Enhanced Load Forecasting for Electric Vehicle Charging Stations Using a Hybrid Random Forest-Convolutional Neural Network Algorithm

Meng Zhang
Vocational Skills and Training Center, Shanxi Vocational & Technical College of Finance & Trade, Taiyuan 030031, China
E-mail: z1216081662@163.com

Keywords: random forest algorithm, convolutional neural network, electric vehicles, charging station, load forecasting

Received: July 12, 2024

Aiming at the problems of low prediction accuracy and single data type in traditional electric vehicle charging station load prediction models, an RF-CNN algorithm based on the combination of random forest algorithm and convolutional neural network is proposed to improve the accuracy and efficiency of electric vehicle charging station load prediction. The research first takes the state of charge of electric vehicles and charging stations as the research object to construct a load prediction model. Then it utilizes the designed algorithm to optimize model performance. Finally, different validation indicators are used to verify the predictive performance of the prediction model in the load of charging stations. During the performance experiment, data from 15 public electric vehicle charging stations in the second quarter of 2022 were used. The experiment followed the steps of data preprocessing, feature selection, model training, and predictive evaluation. The specific results show that the RF-CNN algorithm achieves a prediction accuracy of 92.71%, significantly higher than BP's 85.16% and LSTM's 88.09%. The training loss is 11.34%, which is lower than the 38.06% of BP and 28.52% of LSTM. The average response speed is 42 milliseconds, demonstrating the efficiency of the algorithm. This indicates that the model has higher accuracy and robustness in load forecasting of electric vehicle charging stations, and could improve the operational efficiency of the power system. It has important application value in load analysis and prediction of electric vehicle charging stations.

Povzetek: Predlagan je nov algoritem RF-CNN, ki združuje lastnosti naključnih gozdov (RF) in konvolucijskih nevronskih mrež (CNN) za izboljšanje napovedi obremenitev električnih vozil in s tem optimiranje delovanja polnilnih postaj.

1 Introduction

With the increasingly serious environmental pollution problem, Electric Vehicles (EV) are widely promoted and used as a clean energy transportation tool. However, load analysis and forecasting of EV Charging Stations (CS) have become important topics for improving charging efficiency, optimizing charging strategies, and reducing energy waste. The load analysis and forecasting of EV CSs can help CS managers to arrange the use of charging facilities reasonably, improve charging efficiency, reduce user waiting time, and minimize the load pressure on the power grid [1-3]. However, the complexity of EV charging behavior and the unpredictability of EV user behavior, including factors such as charging time, charging time, and charging volume, pose challenges to load forecasting. In the related research on load analysis of EV CSs, a large number of new technologies have emerged, such as multi-source data integration methods, which can improve prediction accuracy by combining meteorological, traffic flow and other data. The reinforcement learning optimization strategy technology can learn the optimal charging strategy through the interaction between the agent and the environment. Deep learning algorithms have powerful feature extraction capabilities, which are suitable for processing large-scale spatiotemporal data. Although various load forecasting methods have been proposed in existing research, there are still some limitations, such as insufficient handling of uncertainty in EV user behavior and insufficient utilization of multi-source data advantages in predictive models. In response to these issues, the research innovatively combines Random Forest (RF) and Convolutional Neural Network (CNN) to design the RF-CNN algorithm. Compared with Long Short Term Memory (LSTM) networks, CNN is generally more computationally efficient when processing large amounts of data. The convolutional operations of CNNs are highly parallel and can be accelerated with modern computing hardware. In addition, when sliding across the entire input data, the parameters of the CNN convolution kernel are shared, which greatly reduces the number of model parameters, lowers the complexity of the model, and reduces the risk of overfitting. This algorithm combines the characteristics of RFs and the advantages of CNNs to effectively process load data of EV CSs, improving prediction accuracy and efficiency. RFs can select and combine input features, extract important features from load data, reduce the impact of noise, and improve prediction accuracy. CNNs can automatically learn the spatial and temporal features of load data through multi-layer convolution and pooling operations, further improving the accuracy of prediction [4-6]. Predicting the load of CSs will help improve their operational efficiency, further promote the development and application of EVs, and contribute to achieving clean energy transportation.

The necessity of research lies in the fact that with the rapid popularization of EVs, accurately predicting the load of CSs has become particularly important. This is crucial for optimizing power resource allocation, reducing grid pressure, improving charging efficiency, and enhancing user experience. However, existing research has limitations in dealing with the uncertainty and complexity of EV charging behavior, especially in the comprehensive utilization of multi-source data and model generalization ability. The novelty of the research lies in: (1) processing and integrating data from different sources, which improves the model's understanding of charging behavior. (2) Using RF for feature selection and combination, effectively extracting key features from load data. (3) Automatically learning the spatial and temporal characteristics of load data improves prediction accuracy through multi-layer convolution and pooling operations.

The first part of the study constructs a load forecasting model based on the SOC data. The second part optimizes the performance of the model using the RF-CNN algorithm based on the constructed model. The third part tests the function of the constructed model for comment classification through simulation experiments and practical applications. The fourth part summarizes the experimental results and analyzes the advantages and disadvantages of the research method.

2 Related works

With the popularization of EVs and the continuous development of power systems, the analysis and forecasting of CS loads have become increasingly important. Accurate load analysis and forecasting contribute to better power dispatch and optimized resource allocation in the power system. Recently, many scholars have carried out in-depth research in related fields using deep learning algorithms. To provide a more stable EV CS system, Savari et al. designed a real-time forecasting system based on IoT technology collection servers. The system took the cost and time of charging as the research content and was designed using the PHP programming language. The results indicated that the system could markedly enhance the system stability and load forecasting ability of existing CSs [7]. Zhang et al. effectively combined fuzzy control with RBF-NN forecasting model to provide accurate load forecasting for EV CSs. The online correction and performance improvement of fuzzy control were analyzed. The outcomes showcase that the forecasting model could markedly enhanced the load forecasting accuracy of a single model [8]. Zhu et al. proposed a comparison method based on deep learning to solve the load fluctuations during the charging process of plug-in EVs. This method utilized minute level real data from plug-in EV CSs to establish a forecasting model. Several other methods were compared to verify their predictive performance. The outcomes showcased that this method could markedly enhance forecasting performance at different time steps under 12 working conditions, with a maximum forecasting error reduction of 30% [9]. Li et al. proposed an assisted deep learning framework that combined short-term and short-term memory method to predict the charging power of EVs. The study first used a learning framework to capture predictive data, and then trained the data using short-term and short-term memory. The comparison between this learning framework and real charging data showed that it could significantly improve the forecasting performance of charging behavior [10].

