International Journal of Mathematics Trends and Technology
ISSN: 2231-5373/ https://doi.org/10.14445/22315373/IJMTT-V71I3P104
Volume 71 Issue 3, 33-53, March 2025

Original Article

Research on New Energy Power Generation based on Random Forest and LSTM

Xinlong Zhang

School of Mathematics and Information Science, Henan Polytechnic University, Jiaozuo, Henan Province, China.

[1]*Correspongding Author : 212210010010@home.hpu.edu.cn*

Received: 10 January 2025 Revised: 21 February 2025 Accepted: 11 March 2025 Published: 30 March 2025

Abstract *- In the context of globalization, climate change and environmental issues have received increasing attention. For a long time, relying on traditional fossil fuels for electricity production has led to significant greenhouse gas emissions, one of the key factors contributing to global warming. New energy power generation has become important to address global warming and promote sustainable energy development. With the progress of science and technology and the expansion of production scale, the utilization of new energy will improve the conversion efficiency of energy resources and reduce production and conversion costs. This study aims to utilize the methods of Random Forest and Long Short-Term Memory (LSTM) neural networks to research new energy generation power. By collecting and analyzing the historical data of wind power generation and solar power generation, a prediction model based on Random Forest and LSTM neural network was established. The experimental results show that this model can accurately predict the changing trends of new energy generation power. For example, when predicting solar power generation power, the model's accuracy can reach 96.72 percent and 97.37 percent, effectively reducing the prediction errors. Through in-depth research on new energy generation power, we can better understand the utilization potential of renewable energy, provide decision-making support for power grid dispatching and energy planning, promote the development of sustainable energy, improve the power generation efficiency of the power grid, and ensure the safety and stability of the power grid.*

Keywords *- Random Forest, Long Short-Term Memory (LSTM), New Energy, Generation Power Prediction.*

1. Introduction

1.1. Research Background and Significance

With the growing global concern for climate change and environmental protection, China has actively fulfilled its international responsibilities in addressing climate change, demonstrated its responsibility as a major country, actively responded to global climate governance, promoted the construction of a community with a shared future for mankind, and strived to achieve the goals of carbon peaking and carbon neutrality. In the context of global energy transformation, wind power generation and photovoltaic power generation [1-3], as clean and renewable energy sources, have gained favor from various countries and regions due to their characteristics. Carbon emissions can be further reduced in the development path of carbon peaking and carbon neutrality by improving the efficiency and scale of new energy power generation. At the same time, as the key driving force of green transformation, new energy power generation is of great significance in shaping green production and lifestyle and promoting high-quality development. New energy is green energy, which will play an important role in reducing carbon emissions and promoting the transformation of energy structure. These new energy generation methods are clean and environmentally efficient and reduce their dependence on fossil energy, which is renewable and inexhaustible. We will promote transforming and upgrading the energy structure and optimize the supply-side structural reform. New energy power generation is also beneficial to energy security. Since new energy is inexhaustible, as long as there is sunlight and wind, electricity can be continuously generated, which can reduce the risk of energy supply and improve energy security. Vigorously developing new energy not only helps to enhance the competitiveness of the domestic industrial chain but also achieves synergistic effects of pollution reduction and carbon reduction, improves energy security, promotes economic development, and realizes sustainable development.

Against the backdrop of the new normal of the global economy and the scientific and technological revolution, the traditional productivity structure is undergoing fundamental changes. The wide application of emerging information, science, and technology has promoted the leap of productivity to quality and the expansion of quantity, so the concept of new quality productivity came into being. It represents that the productive forces have entered a new stage and are a major supplement to the development of traditional theory. New quality productivity emphasizes the use of artificial intelligence, big data, and other new technologies to enable traditional industries to optimize production, distribution, circulation, and consumption comprehensively and maximize the efficiency of resource allocation. With the improvement of environmental protection awareness and the adjustment of energy structure, China's power structure will develop in a cleaner, more efficient and renewable direction. Currently, China has adopted the energy structure of thermal power generation and continues to increase the diversified power generation methods such as hydraulic power, nuclear energy, new energy and so on. Firstly, new energy power generation, such as solar, wind, and hydropower, are all green and low-carbon energy forms. The advantages of new energy power generation lie in its clean, low-carbon, and renewable characteristics. Solar photovoltaic power generation uses solar cells to convert light energy into electrical energy, which is clean and efficient; wind power generation converts the kinetic energy of the wind into mechanical kinetic energy and then into electrical energy, with the advantages of being renewable and pollution-free. Secondly, new energy power generation also has significant economic advantages. At present, the utilization efficiency of new energy is high; for example, the utilization efficiency of solar power generation can reach more than 30 percent, while that of traditional thermal power generation is only about 30 percent. Using new energy will greatly improve efficiency and save costs. At the same time, with the improvement of photovoltaic and wind power generation technology, its power generation accounts for a substantial increase in the total power generation of the system.

China has rich resources and a broad market in these two fields, but it also faces many challenges in technology and economy that need to be addressed urgently. According to the twin goals of carbon peak and carbon neutrality, China's new energy industry is developing rapidly, and fields such as photovoltaic and wind power have shown strong momentum. Since the twin goals of carbon peak and carbon neutrality were put forward, the installed capacity of wind power and photovoltaic has exceeded 100 million kilowatts for three consecutive years, reaching a new historical high. Among them, distributed photovoltaics have developed rapidly, with an annual growth rate of 47.68 percent, exceeding the growth rate of centralized photovoltaics by nearly 20 percent. Since 2021, the newly connected grid scale of distributed photovoltaic has exceeded that of centralized photovoltaic, and the two show a parallel development trend.

1.2. Research Status

In recent years, new energy power generation in China has developed rapidly. However, due to the limitations of technology, cost and other factors, its proportion in the power structure is still relatively small. The construction of new energy power generation infrastructure is not yet perfect. Insufficient investment in technology research and development has led to fragile infrastructure, difficulties in breaking through key core technologies, and the dependence of new energy power generation equipment on imports. In addition, natural factors such as weather and temperature greatly affect new energy power generation, resulting in an unstable power supply. In the new energy power generation field, especially in the practical application of wind and solar power generation, challenges such as insufficient prediction accuracy, complex resource allocation, and difficult efficient control are often faced. These problems have greatly increased the risk to the safety and stability of the power grid. To ensure the smooth operation of power dispatching and effectively reduce the daily maintenance costs of the power grid, it is urgent to strengthen the improvement of the safety and stability of the power grid.

