Editing review & editing and, Supervision. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was financially supported by the Laboratory of Transport Pollution Control and Monitoring Technology (Grant No. (2024)JH-F003) and the Fundamental Research Funds for the Central Universities (Grant No. 2023JC007).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author due to some commercial requirement.

**Acknowledgments:** The authors are immensely grateful to the editors and reviewers of the journal for their constructive comments on this manuscript.

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

## References

1. Li, N.; Ma, H. Research on Electric Vehicle Ownership and Load Prediction Method Based on Comprehensive Prediction Model and Monte Carlo. *Mot. Control Appl.* **2022**, *49*, 74–80.
2. Niu, M.; Liao, K.; Yang, J.; Xiang, Y. Multi-time-scale electric vehicle load forecasting model considering seasonal characteristics. *Power Syst. Prot. Control* **2022**, *50*, 74–85.
3. Hua, Y.; Wang, Y.; Han, D.; Bu, F.; Wang, H.; Jia, Y. Long-term charging load forecasting of electric vehicles in residential areas considering orderly charging. *J. Electr. Power Sci. Technol.* **2022**, *34*, 142–150.

4. Zhang, M.; Ye, R.; Yang, X.; Sun, Q. Charging station planning for electric vehicle based on charging load forecast by MCMC method in multi-dimensional state space. *J. Electr. Power Sci. Technol.* **2022**, *37*, 78–87.
5. Huang, N.; Liu, D.; Cai, G. Interval prediction of electric vehicle charging load based on scene generation with multiple correlation days. *Proc. CSEE* **2021**, *41*, 10.
6. Tian, J.; Lv, Y.; Zhao, Q.; Gong, Y.; Li, C.; Ding, H.; Yu, Y. Electric vehicle charging load prediction considering the orderly charging. *Energy Rep.* **2022**, *8*, 124–134. [[CrossRef](#)]
7. Zhang, L.; Huang, Z.; Wang, Z.; Li, X.; Sun, F. An urban charging load forecasting model based on trip chain model for private passenger electric vehicles: A case study in Beijing. *Energy* **2024**, *299*, 130844. [[CrossRef](#)]
8. Bian, H.; Ren, Q.; Guo, Z.; Zhou, C.; Zhang, Z.; Wang, X. Predictive Model for EV Charging Load Incorporating Multimodal Travel Behavior and Microscopic Traffic Simulation. *Energies* **2024**, *17*, 2606. [[CrossRef](#)]
9. Zhang, M.; Wu, Z.; Yang, X. Electric Vehicle Charging Load Prediction Based on Dynamic Energy Consumption Model and User Psychology. *Mod. Electr. Power* **2022**, *39*, 710–719.
10. Long, X.; Yang, J.; Wu, F.; Zhan, X.; Lin, Y.; Xu, J. Prediction of Electric Vehicle Charging Load Considering Interaction Between Road Network and Power Grid and User's Psychology. *Autom. Electr. Power Syst.* **2020**, *44*, 86–93.
11. Zhang, M.; Sun, Q.; Yang, X. Electric Vehicle Charging Load Forecasting Considering Real-time Interaction of Multi-source Information and Consumer Regret Psychology. *Power Syst. Technol.* **2022**, *46*, 632–645.
12. Fang, Y.; Junhao, T.; Shenghua, C. Charging load prediction method for electric vehicles based on an ISSA-CNN-GRU model. *Power Syst. Prot. Control* **2023**, *51*, 158–167.
13. Lin, X.; Zhang, H.; Ma, Y.; Chen, L. Electric vehicle charging load prediction based on improved LSTM neural network. *Mod. Electron. Tech.* **2024**, *47*, 97–101.
14. Zhang, Y.; Liu, C.; Rao, X.; Zhang, X.; Zhou, Y. Electric Vehicle Charging Load Prediction Based on Dynamic Adaptive Graph Neural Network. *Autom. Electr. Power Syst.* **2024**, *48*, 86–93.