Chung et al. designed an RF-based forecasting charging method to predict the charging behavior of EVs. This study collected historical charging data and energy consumption data. The defined entropy and sparsity ratio were used as evaluation indicators. Compared with non-coordinated charging, the combination of charging scheduling and algorithm forecasting could reduce peak load by 27%, load changes by 10%, and cost by 4% [11]. Wumaier et al. proposed a dynamic short-term Traffic Flow (TFL) forecasting method based on RF algorithm to solve traditional TFL forecasting. This method removed invalid data from the collected data, normalized available data, and completed data preprocessing before TFL forecasting. The outcomes showcased that the forecasting time of this method was relatively short, always less than 0.5 seconds. The forecasting accuracy was high, reaching 97% [12]. Huang et al. developed a hybrid model combined CNNs and variational modal decomposition to enhance effective regulatory capabilities and increase returns in power grid systems. This model collected historical data and extracted features from it, while inputting the data into the model for analysis and comparison. The outcomes demonstrated that the MAPE and RMSE of the model for the four seasonal averages were 0.730% and 0.453, respectively, indicating the feasibility and high accuracy of the model in predicting short-term electricity prices [13]. To effectively predict the electricity consumption in residential areas, Khan et al. designed a two-step method to predict the electricity load of residential buildings. This model first refined and trained the electricity consumption data, and then processed and analyzed the data using CNNs and multi-layer bidirectional gated loop units. The outcomes showcased that the model reduced the errors in the Individual Household Electricity Consumption Prediction dataset (IHEPC), namely RMSE (5%), MSE (4%), and MAE (4%), as well as the Power Load Prediction dataset, namely RMSE (2%) and MAE (1%) [14].

In summary, the research on load forecasting of EV CSs based on RF-CNN is of great significance. The combination of RF and CNN can effectively analyze and process the load data of CSs, and obtain load forecasting results. This study aims to provide stronger support for the EVs and enhance the development capacity of the new energy industry. The comparative analysis of related work and research methods is shown in Table 1.

Table 1: Comparison of related work and research methods

Authors & Year	Main technology	Key findings	Research method gap
Savari et al. [7]	IoT-based forecasting system	Enhanced system stability and load forecasting	Limited generalization due to lack of algorithmic details
Zhang et al. [8]	Fuzzy control with RBF-NN	Improved load forecasting accuracy	Miss the advantage of deep learning in feature extraction
Zhu et al. [9]	Deep learning comparison method	30% reduction in maximum forecasting error	Need further exploration of adaptability to charging patterns
Li et al. [10]	Assisted deep learning framework	Significantly improved predictive performance	Require optimization for large-scale datasets
Chung et al. [11]	RF-based predictive charging method	Reduced peak load and cost	Need further research on spatiotemporal dynamics of charging
Wumaier et al. [12]	Stochastic forest algorithm for traffic flow forecasting	High forecasting accuracy and speed	Need to verify the stability of real-world traffic flow
Huang et al. [13]	A hybrid model of CNN and variational mode decomposition	Low MAPE and RMSE for electricity price forecasting	Impact of market changes on forecasting needs further study
Khan et al. [14]	Two-step method for residential electricity load forecasting	Reduced prediction errors in datasets	Requires verification and adjustment for different building types

In the existing technology, although various methods have been proposed for load prediction of EV CSs, they still have some limitations. For example, some models may perform poorly in dealing with uncertain EV user behavior, or fail to fully utilize the advantages of multi-source data to improve prediction accuracy. In addition, existing technologies may lack computational efficiency and model generalization ability, which is particularly evident in large-scale datasets and dynamically changing charging environments. RF-CNN combines strong feature selection ability and robustness to noise, and further improves prediction accuracy by automatically learning spatial and temporal features of load data. Applying it to load prediction of EV CSs can effectively improve the prediction quality.

3 Load analysis and prediction model construction of EV CSs on the ground of RF-CNN algorithm

The load analysis and prediction model for EV CSs based on the RF-CNN algorithm is a deep learning algorithm. It can learn the characteristics and patterns of load from historical load data and used to predict future load situations.

3.1 Data definition of prediction model

When the load analysis and prediction model of EV CS is studied, the data involved in the model include charging state data, CS data, power data, time data, and environmental data. The charging state data refers to the charging state of the EV and reflects the current level of charge of the battery. CS data refers to the number of CSs, their location, and the status of currently available CSs. Power data refer to the charging power of the CS, including current power and historical power data. Time data refers to the specific time when the charging behavior occurs, including time characteristics such as hour, day, and week. Environmental data refers to environmental factors that may affect charging behavior, such as weather conditions, temperature, etc.

3.2 Design of load data processing method of electric vehicle charging station

To conduct load analysis on EV CSs, this study first collects and processes load data from the CSs. On this basis, the construction of the load model is completed, providing data support for the CS load prediction model. The constructed load model for EV CSs first requires first calculates the usage time of the charging pile, which can determine the operating rules of the CS, analyze the charging behavior of EVs, and optimize charging strategies. The missing values found in the dataset are processed through two main methods. For missing charging status data in the time series, linear interpolation or the average of the previous and subsequent data points is used for estimation. This method is based on the continuity of time series data, ensuring the integrity of the data and the accuracy of the prediction model. In order to eliminate the influence of different dimensions and scales, the Z-Score standardization method is used to standardize all data in the study. Z-Score standardization is achieved by calculating the difference between each data point and the mean, and then dividing it by the standard deviation to ensure that the mean of the data is 0 and the standard deviation is 1. For EV CSs, the charging behavior of EVs exhibits significant randomness, and there are also differences in charging time and power among different EVs. The study utilizes the State of Charge (SOC) to indicate the initial charge level of EVs, as this value is critical in influencing the charging duration and efficiency during the EV's charging session at the CS [15]. Meanwhile, the charging process of EVs is a constant current constant charge process. The charging duration can be expressed by formula (1).

$$t_c = T_{cc} + T_{cv} \tag{1}$$

In formula (1), T_{cc} represents the charging time during the constant current stage of the charging process. T_{cv} represents the charging time during the constant voltage stage of the charging process. The SOC is also affected by the charging duration, and the corresponding SOC can be represented by formula (2).

$$SOC(t_s) = SOC_0 + \frac{It_c}{C_{ap}} \tag{2}$$

In formula (2), SOC_0 represents the initial SOC of the EV battery. I represents the charging current value. C_{ap} represents the rated charging capacity of the EV battery. Numerous analyses have shown that the constant voltage charging time of EV batteries is very short throughout the entire charging process, accounting for only 1% of the total charging time. The processing of CS load data is shown in Figure 1.