On the other hand, more and more new energy sources are connected to the power system, and the power grid operation faces the problems of peak regulation imbalance between supply and demand and consumption. New energy power prediction has become a key technology to solve these problems. Although existing research mainly focuses on the power prediction of centralized power stations, there are few research results on the power prediction of regional power plant clusters. The accuracy of the prediction systems put into operational use still needs to be improved, and they cannot meet the requirements of relevant dispatching and trading services. Therefore, studying high-precision new energy power prediction applications is of great significance.

The prediction of new energy power generation can be subdivided into multiple time-period categories according to the period of the prediction and the control strategies adopted, including ultra-short-term prediction, short-term prediction, medium-term prediction, and long-term prediction. Each category serves specific management and optimization needs. There are many methods for short-term wind power and solar power prediction.

Wavelet algorithms, Support Vector Machines (SVM), Markov chains, Wavelet transform LS-SVM[4], etc, are all used to conduct short-term predictions based on the existing wind and solar power generation data. Due to the intermittency and volatility of renewable energy, its power generation is greatly affected by natural, climatic, and geographical factors such as weather and seasons. The randomness, disturbance, and intermittency of wind farms have a huge impact on the voltage stability, transient stability, safety, and power quality of the power grid.

Conducting short-term or long-term predictions of power generation is the main way to solve this problem. Through long-term prediction, the power generation of renewable energy can be more accurately evaluated, providing strong support for power grid dispatching, reducing the system risks caused by the fluctuations of renewable energy power generation, helping the power grid dispatching department to understand the changing trends of future power demand in advance, so as to reasonably arrange power generation plans and resource allocation, and ensuring the stability and reliability of power supply. Short-term power generation prediction cannot predict the power generation in a relatively long-term future, usually covering trends in several days, weeks, or even months. Therefore, long-term prediction is of great significance for the stable operation of the power system, optimal resource allocation, and economic decision-making.

Domestic and foreign scholars have carried out many studies on the prediction of new energy power generation represented by wind energy and solar energy. Blonbou R [5]used adaptive Bayesian learning and neural network methods to predict wind power generation. Zhe S, Jiang Y, and Zhang Z [6] used Bayesian and Markov switching models for prediction. At the same time, the support vector machine (SVM) is also a popular algorithm. Yang L et [7] used an SVM-enhanced Markov model to predict wind power generation. Literature [8] adopted a wind power prediction method based on random forest. New energy power generation prediction can be divided into physical modeling methods, statistical analysis methods, and deep learning methods according to different historical data sources.

Physical modeling methods model new energy and power electronic converters and use numerical weather forecasts as independent variables to achieve prediction [9]. Physical methods mainly rely on numerical weather forecasts for prediction. The disadvantage is that the accuracy is insufficient, and the modeling process is complex and requires high computational power. In statistical learning models, by combining factors such as light intensity, wind speed, and historical power generation, better prediction results can be achieved than other methods, but the modeling requirements are relatively high. With the help of big data and artificial intelligence technology, learning methods effectively improve the accuracy and effectiveness of solar and wind power prediction, providing support for the power grid's safe, stable, and economical operation. Machine learning, as the core technology of artificial intelligence, shows great potential in energy prediction. By learning and modeling historical data, machine learning algorithms can predict key indicators such as energy consumption and power generation, providing decision-making support for optimized energy system management. Among them, unique deep learning models, such as the Long-Short-Term Memory network (LSTM), can effectively capture the long-term dependence and non-linear patterns of energy data, laying a foundation for improving prediction accuracy. Its excellent performance benefits from its unique design and structure, enabling it to process complex sequence data and capture long-term dependencies. The long-short-term memory network, abbreviated as LSTM (Long Short-Term Memory) neural network, is a unique Recurrent Neural Network (RNN) type. It can solve the problems of gradient explosion and gradient disappearance that occur when traditional RNNs process long-sequence data.

At the same time, the random forest algorithm is a powerful machine learning tool. The random forest algorithm can make accurate predictions in many highly accurate scenarios [10]. It generates accurate and stable prediction results of multiple decision trees, reducing the risk of overfitting. The random forest is composed of multiple decision trees and is more robust to outliers and noise in the data. The random forest can evaluate the important features that impact the prediction results the most. The random forest algorithm can also be easily parallelized, improving calculation efficiency. However, the random forest algorithm also has its own defects. The computational complexity is relatively high. Since multiple decision trees need to be trained, the computational complexity of the random forest is usually higher than that of a single decision tree. When the data set is very large, or the number of decision trees is very large, the training time may be long and require a large amount of memory space. If the memory is insufficient, it may lead to performance degradation or inability to run. In some specific types of problems, such as certain time-series problems or problems that require capturing long-term dependencies, the random forest may not be as effective as other algorithms [11-14], such as recurrent neural networks. This paper proposes combining the random forest and the Long-Short-Term Memory network (LSTM) to predict new energy power generation, which can form a hybrid model. This model combines the advantages of the two algorithms and has some unique advantages: it has a strong prediction ability. The random forest is known for its high accuracy and stability, while the LSTM is good at processing time-series data and capturing long-term dependencies. Combining the two can enhance the processing ability of time-series data while maintaining the prediction accuracy of the random forest to provide more accurate and comprehensive predictions, making the model more flexible and effective in processing complex data sets, improving the generalization ability of the model to a certain extent, reducing the risk of overfitting, and enabling long-term prediction and improving prediction accuracy [15,16].

1.3. Research Ideas

This study collected and analyzed the historical data of wind power generation and solar power generation from 2021 to 2023 in two regional power plant clusters and used the random forest and LSTM algorithms to predict solar power generation and wind power generation, respectively. The specific research ideas are as follows:

(1) Since the data may be abnormal, the data is pre-processed first.

(2) The optimal model is selected through parameter tuning of the random forest and LSTM algorithms for model training to predict solar power generation and wind power generation.