15. Zhang, Y.; Liu, C.; Rao, X.; Zhang, X.; Zhou, Y. Spatial-temporal load forecasting of electric vehicle charging stations based on graph neural network. *J. Intell. Fuzzy Syst.* **2024**, *46*, 821–836. [[CrossRef](#)]
16. Chuang, C. *Research on Electric Vehicle Charging Load Forecasting Based on Graph Convolutional Networks and Queuing Theory*; Harbin Institute of Technology: Shenzhen, China, 2022.
17. Hu, B.; Zhang, P.; Huang, E.; Liu, J.; Xu, J.; Xing, Z. Graph wavenet-based charging load forecasting of electric vehicle. *Autom. Electr. Power Syst.* **2022**, *46*, 207–213. (In Chinese)
18. Shen, X.; Zhao, H.; Xiang, Y.; Lan, P.; Liu, J. Short-term electric vehicles charging load forecasting based on deep learning in low-quality data environments. *Electr. Power Syst. Res.* **2022**, *212*, 108247. [[CrossRef](#)]
19. Feng, J.; Yang, J.; Li, Y.; Wang, H.; Ji, H.; Yang, W.; Wang, K. Load forecasting of electric vehicle charging station based on grey theory and neural network. *Energy Rep.* **2021**, *7*, 487–492. [[CrossRef](#)]
20. Zhang, J.; Liu, C.; Ge, L. Short-Term Load Forecasting Model of Electric Vehicle Charging Load Based on MCCNN-TCN. *Energies* **2022**, *15*, 2633. [[CrossRef](#)]
21. Li, Q.; Xiong, H. Short-term Load Prediction of Electric Vehicle Charging Stations built on GA-BPNN Model. *J. Phys. Conf. Ser.* **2023**, *2474*, 012023. [[CrossRef](#)]
22. Cheng, N.; Zheng, P.; Ruan, X.; Zhu, Z. Electric vehicle charging load prediction based on variational mode decomposition and Prophet-LSTM. *Front. Energy Res.* **2023**, *11*, 1297849. [[CrossRef](#)]
23. Mekkaoui, D.E.; Midoun, M.A.; Shen, Y. LA-RCNN: Luong attention-recurrent- convolutional neural network for EV charging load prediction. *Appl. Intell.* **2024**, *54*, 4352–4369. [[CrossRef](#)]
24. Wang, H.; Li, X. Research on Electric Vehicle Charging Load Forecasting Using an Improved Bidirectional Long Short-Term Memory Network with Fusion Attention Mechanism. *Electr. Power Clean Energy* **2022**, *38*, 104–112.
25. Shi, L.; Li, Y.; Liu, J.; Wang, Z.; Wang, W. Improved GRU Method for Ultra-Short-Term Charging Load Forecasting of Electric Vehicle Charging Stations. *Power Supply Consum.* **2023**, *40*, 42–47.
26. Gan, L.; Chen, F.; Sun, X.; Li, R.; Wang, C.; Xu, T. Load forecast of new energy vehicle charging stations based on QEM-Elman model. *J. Terahertz Sci. Electron. Inf. Technol.* **2022**, *20*, 378–384.
27. Guo, G. Load calculation and influence factors of electric vehicle charging station. *Electr. Eng.* **2019**, *20*, 93–97.
28. Ma, R.; Zhou, X.; Peng, Z.; Liu, D.; Xu, H.; Wang, J.; Wang, X. Data mining on correlation feature of load characteristics statistical indexes considering temperature. *Proc. CSEE* **2015**, *35*, 43–51.
29. Jiang, C. *Research of Thermal Characteristics of Battery Pack in Pure Electric Vehicle*; Hefei University of Technology: Hefei, China, 2015.
30. Vermaak, J.; Botha, E.C. Recurrent neural networks for short-term load forecasting. *IEEE Trans. Power Syst.* **1998**, *13*, 126–132. [[CrossRef](#)]
31. Shi, X.; Chen, Z.; Wang, H.; Yeung, D.Y.; Wong, W.K.; Woo, W.C. Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting. *Adv. Neural Inf. Process. Syst.* **2015**, *28*. Available online: <https://webofscience.clarivate.cn/wos/woscc/full-record/WOS:000450913102103> (accessed on 23 July 2024).
32. Chen, J. *Research on Operation Scheduling Strategy for Electric Vehicle Charging Station*; Southeast University: Nanjing, China, 2016.

**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.