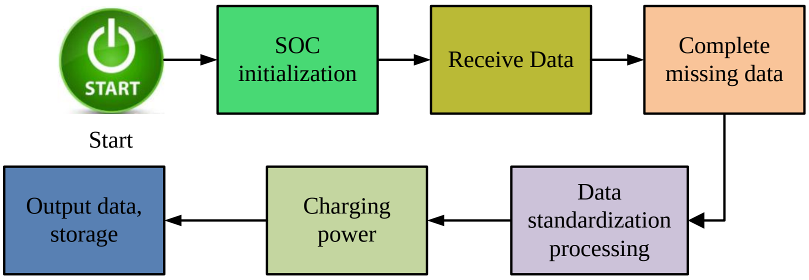


Figure 1: CS load data processing flow chart

Based on Figure 1 and the characteristics of constant current charging at CSs, this study only considers the constant current charging process of EVs during charging at CSs. That is, the study considers the constant current charging process as the entire charging time. During the data collection phase, some data may be missing for various reasons. Based on the adjacent data points of the time series, the missing value is estimated by linear interpolation or the average value of the data points before and after, and the missing data is completed. The corresponding charging time can be represented by formula (3).

$$t_c = \frac{(SOC_{end} - SOC_0) \cdot C_{ap}}{I_{cc}} \tag{3}$$

In formula (3), SOC_{end} represents the SOC of the battery after charging is completed. For ease of calculation, the final SOC is calculated as 100% to obtain the charging time of the EV.

3.3 Load Model construction of electric vehicle charging station

During the actual operation of the CS, there are different numbers of EVs charging simultaneously. Therefore, the study describes all charging vehicles in the CS according to the Poisson distribution. The charging process of each EV is an independent event, and the EVs charged at the CS follow an exponential distribution. At this time, the total charging vehicles during the total time period can be represented by formula (4).

$$S(t) = \begin{cases} \{s_1, s_2, s_3, \cdots, s_n\} \sim P(\lambda) \\ n = \dfrac{T_{total}}{T_{interval}} \end{cases} \tag{4}$$

In formula (4), s_n represents the sequence values of vehicles charged during n period. n is a continuous integer used to identify a specific time period, and $n = 1, 2, 3, \cdots$. P represents the probability of charging. λ represents the number of vehicles arriving at the CS per unit time. T_{interval} represents the unit time of vehicle charging. T_{total} represents the total charging time. In the actual operation of CSs, it may also not require queuing for charging. The start time of charging can be represented by formula (5).

$$T_{start-nm} = q \cdot T_{\text{interval}} + \frac{T_{\text{interval}}}{s_q} \cdot m \qquad (5)$$

In formula (5), q represents the time period serial number in the CS, which usually refers to a unique number used to identify a specific time period in data analysis, time series analysis, or any situation where data needs to be recorded in chronological order. s_q represents the quantity of EVs being charged during the q th time period. m represents the vehicle signal that reaches the CS in the q -th time series. The study obtains the charging data of all charging vehicles in the entire CS. Due to the large amount and complex types of charge data in the entire CS, to ensure the reliability of the data used for research, the charging time and charge state are processed in the study. The processing is divided into supplementing the collected data with missing information, and standardizing all data [16-17]. To ensure the integrity of the data, the study uses data with the same date and type to fill in missing data. The filling rules can be represented by formula (6).

$$\begin{aligned} &x(d,t) = \\ &\omega_1 x(d,t_1) + \omega_2 x(d,t_2) + \omega_3 x(d_1,t) + \omega_4 x(d_2,t) \end{aligned} \qquad (6)$$

In formula (6), $x(d,t)$ represents the charging load value at time t on day d in the CS. $x(d,t_1)$ and $\omega_2 x(d,t_2)$ respectively represent the charging load values before and after t on day d in the CS. $x(d_1,t)$ and $\omega_4 x(d_2,t)$ represent the time when the charging charge in the power station is the same before day d and after day d. ω represents the strengthening coefficient, with a value range of [1,2,3,4]. After completing the data filling, the data can be standardized, and Z-Score standardization is used to standardize the data. This standardization can convert data of various magnitudes into a unified Z-Score score for comparing. Standardization can be represented by formula (7).

$$Z_i = \frac{A_i - \mu}{\sigma} \qquad (7)$$

In formula (7), Z_i represents the normalized Z value of the i -th sample. A_i represents the i -th sample in the dataset. μ represents the mean of the total sample. σ represents the standard deviation of the total sample. In summary, the power load of the CS is obtained by adding up the charging power of all EVs entering the station for charging. Based on the actual operation of existing CSs, the losses of the entire CS can be basically ignored. Currently, the power batteries of EVs on the market are mainly lithium batteries. Therefore, the study applies the charging power of lithium batteries to construct load models. Due to only considering the constant current charging process during the charging, the charging power of the lithium battery can be represented by formula (8).

$$p_{batt}(t) = U_{batt}(t) I \qquad (8)$$

In formula (8), $p_{batt}(t)$ represents the charging power. $U_{batt}(t)$ represents the voltage value at the battery end. At this point, the charging power of the corresponding n -th charging vehicle can be represented by formula (9).

$$\begin{aligned} &P_{in}(k) = \begin{cases} P_{charge}(t - t_{start-ij} + t_0) \\ 0 \end{cases} \\ &t \in \left[0, t_{start-ij} + t_{charge} - t_0\right] \end{aligned} \qquad (9)$$

In formula (9), P_{charge} represents the charging power curve value. t represents the time point. $t_{start-ij}$ represents the time point corresponding to the start of charging. t_0 represents the time when charging begins. t_{charge} represents the completion time of charging. The total power value after superposition can be obtained, which can be represented by formula (10).

$$P_{EVstation}(k) = \sum_{1}^{n} P_{in}(k) \qquad (10)$$

In formula (10), $P_{in}(k)$ represents the charging power of the n -th charging vehicle. In summary, the study utilizes the total charging power of cars charged in CSs to obtain the load model of the entire CS. Figure 2 shows the flowchart for constructing a load model for EV CSs.