(3) The best accuracy is analyzed through error calculation, and appropriate parameters are used to establish a suitable prediction model.
(4) The predicted values and real-value data are visualized to observe the results.

2. Theoretical Basis of the Model

2.1. Random Forest Algorithm

As a machine-learning strategy, the random forest belongs to the category of ensemble learning. Its core lies in gathering the power of multiple models to improve prediction performance. In machine-learning tasks, the goals can be roughly divided into two categories: classification and regression. The uniqueness of the random forest lies in its flexibility in dealing with these two types of tasks. The operation mechanism of this algorithm is based on a collection of decision trees. Each decision tree is independently trained and adopts random subsets of data samples and features during the generation process to enhance the diversity and generalization ability of the model. During prediction, the outputs of all decision trees are aggregated. Usually, majority voting (for classification) or mean calculation (for regression) is used to obtain the final prediction conclusion. This collective decision-making method improves the accuracy and stability of the overall prediction.

To understand the RF algorithm, it is necessary first to understand the meaning of the decision tree. The decision tree divides the data set into smaller subsets around the input variables. The division is performed to give the formed subsets a more minor variance in the result values. Each division can be regarded as a branch of the tree, and each data subset can be regarded as a leaf. The data is gradually divided until certain final conditions are met. The termination conditions can be that the maximum number of divisions has been performed or the standard deviation of the subset has dropped below a cut-off value. The average value of the results in the terminal leaves is the predicted value of the set of input variables.

For example, a decision tree is used to predict the power generation as a function of two key variables: wind speed and temperature. It is created by tree splitting with branches of wind speed < 10 m/s and wind speed ≫ 10 m/s. These branches are further divided for temperature < 350K and temperature ≫ 350K. Then, to predict the power generation under certain conditions, such as wind speed = 4.5 and temperature = 300K, the average value of the power generation of the leaf wind speed < 7 → temperature < 350K is the predicted value of this decision tree. RF is composed of a large number of decision trees. The decision trees are formed based on random subsets of the training data through replacement and using random subsets of features. The RF algorithm reports the weighted average of the predictions of all decision trees. Generally, using the results of many machine-learning models for the final prediction is called ensemble learning, and it has been proven to improve prediction performance significantly. Therefore, RF is an ensemble learning algorithm based on numerous decision trees. Thus, in the above example, power generation is the output variable, and wind speed and temperature are the input variables.

2.2. LSTM

A special type of cyclic neural network, long-term and short-term memory network (Long Short Term Memory networks), is called “LSTM” for short, which can explain the long-term dependence of variables. Hochreiter and Schmidhuber introduced them, and they have been improved and popularized for predicting feasibility.

LSTM belongs to a type of Recurrent Neural Networks (RNNs). Compared with traditional RNNs, LSTM has stronger memory ability and the ability to model long-term dependencies. LSTM solves the problems of gradient vanishing and gradient explosion in RNNs by introducing a mechanism called “gates”. These gates include the input gate, forget gate, and output gate, which control the flow of information by determining whether the information is to be remembered, forgotten, or output.

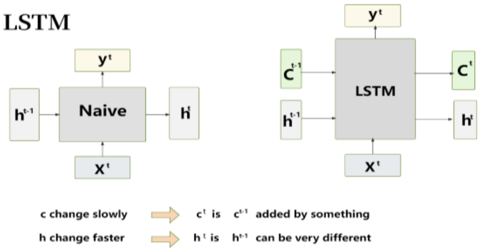


Fig. 1 Structure diagram of LSTM (right) and RNN (left)

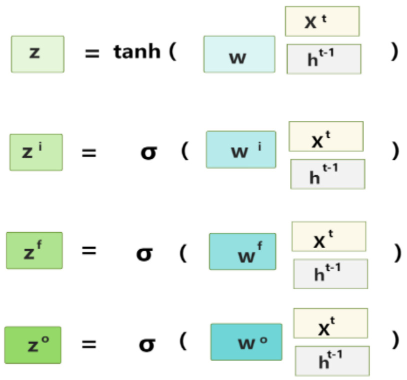
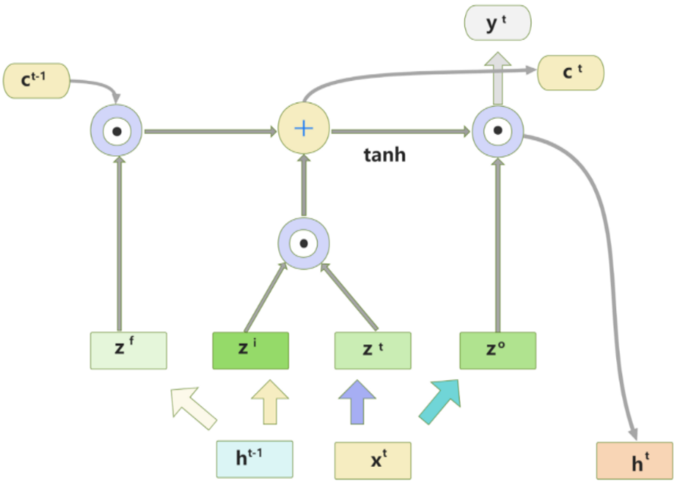


Fig. 2 State diagram of the internal structure of LSTM



$$c^t = z^f \odot c^{t-1} + z^i \odot z$$

$$h^t = z^o \odot tanh(c^t)$$

$$y^t = \sigma(W'h^t)$$

Fig. 3 LSTM operation flow chart

Analyzing the internal structure of LSTM as shown in Figure 1, input it into the LSTM formula and train it with the passage from the previous state to obtain four values. Among them, z^f, z^i, z^o it is multiplied by the splicing vector by the weight matrix and then converted into a value between 0 and 1 by a sigmoid activation function, which is used as a gated state. Itzconverts the result to a value between -1 and 1 through a tanh activation function (tanh is used here because it is used as input data, not as a gated signal).

Figure 3 is Hadamard Product, $\odot$which represents the multiplication of the corresponding elements in the operation matrix, so the two multiplication matrices are required to be of the same type. $\oplus$ is represented by matrix addition.