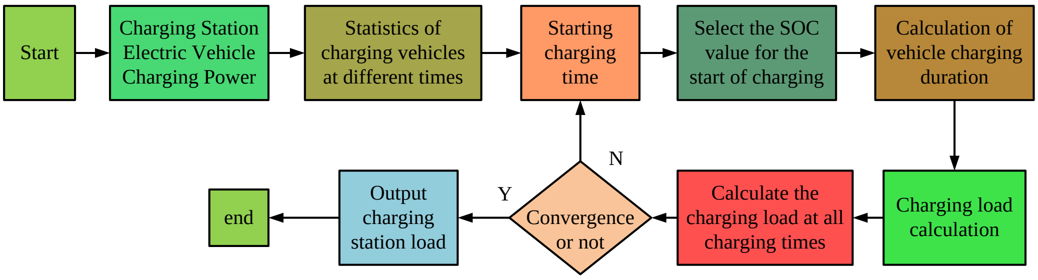


Figure 2: Flow chart for constructing load model of EV CS

3.4 Design of load forecasting model for EV CSs based on RF-CNN algorithm

The preliminary forecasting of data by the constructed CS load model indicates that although the model covers the required charging load data, its accuracy, evaluation ability, and generalization ability are weak. This leads to unsatisfactory predicted results. Therefore, this study combines RF and CNN to design an RF-CNN algorithm. This algorithm combines the characteristics of RF and the advantages of CNN to effectively process load data of EV CSs, improving prediction accuracy and efficiency. RF can select and combine input features, extract load data information from each CS, and use these information to construct a prediction spatiotemporal matrix, thereby improving prediction accuracy. CNN can automatically learn the spatial and temporal characteristics of load data through multi-layer convolution and pooling operations, predicting the CS load. The RF algorithm can use multiple decision trees to eliminate data features, which can accelerate the speed of the prediction model during the training process and improve performance [18-20]. The study uses the Gini index to measure the designed decision tree. The more branch points a decision tree has, the stronger its ability to reduce errors. The research sets the load forecasting data as a dataset. The accuracy of the dataset can be represented by the Gini index, as shown in formula (11).

$$Gini(D) = 1 - \sum_{k=1}^{y} p_k^2 \tag{11}$$

In formula (11), p_k represents the probability value of inconsistency in the k-th sample in the dataset. The smaller the Gini index of two randomly selected samples in the dataset, the higher the accuracy of the data in the dataset. To define the Gini index of a specific data in the dataset, the definition can be represented by formula (12).

$$Gini\_index(D,a) = \sum_{v=1}^{V} \frac{|D|^v}{|D|} Gini(D^v) \tag{12}$$

In formula (12), D represents the load forecasting dataset. a represents the characteristic value. v represents the value of a discrete attribute. The process diagram of RF algorithm decision-making is shown in Figure 3.

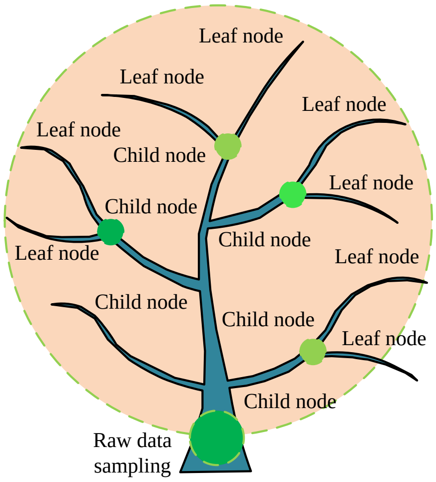


Figure 3: Process diagram of RF algorithm for decision-making

Figure 3 shows the decision-making process of the RF, which includes child nodes and leaf nodes. The child nodes represent the intermediate decision points for data segmentation in the decision tree. The leaf nodes represent the final prediction results. The process starts with the sampling of the original data. The data is continuously segmented through the child nodes until it reaches the leaf node representing the final decision. The branching structure in the figure shows that the data is divided into different paths at each child node according to the established rules, and finally reaches the leaf node. RF sets the number of decision trees in the forest to 100, with a maximum depth of 20 per tree, to prevent overfitting. In addition, Gini impurity is used as a indicator to measure the quality of splitting. CNN can process high-dimensional and nonlinear data. The maximum pooling operations are performed to handle convolutional operations in CNN. To ensure consistency between the input and output of the data, the maximum element value is selected in the downsampling layer to replace the output value at that location. This not only ensures data consistency, but also enhances sensitivity to feature data information. The maximum pooling function introduced in the maximum pooling operation can be represented by formula (13).

$$O(b,c) = Max(\beta(b,c)) \quad (13)$$

In formula (13), $O(b,c)$ represents the matrix in the CNN that has undergone maximum pooling output. $\beta(b,c)$ represents the region with pooling in the corresponding matrix. $Max(\cdot)$ represents the element value output by the function after maximizing the input elements in the CNN. The schematic diagram of maximum pooling in CNN is shown in Figure 4.

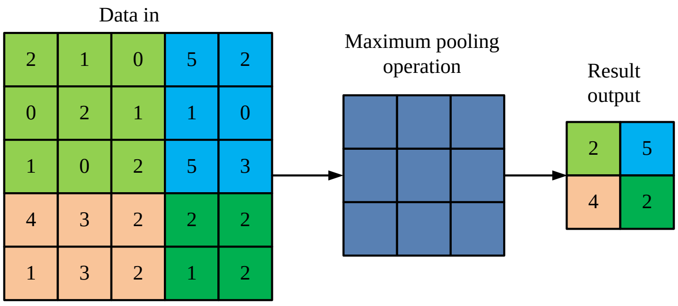


Figure 4: Schematic diagram of maximum pooling in CNN

In the CNN, the convolutional layer uses 32 3×3 convolutional kernels, followed by a 2×2 max pooling layer. Before the fully connected layer, ReLU activation function is used for nonlinear transformation. In summary, the study aims to construct a CS load forecasting model by integrating RF and CNN. The specific construction process is as follows. First, it uses the RF method to predict the load of all EVs in the CS. Then, it uses RF to predict the load of each charging EV. The collected charging load forecasting values are input into the corresponding spatiotemporal matrix, which can be represented by formula (14).

$$X = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,n} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m,1} & x_{m,2} & \cdots & x_{m,n} \end{bmatrix} \quad (14)$$

In formula (14), $x_{m,n}$ represents the charging composite data values corresponding to positions m and n in the matrix. The spatiotemporal matrix for predicting charge values is input into the CNN. The convolutional layer throughout the entire CNN will continuously convolution the size of the matrix. When the requirements are met, the changes will be transmitted to the pooling layer. The pooling layer utilizes maximum pooling operations to expand the pooled channels. This will significantly increase the amount of data that can reach the fully connected layer. By extracting and weighting these data features, the average value of the fully connected layer can be calculated. The average output of the fully connected layer is the predicted load value of the EV CS. The flow chart of the load forecasting model for EV CSs based on the RF-CNN is shown in Figure 5.

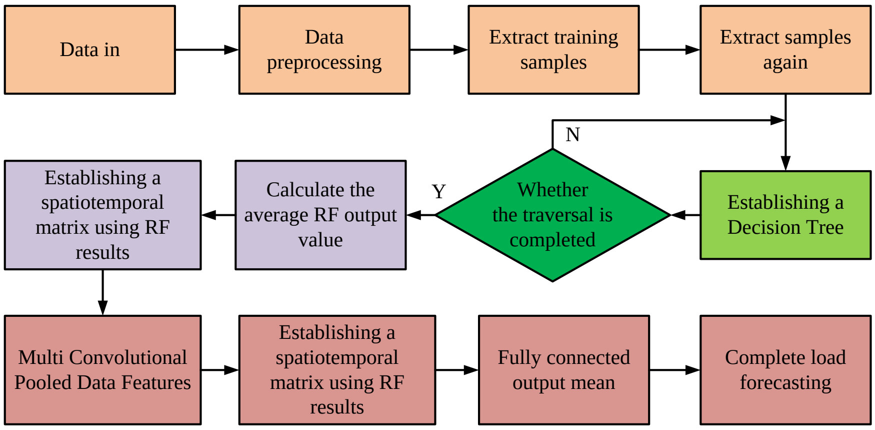


Figure 5: Flow chart of EV CS load forecasting model on the Ground of RF-CNN algorithm

The diversity of initial SOCs has been considered in the model. By analyzing a large number of charging data under different initial SOCs, the model learns and predicts the load changes of CSs under different charging demands. The influence of different initial SOC on load forecasting of CS is mainly reflected in two aspects: charging time and charging power. The research method can automatically extract the important features of the load data, and effectively reduce the influence of noise. Therefore, the load forecasting task of EVs can be effectively completed. In practical operation, the research method first collects basic data, and then cleans and standardizes the collected data. Based on the characteristics of charging behavior, key features are extracted from raw data to provide input for model training. A load forecasting model is constructed and trained by cross-validation method. The model structure is optimized by adjusting hyper-parameters and using Adam optimizer to minimize the loss function. The trained model is deployed to simulated or actual EV CS environments for real-time load forecasting and charging scheduling.

4 Load analysis and forecasting model performance analysis of EV CSs

For verifying the performance of the load forecasting model, data collection is conducted on a certain CS. The collected data are used to construct a dataset for model performance verification.

4.1 Performance analysis of CS load forecasting model

To analyze the performance of the CS load forecasting model, a comparison is conducted between the Back Propagation Network (BP), LSTM, and RF-CNN to verify the performance of the forecasting model. The dataset used in the study comes from 15 public EV CSs in Nanjing, Jiangsu Province, China. The data collection took place in the second quarter of 2022 and covered daily operational data from the CSs. The specific data includes the SOC of the EV, the number of CSs, the charging power, and the geographical location of the CSs. The prediction accuracy of SOC data is used to validate the model's ability to collect data. Figure 6 compares the prediction accuracy of SOC data using three methods.

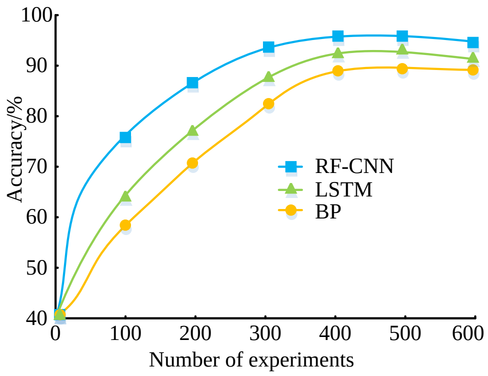
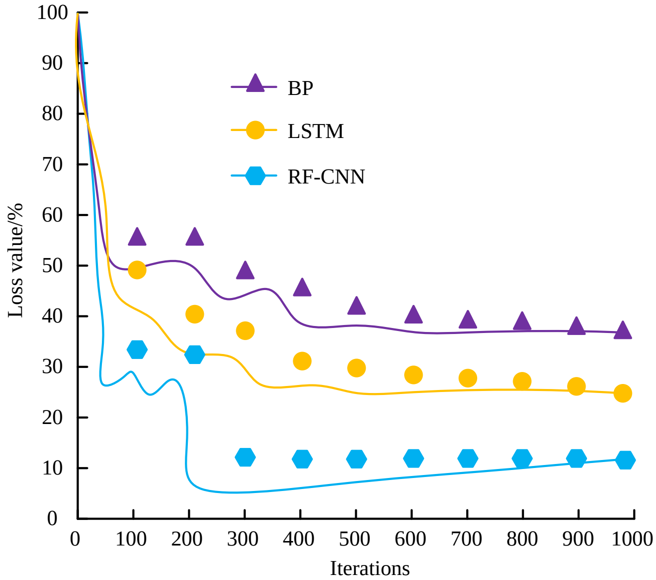
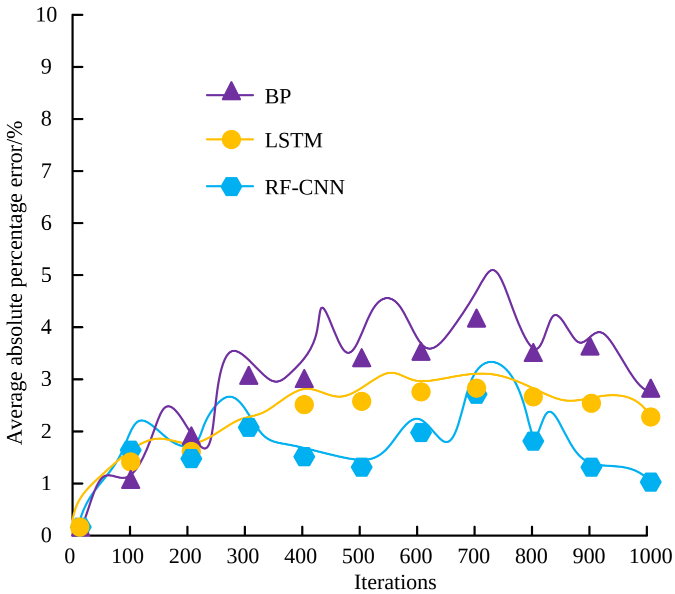


Figure 6: Comparison results of three methods for predicting accuracy of SOC data

In Figure 6, as the number of experiments increases, the accuracy of SOC data prediction for the three methods first increases and then tends to stabilize. The SOC prediction accuracy of RF-CNN, LSTM, and BA was 92.71%, 88.09%, and 85.16%, respectively. To verify the data processing ability of the RF-CNN forecasting model, the study uses training loss value and average absolute percentage error as indicators. The comparison results of the training loss values and average absolute percentage errors of the three methods are shown in Figure 7.