3. Data processing and prediction

3.1. Forecast of Random Forest Solar and Wind Power

First, the solar and wind data are preprocessed, a random forest model is constructed to process the data set to get the predicted power, the parameter network is set, the cross-verification and parameter selection is carried out, and the prediction is carried out through the training model. Finally, the results are evaluated by error calculation, and the real and predicted values graphs are drawn.

3.1.1. Data Preprocessing

The general steps of data preprocessing are cleaning and dealing with data problems such as missing values, outliers and repeated values. Cleaning is carried out by filling in missing values, deleting outliers and repeating values. Feature selection: select features with high correlation with target variables. Choices can be made based on statistical methods such as correlation coefficients or the feature importance of the model. The common coding methods are One-Hot Encoding, Label Encoding and so on. Feature scaling: continuous features are scaled to the same scale. Common scaling methods, such as Standardization and Normalization, are used to eliminate dimensional effects. Data split: divide the dataset into training sets and test sets. The training set is used to build the model, and the test set is used to evaluate the model's performance.

3.1.2. Random Forest Parameters

The random forest library provides two core classes: Random Forest Classifier for classification tasks and Random Forest Regressor for regression tasks. Similarly, its variant Extra Trees also provides Extra Trees Classifier and Extra Trees Regressor. The core idea of stochastic forest regression is to build a more powerful model by combining multiple independently trained decision trees. These decision trees will consider different data subsets and feature subsets in the training process to reduce the model variance and improve the generalization ability. In terms of parameter adjustment, although random forests usually do not need a large number of parameter adjustments, the number of decision trees controlled by n_estimators parameters is a key parameter. If the n_estimator setting is too small, it may cause the model to underfit, while setting too large may not significantly improve performance. By default, the parameter value is 100. Another important parameter is the max_depth of the decision tree, which limits the maximum depth of the tree. In some cases, if the data sample size or the feature is large, it is beneficial to limit the depth of the tree, while when the sample size or feature is small, it may not need to be limited. It is worth noting that random forest is a decision tree model based on a bagging framework. Therefore, the parameter tuning involves two levels: one is the parameter tuning of the RF framework, such as n_estimators, and the other is the parameter tuning of the RF internal decision tree, such as max_depth. Understanding these parameters is the premise of effective model tuning.

3.1.3. Generation Power Prediction

In this paper, we introduce the prediction method of wind power and solar power rates based on the RF random forest algorithm. After determining the influence factors of solar energy and wind power, the pre-prediction mathematical model of solar energy and wind power is established.

A photovoltaic power prediction system constitutes a comprehensive solution that closely integrates the three cores of data monitoring, power prediction models, and advanced software platforms. Among them, the data monitoring link, as the cornerstone of prediction, not only closely tracks and records the real-time changes of meteorological variables but also continuously monitors the operation status of photovoltaic power stations to ensure the comprehensiveness and accuracy of information. The system is equipped with optical power prediction technology, which is good at carrying out short-term and ultra-short-term prediction tasks and accurately fits the diversity of photovoltaic enterprises' forecasting needs on different time scales. This ability is very important for lighting conditions and helps enterprises to dispatch resources and optimize production capacity efficiently. The supporting software platform transforms complex data into intuitive graphics and reports, provides management and decision-making methods, and further excavates data value through in-depth data analysis. The platform strictly follows the standards and specifications of power grid operation, and its design essence lies in high efficiency, accuracy and intelligence, which lays a foundation for daily power generation planning and operation and maintenance strategy and ensures the stability and efficiency of the power system.

3.1.4. Model Evaluation

The loss function is obtained by the Mean Square Error (MSE) of the fitting degree between the predicted value of the evaluation model and the initial data set and the fitting degree of error analysis. The smaller the loss function value is, the better the model fitting is and the more accurate the prediction is. The average absolute error (MAE) index evaluates the prediction results, and the model accuracy is evaluated by Root Mean Square Error (RMSE) and correlation coefficient R^2. The calculation formula is as follows:

$$MAE = \frac{1}{n_s}|\hat{y}_i - y_i| \tag{1}$$

$$MSE = \frac{1}{n_s}\sum_{i=1}^{n_s}(\hat{y}_i - y_i)^2 \tag{2}$$

$$RMSE = \left(\frac{1}{n_s}\sum_{i=1}^{n_s}(\hat{y}_i - y_i)^2\right)^{\frac{1}{2}} \tag{3}$$

Where n_s is the number of amples, $\hat{y}_i$ is the predicted value of generating power, and y_i is the actual value of generating power.

3.2. Prediction of Solar and wind Power Generation in the LSTM Network

First, we preprocess the data set, create features and target values, build an LSTM model, optimize the parameters to find the best hyperparameters, and finally compare the predicted and the real values.

3.2.1. Data Preprocessing

LSTM is a special recurrent neural network. When designing neural network prediction models, especially in processing dynamic data such as solar and wind power, the working range of non-linear activation functions must be carefully considered to prevent neurons from experiencing saturation. To this end, the normalization of the data is a necessary step, and the process ensures that all data points fall into a uniform range from 0 to 1 through scaling, which is essential to maintain the optimal efficiency of the activation function. It is worth noting that if the initial value of the data point exceeds the boundary of -1 to 1, it should be identified and eliminated in the preprocessing stage to ensure the effective implementation of normalization. In addition to normalization, data preprocessing includes reasonable classification of input data. For the case of wind power prediction, it is a standard operation to divide the data set into two independent parts: the training set and the test set, which not only helps the model to learn historical patterns but also evaluates the generalization ability of the model by retaining unknown data. Then, the model training starts, during which the architecture design and parameter configuration of the neural network are constantly fine-tuned, and optimization algorithms such as gradient descent are used to reduce the prediction error gradually. LSTM network, with its unique advantage in capturing long-term dependence on time series data, has become an ideal choice for this kind of prediction task. With the deepening of training, the LSTM network gradually refines its internal mechanism to maximize the consistency between the predicted output and the actual observations. Finally, the reserved test set is used to verify the trained LSTM model to test the accuracy and reliability of the model strictly to predict the future wind power rate. This series of fine operation flows combines the rigour of data preprocessing and the flexibility of model training to build a strong and accurate prediction system, which provides solid support for the efficiency analysis and strategic planning of solar and wind power generation.