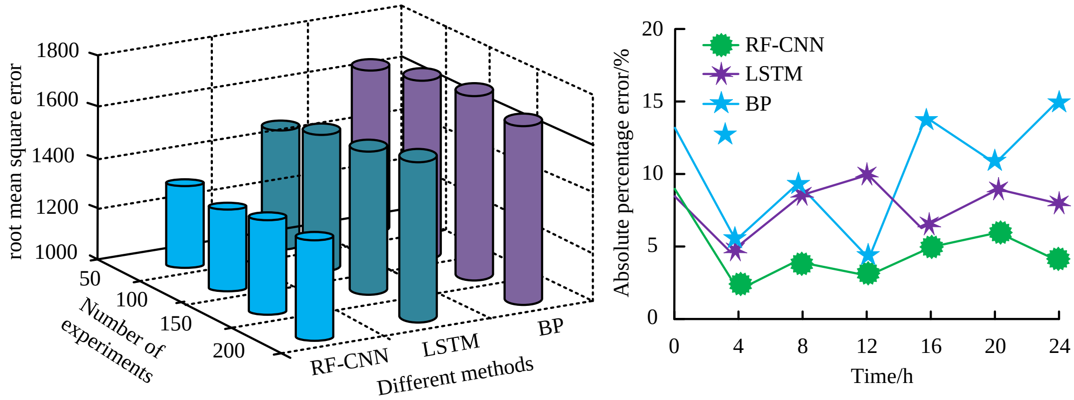
(a) Three models for training loss functions



(b) Average absolute percentage error

Figure 7: Comparison results of training loss values and average absolute percentage errors of three methods

In Figure 7 (a), as the training iteration progresses, the difference in loss values among the three methods also increases. The training loss values of RF-CNN, LSTM, and BA were 11.34%, 28.52%, and 38.06%, respectively. In Figure 7 (b), the average absolute percentage errors of the three methods were 1.82%, 2.16%, and 3.61%, respectively. This indicates that the RF-CNN used in the study has better performance in processing CS data. To further test the processing ability of RF-CNN in CS data, the study also uses root mean square error and data extraction error as validation indicators. The comparison outcomes of the root mean square error and data extraction error of the three methods are shown in Figure 8.

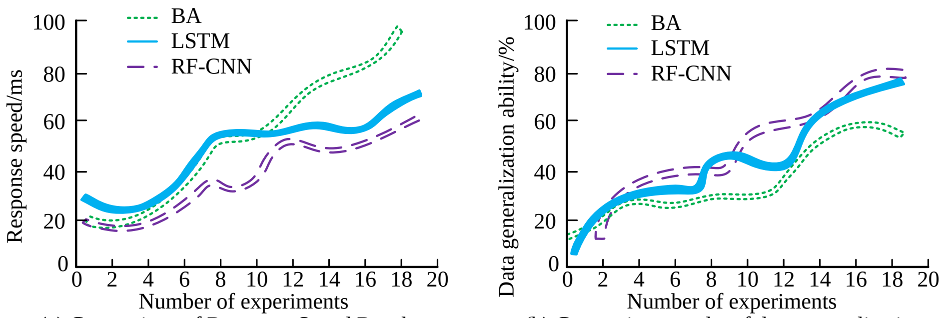


(a) Comparison results of root mean square error of three methods

(b) Comparison results of root mean square error of three methods

Figure 8: Comparison results of root mean square error and data extraction error of three methods

In Figure 8 (a), there was a certain difference in the root mean square error among the three methods, with RF-CNN having the smallest root mean square error, followed by LSTM and BP. The root mean square errors of RF-CNN, LSTM, and BA were 1261, 1428, and 1608, respectively. In Figure 8 (b), there were also certain differences in extraction errors among the three methods in the process of extracting charging data. The BP with the highest extraction error and the RF-CNN with the lowest error indicate that the RF-CNN model constructed in the study has good robustness. To verify the performance of the forecasting model in predicting load, the response speed and data generalization ability of the model are used as indicators. The comparison results of the response speed and data generalization ability of the three methods are shown in Figure 9.



(a) Comparison of Response Speed Results of Three Methods

(b) Comparison results of data generalization ability among three methods

Figure 9: Comparison results of response speed and data generalization ability of three methods

In Figure 9 (a), all three methods had certain efficiency in predicting charges. The faster the response speed, the better the performance in the forecasting process. The average response speeds of RF-CNN, LSTM, and BA were 42ms, 58ms, and 83ms. In Figure 9 (b), the data generalization ability of the RF-CNN, LSTM, and BA methods was 83.61%, 69.84%, and 41.59%, respectively. This indicates that the detection efficiency and resource optimization allocation ability of the RF-CNN forecasting model are also significantly superior to comparison methods. The other performance comparisons of the research method are shown in Table 2.

Table 2: Other performance comparisons

Performance metric	RF-CNN model	BP model	LSTM model
F1 score	0.91	0.8	0.85
Precision	0.92	0.83	0.88
Recall	0.9	0.79	0.84

AUC	0.95	0.87	0.91

As shown in Table 2, the F1 score of the RF-CNN model reached 0.91, with a precision of 0.92, a recall rate of 0.90, and an AUC as high as 0.95, indicating that the model has high accuracy and reliability in load forecasting tasks. In contrast, although the BP model and LSTM model also perform well, they are lower than the RF-CNN model in all indicators. This further proves the superior performance of the research method.

4.2 Application performance analysis of CS load forecasting model

To further validate the application performance of the CS load prediction model, this study compares the predicted results of RF-CNN, LSTM, and BA with the actual values. Figure 10 displays the comparison outcomes between predicted values and actual values.

In Figure 10, the true load value of the CS ranged from [151.31, 237.59] kW. The load prediction ranges of RF-CNN, LSTM, and BA were [152.05, 231.81] kW, [143.49, 245.01] kW, and [157.53, 265.07] kW, respectively. The load forecasting value closest to the actual value is RF-CNN, with a difference of 0.74kW and 5.78kW between the minimum and maximum values. The trend of the prediction curve is also roughly the same. This indicates that the EV CS forecasting model constructed using the RF-CNN algorithm has good predictive ability. To verify the predictive ability of the forecasting model at different time periods, this study conducts load forecasting for the CS in mid April, May, June, and July, and compares the predicted values with the actual values of the day. Figure 11 shows the comparison outcomes between the predicted value and the actual values.

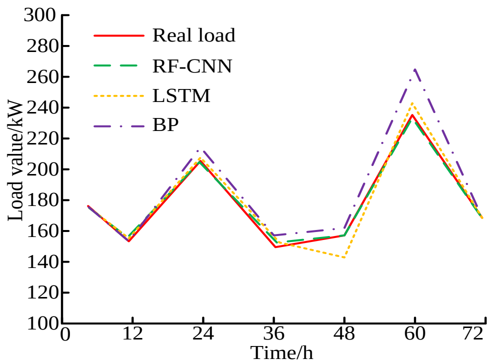
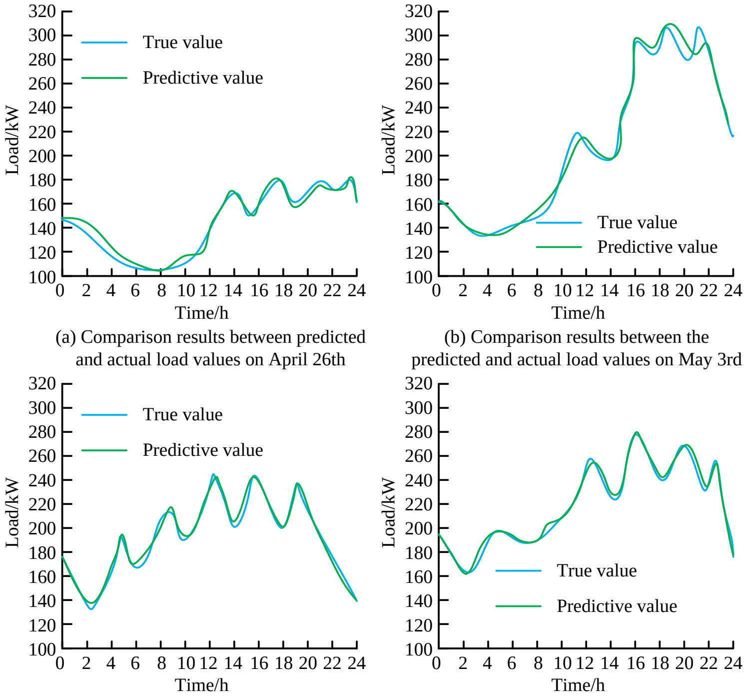


Figure 10: Comparison results between predicted values and true values using three methods



(a) Comparison results between predicted and actual load values on April 26th

(b) Comparison results between the predicted and actual load values on May 3rd

(c) Comparison results between the predicted and actual load values on June 26th

(d) Comparison between predicted and actual load values on July 28th

Figure 11: Comparison results between predicted values and true values of the RF-CNN model

Figures 11 (a), (b), (c), and (d) show that there are significant differences in the load of CSs on different dates. Among them, May 3rd was a holiday, with a large number of people traveling and more EV mileage, which increased the need for charging. The other three days are all working days, and the load of the CS wa relatively small. The difference between the predicted value and the true value of the forecasting model constructed during the research was very small. The load curve was the same. This indicates that the forecasting model constructed in the research can effectively predict the load of CSs. To further validate the performance of the RF-CNN forecasting model, this study conducts research on different charging methods for EVs. Disordered charging, positive sequence valley time charging, reverse sequence valley time charging, and sequential valley time charging are taken as indicators. Figure 12 shows the predicted results of four charging methods in EV CSs.

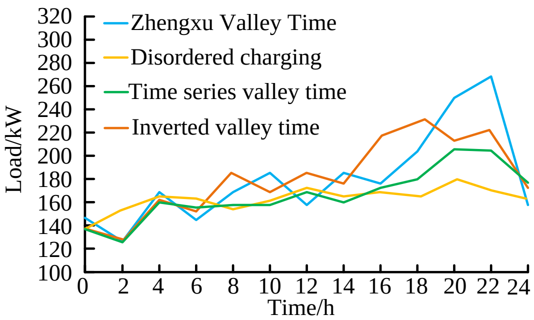


Figure 12: Prediction results of four charging methods in EV CSs

In Figure 12, there was no specific time limit for disordered charging, and the cost was paid based on the actual usage time, resulting in higher charging cost. Therefore, the load variation of the CS under this charging method is relatively small. Positive sequence valley time charging is carried out during the low electricity consumption period of the power system. Due to the relatively low charging cost during this stage, the CS load is relatively large. Reverse valley charging is carried out during peak periods of power consumption in the power system, and the charging cost is relatively high under this charging method. Therefore, the quantity of EVs charged at the charging point is relatively small, and the load change is not significant. Time series valley charging is carried out according to the user's usage habits and needs. Under this charging method, the CS load is relatively stable and does not change much. In summary, the RF-CNN model constructed in the study can effectively predict the specific CSs load, indicating that the forecasting model has high practicality. The results are shown in Table 3.

Table 3: Comparison of all results

Performance index	RF-CNN	BP	LSTM

Prediction accuracy (%)	92.71	85.16	88.09
Training loss (%)	11.34	38.06	28.52
Average percentage error (%)	1.82	3.61	2.16
Response speed (ms)	42	83	58
Data generalization capability (%)	83.61	41.59	69.84

From Table 3, RF-CNN predicted accuracy, training loss, average percentage error, response speed, and data generalization. The performance is better than BP and LSTM. In order to further evaluate the robustness of the model, a sensitivity analysis is conducted to explore the impact of different parameters on the model performance. By simulating different numbers of CSs, as the number of CSs increases, the prediction accuracy of the model slightly decreases. However, due to the integrated nature of the RF-CNN algorithm, this impact is limited and the decrease remains within 5%. The change in charging power has a direct impact on the model performance. When there is a significant fluctuation in the charging power data, the prediction accuracy of the model may decrease. By adjusting the model parameters, it can quickly recover above 98%. Time factors, such as the different time distributions of charging behavior throughout the day, have a significant impact on model predictions. The model performs better in handling data during peak hours, which may be related to the standardization of charging demand.

5 Discussion

The RF-CNN model proposed in the study shows significant performance improvement in load forecasting of EV CSs compared with existing technologies. The RF-CNN model achieved a prediction accuracy of 92.71%, significantly higher than the 85.16% and 88.09% of BP and LSTM models, and also superior to methods such as Fuzzy control based on RBF-NN and assisted deep learning frameworks. This difference may be attributed to the fact that RF-CNN combines the feature selection ability of RF with the spatiotemporal feature learning ability of CNN, effectively improving the model's ability to capture complex charging behaviors. The novelty of the RF-CNN model lies in its innovative algorithm design, which improves the accuracy and robustness of load forecasting for CSs by integrating two powerful learning mechanisms. The scalability of the RF-CNN model is particularly important in large-scale deployment across multiple CSs. The model can adapt to charging networks of different scales, improve processing capabilities through distributed computing, and achieve real-time analysis and prediction of CS data with wide geographical distribution. In large-scale deployment, the distributed architecture of the RF-CNN model allows for the simultaneous processing of data from different geographical locations, reducing the latency in data transmission and processing. Compared with traditional models, it has significant advantages in computational efficiency.

6 Conclusion

The study solved the single data and low accuracy in traditional CS load forecasting models by predicting the load of charging vehicle CSs. This study aims to use SOC as the data collection object and combine RF with CNN to improve the performance of the model. The results indicated that the predicted value range of the forecasting model for the CS was [152.05, 231.81] kW, and the true load range of the CS was [151.31, 237.59] kW. The difference between the minimum and maximum predicted values of the model and the actual values was 0.74kW and 5.78kW. Meanwhile, the forecasting model also accurately predicted the load of CSs at different periods and charging methods, and the trend was very close to the true value. This indicates that the RF-CNN forecasting model has better generalization performance and adaptability, which can better capture the complexity and uncertainty of load changes. In summary, the forecasting model constructed can assist EV CS managers in load analysis and prediction. It can accurately predict future load conditions and provide valuable references for the planning and scheduling of CSs.