3.2.2. Determine the Parameters of the LSTM Neural Network

Parameters are the most important learning goal of our training neural networks. The purpose of our training is to find a set of good model parameters to predict the results of the position. These parameters are generated automatically during model training. We adjust and combine the superparameters n_estimators and max_depth through the “grid method” and then find the best superparameters by the score of each parameter combination.

3.2.3. Generation Power Prediction

Photovoltaic power prediction system is a technology based on data analysis and model prediction. Through the comprehensive analysis of historical data, weather forecasts and power station parameters, the photovoltaic power generation power of photovoltaic power stations in the future can be predicted. It is like the “intelligent brain” of a power station, which can predict the generating capacity of the power station in advance. The optimized operation of the station provides an important basis.

3.2.4. Model Evaluation

The loss function is obtained by evaluating the fitting degree between the predicted value of the model and the initial data set and the MSE of the fitting degree of the error analysis. The smaller the value of the loss function, the better the model fitting and the more accurate the prediction. The calculation formula is as follows:

$$MSE = \frac{1}{n_s}\sum_{i=1}^{n_s}(\hat{y}_i - y_i)^2 \tag{4}$$

4. Case Analysis

The data set studied in this paper comes from the special competition of “Smart Energy Competition”. The training data in the data set are weather history data, weather forecast data, and power generation time series data for one field group. The time span of the sample is two years, and the time resolution is fifteen minutes, including solar power generation and wind power generation. The input variables are shown in Table 1.

Table 1. Design of input parameters of regional field group power generation

Prediction of solar power generation		Prediction of wind power generation	
Field name	Field Type	Field name	Field Type
Time stamp	string	Time stamp	string
Historical photovoltaic power generation	float64	Historical wind power	float64
Observation and irradiation of historical field station	float64	Observation of wind speed at historical stations	float64
Installed capacity of station	float64	Installed capacity of station	float64
Inclination angle	float64	Temperature	float64
Azimuth angle	float64	Air pressure	float64
Temperature	float64	Wind speed	float64
Air pressure	float64	Precipitation amount	float64
Precipitation amount	float64		
Total cloud cover	float64		
Total solar radiation	float64		
Wind speed	float64		

The provided dataset includes meteorological information from two meteorologists, and the selection of features will be carried out according to the importance of the features in the study to optimize the accuracy of the forecast further. Random forest has the advantages of robustness, high accuracy, and the ability to process and analyze high-dimensional and large-scale data, and it is not easy to over-fit, which makes it one of the most powerful algorithms in machine learning. When using the random forest algorithm, we adjust its parameters, including the maximum depth of the decision tree max_depth and the number of decision trees n_estimators. The range of values is shown in Table 2.

Table 2. Random forest parameters

Parmeters	*max_depth*	*N_estimators*
	None	10
	5	20
	10	30
	15	40
	20	50
		60
		70
		80
		90
		100
		150
		200

The parameters of LSTM are shown in Table 3. Batch_size indicates the number of samples included in each batch of training. When training the neural network, the data set is divided into several batches, and each batch has a certain number of samples. The gradient is calculated for each batch, and the model parameters are updated to improve the training speed and the stability of the model. Epochs represent the number of iterations of model training and the number of times the entire data set is entered into the model for training. In each epoch, the model carries forward propagation and backpropagation to the whole data set and updates the model parameters according to the gradient calculated by backpropagation. Optimizer is an optimization algorithm for updating model parameters. Common optimization algorithms include Stochastic Gradient Descent, Adam, Root Mean Square Propagation and so on. Choosing the appropriate optimization algorithm has an important impact on the convergence speed and performance of the model.

4.1. Forecast of Solar Power

4.1.1. Random Forest

Solar energy has attracted much attention as the representative of new and clean energy. Although solar power now accounts for only six percent of power generation, according to foreign forecasts, solar energy will account for fifty percent or more of China’s electricity generation in 2050.

The software used in the prediction process is Pycharm and is based on Python 3.10. Figure 4 shows the resulting diagram of parameter optimization of the random forest algorithm. As can be seen from the figure, the accuracy of the model prediction is high, above 90 per cent. The optimal parameter combination is max_depth = None, n_estimators = 150, and the accuracy is 96.72 percent. The test set is predicted using the optimal model, and the result is shown in Figure 4. As can be seen from Figure 5, the real and predicted values of solar power are concentrated in the middle part, indicating that this model can well predict the power generation power.

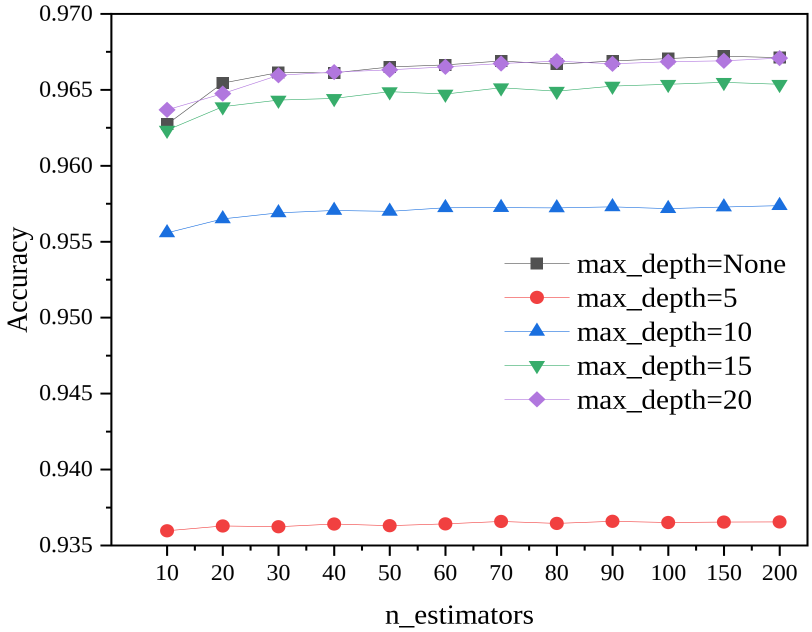


Fig. 4 Optimization of solar power parameters based on random forest.