Future research can apply the model to CSs under more diverse geographical and climatic conditions to validate and enhance the model's generalization ability and adaptability. Meanwhile, integration methods for real-time data streams can also be explored to enable the model to respond to real-time changes in charging demand and provide support for dynamic charging scheduling.

Table 4: Nomenclature table

Full name	Abbreviation
Charging Station	CS
Electric Vehicle	EV
Random Fores	RF
Convolutional Neural Network	CNN

Long Short-Term Memory	LSTM
Traffic Flow	TFL
Individual Household Electricity Consumption Prediction	IHEPC
State of Charge	SOC
Back Propagation	BP

References

[1] Y. Shi, Z. Liu, D. Feng, Y. Zhou, K. Y. Zhang, and H. J. Li, “Electric vehicle charging situation awareness for ultra-short-term load forecast of charging stations,” Journal of Shanghai Jiaotong University (Science), vol. 28, no. 1, pp. 28-38, 2023. https://doi.org/10.1007/s12204-023-2566-2

[2] D. Jaraniya, and S. Kumar, “Integration of charging station to the smart grid power network by using robust immune feedback adaptive and Rao optimization approaches,” International Journal of Numerical Modelling: Electronic Networks, Devices and Fields, vol. 36, no. 5, pp. 30-45, 2023. https://doi.org/10.1002/jnm.3045

[3] G. T. Guo, “Load calculation and influence factors of electric vehicle charging station,” Electrical Engineering, vol. 20, no. 3, pp. 93-97, 2019. https://doi.org/CNKI:SUN:DQJS.0.2019-03-023

[4] H. Wang, Y. Zhang, and H. Mao, “Charging load forecasting method based on instantaneous charging probability for electric vehicles,” Electric Power Automation Equipment, vol. 39, no. 3, pp. 207-213. https://doi.org/10.16081/j.issn.1006-6047.2019.03.033

[5] S. Lai, J. Qiu, Y. Tao, and J. Zhao, “Pricing for electric vehicle charging stations based on the responsiveness of demand,” IEEE Transactions on Smart Grid, vol. 14, no. 1, pp. 530-544, 2022. https://doi.org/10.1109/TSG.2022.3188832

[6] M. Akil, E. Dokur, and R. Bayindir, “Modeling and evaluation of SOC-based coordinated EV charging for power management in a distribution system,” Turkish Journal of Electrical Engineering and Computer Sciences, vol. 30, no. 3, pp. 678-694, 2022. https://doi.org/10.3906/elk-2105-100

[7] G. F. Savari, V. Krishnasamy, J. Sathik, Z. M. Ail, and S. H. E. Abdel Aleem, “Internet of Things based real-time electric vehicle load forecasting and charging station recommendation,” ISA Transactions, vol. 97, no. 6, pp. 431-447, 2020. https://doi.org/10.1016/j.isatra.2019.08.011

[8] J. Zhang, C. Liu, and L. Ge, “Short-term load forecasting model of electric vehicle charging load based on MCCNN-TCN,” Energies, vol. 15, no. 7, pp. 26-33, 2022. https://doi.org/10.3390/en15072633

[9] J Zhu, Z. Yang, M. Mourshed, Y. Guo, and S. Feng, “Electric vehicle charging load forecasting: A comparative study of deep learning approaches,” Energies, vol. 12, no. 14, pp. 87-92, 2019. https://doi.org/10.3390/en12142692

[10] Y. Li, S. He, Y. Li, L. Ge, S. Lou, and Z. Zeng, “Probabilistic charging power forecast of EVCS: Reinforcement learning assisted deep learning approach,” IEEE Transactions on Intelligent Vehicles, vol. 8, no. 1, pp. 344-357, 2022. https://doi.org/10.1109/TIV.2022.3168577

[11] Y. W. Chung, B. Khaki, T. Li, C. Chu, and R. Gadh, “Ensemble machine learning-based algorithm for electric vehicle user behavior prediction,” Applied Energy, vol. 254, no. 15, pp. 1-12, 2019. https://doi.org/10.1109/GlobConHT56829.2023.10087780

[12] H. Wumaier, J. Gao, and J. Zhou, “Short-term forecasting method for dynamic traffic flow based on stochastic forest algorithm,” Journal of Intelligent and Fuzzy Systems, vol. 39, no. 2, pp. 1501-1513, 2020. https://doi.org/10.3233/JIFS-179924

[13] C. J. Huang, Y. Shen, Y. H. Chen, and H. Chen, “A novel hybrid deep neural network model for short-term electricity price forecasting,” International Journal of Energy Research, vol. 45, no. 2, pp. 2511-2532, 2021. https://doi.org/10.1002/er.5945

[14] Z. A. Khan, A. Ullah, W. Ullah, S. Rho, M. Lee, and S. W. Baik, “Electrical energy prediction in residential buildings for short-term horizons using hybrid deep learning strategy,” Applied Sciences, vol. 10, no. 23, pp. 1-15, 2020. https://doi.org/10.3390/app10238634

[15] A. Almaghrebi, F. Aljuheshi, M. Rafaie, K. James, and M. Alahmad, “Data-driven charging demand prediction at public charging stations using supervised machine learning regression methods,” Energies, vol. 13, no. 16, pp. 31-42, 2020. https://doi.org/10.3390/en13164231

[16] I. Rizianiza, and D. M. Shoodiqin, “Prediction of State of Charge (SOC) of battery electric vehicle,” Advances in Science and Technology, vol. 104, no. 11, pp. 65-71, 2021. https://doi.org/10.4028/www.scientific.net/AST.104.65

[17] D. Yan, and Y. Chen, “A distributed online algorithm for promoting energy sharing between EV charging stations,” IEEE Transactions on Smart Grid, vol. 14, no. 2, pp. 1158-1172, 2022. https://doi.org/10.1109/TSG.2022.3203522

[18] R. K. Agrawal, F. Muchahary, and M. M. Tripathi, “Ensemble of relevance vector machines and boosted trees for electricity price forecasting,” Applied Energy, vol. 250, no. 2, pp. 540-548, 2019. https://doi.org/10.1016/j.apenergy.2019.05.062

[19] X. Guo, Q. Zhao, and D. Zheng, “A short-term load forecasting model of multi-scale CNN-LSTM hybrid neural network considering the real-time electricity price,” Energy Reports, vol. 6, no. 1, pp. 1046-1053, 2020. https://doi.org/10.1016/j.egyr.2020.11.078

[20] Y. Ishii, “Electricity demand and supply curve estimation with neural network,” IEEJ Transactions on Power and Energy, vol. 141, no. 5, pp. 384-390, 2021. https://doi.org/10.1541/ieejpes.141.384