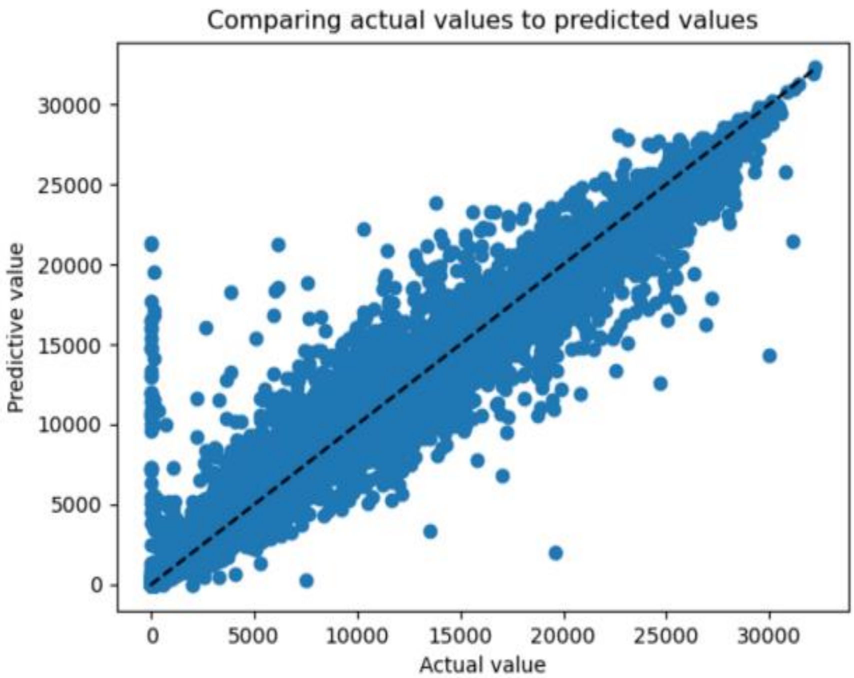


Fig. 5 Comparison between real value and predicted value of solar power generation based on random forest

4.1.2. LSTM

Table 3. LSTM Parameters

Parameters	*Batch_size*	*epochs*	*optimizer*
	16	8	adam
	20	10	Adadelta

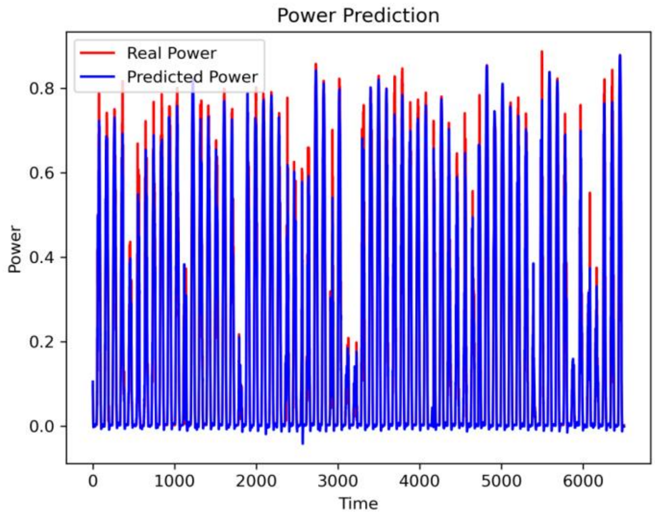


Fig. 6 Real and predicted solar power based on LSTM

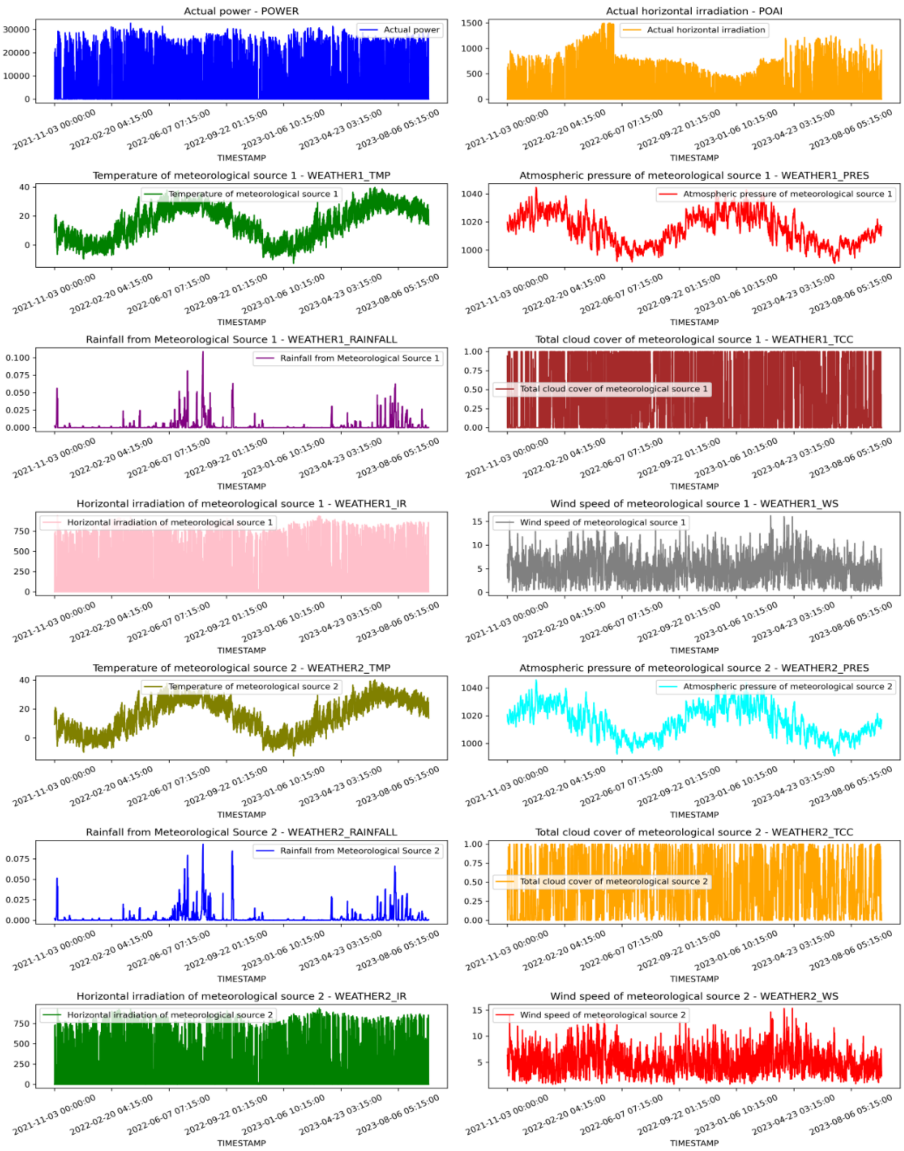


Fig. 7 Solar energy data visualization

LSTM has the advantages of dealing with long-term dependence, adapting to multiple sequence lengths, strong expression ability, end-to-end learning and flexible input and output, which makes it an effective tool for processing and predicting time series data. Next, you will use LSTM to predict the data.

In order to understand our data set, each feature is drawn. This shows that between 2021 and 2023, there are different patterns for each feature. It also shows the location of exceptions that will be resolved during the normalization process.

The number of data points of solar energy is 66913, and we divide the test set and training set according to 0.715, which is 71.5 percent. Five is the training set, and the rest is the test set. Track the data of the past 720 timestamps, which are used to pre-measure the power after 72 times tamps. Because each function has a different range, it is normalized to limit the eigenvalues within the range to train the neural network. This is done by subtracting the average and dividing by the standard deviation of each feature.

In addition, the heat map of the dataset is shown in Figure 8. We will further select the features according to the heat map because some parameters are redundant.

According to the set label, the training number starts from the 792nd observation value (720 + 72). The timeseries_dataset_from_array function is used to input equal intervals and time series parameters, such as sequence or window and so on, to generate batch sub-time series inputs and targets sampled from the main time series. The validation data set must not contain the last 792 rows because we do not have these records, so we must subtract 792 from the end of the data.

Validating the tag dataset must start at 792 after train_split, so we must add past and future to label_start. As a result, the input and output input shapes (256, 120, 7) and target shapes (256, 1) are obtained.

We will use callbacks to save checkpoints periodically, and callbacks that interrupt training when verification fails are no longer improved and save model_checkpoint.h5.

In addition, we define the visualize_loss(history, title) function to visualize the loss of training and testing, as shown in Figure 9. Train loss continues to decline, and test loss continues to decline, indicating that the network is learning. Train loss continues to decline, test loss tends to be constant, indicating that network overfitting train loss tends to be unchanged, test loss tends to remain unchanged, indicating that learning encounters bottlenecks, need to reduce learning Erate or batch size train loss tends to be constant, test loss continues to decline, indicating that one hundred percent of data sets have problems train loss continues to rise, test loss continues to rise, eventually become NaN, may be due to improper network structure design and improper setting of training hyperparameters Caused by a problem such as a program bug. It can be seen from Figure 9 that the whole model is running normally.

Finally, five groups of values are predicted and verified using the trained model, and the results are shown in Figure 10. The changing characteristics, true values and predicted values of historical data can be seen in Figure 10. The real value and the predicted value are not exactly the same, indicating that the prediction model can not accurately predict some specific data values. Further tweaking and optimization are needed.

Therefore, the model is further improved. let batch_size= 256, epochs = 20, optimizer = adam. We will further adjust these parameters, as shown in Table 3. The grid search function is used to optimize the parameters, and the best super parameter is {batch_size': 16, 'epochs': 8, 'optimizer':'Adadelta'} Figure 6 is the solar power prediction result of the LSTM algorithm, and the model can achieve 97.37 percent prediction accuracy. The accuracy of LSTM's model is higher than that of the random forest model.

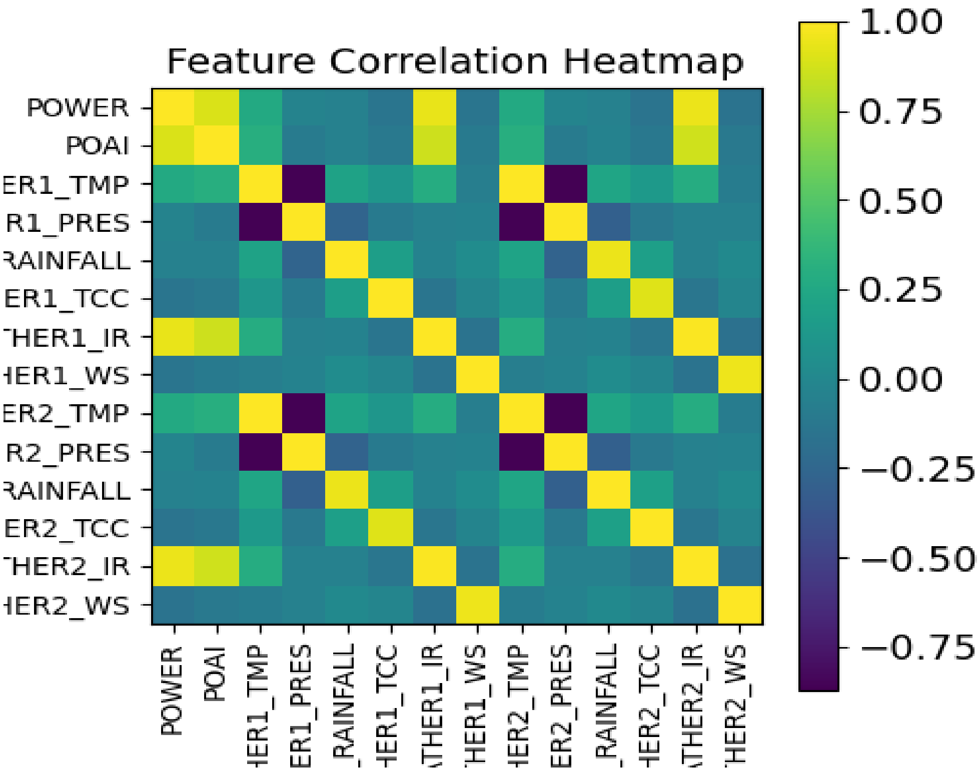


Fig. 8 Solar energy data heat map

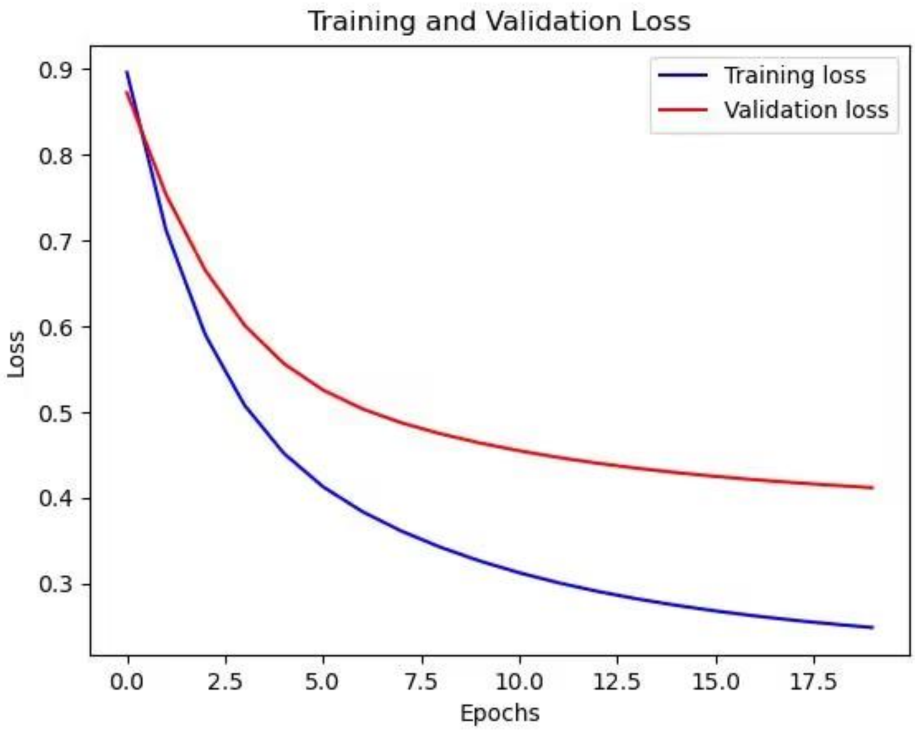


Fig. 9 Visualization of power function loss of solar power generation based on LSTM

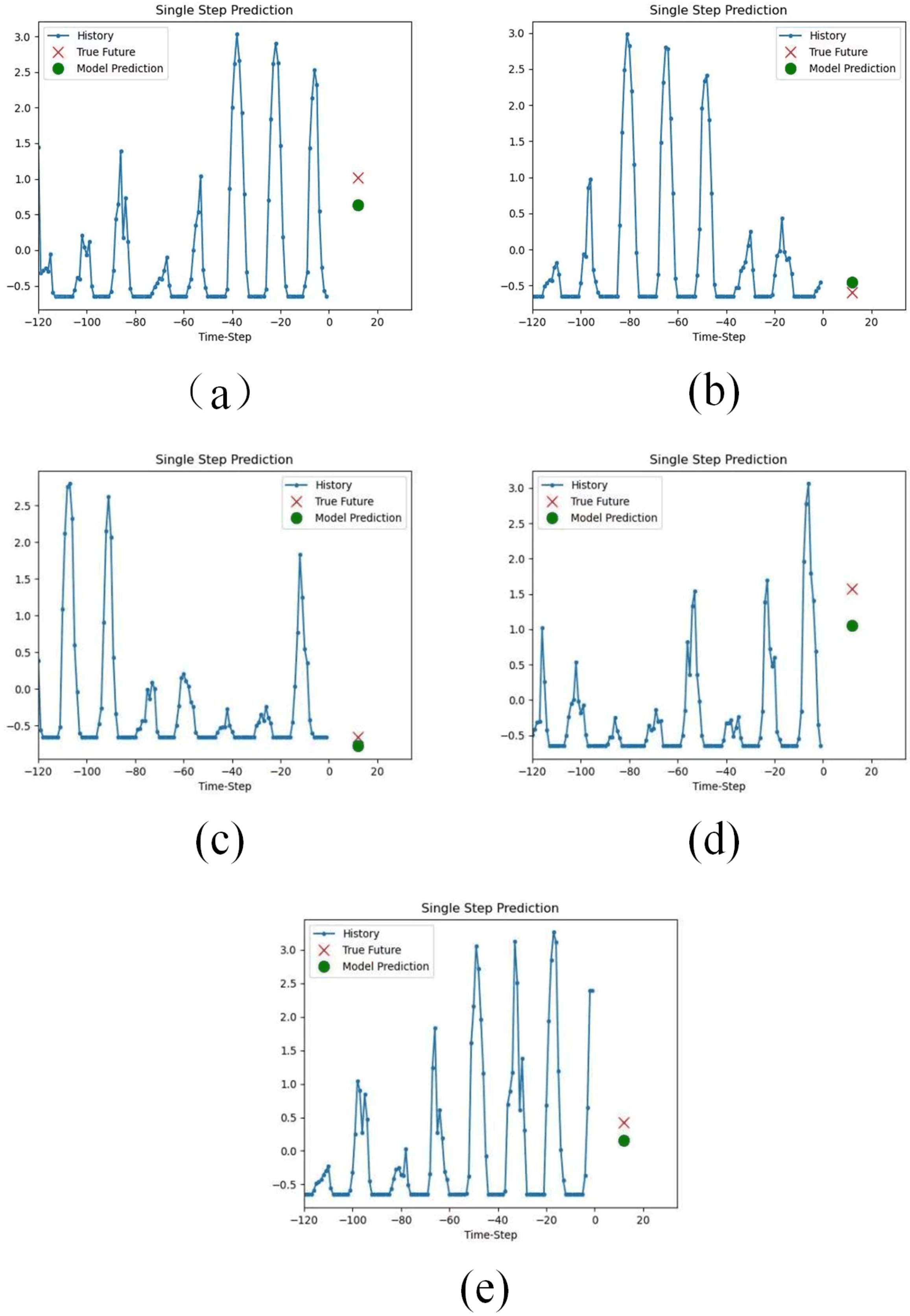


Fig. 10 Single-step Prediction of Solar Power based on LSTM

4.2. Forecast of Wind Power

4.2.1. Random Forest

China's wind power market is expanding after nearly a decade of development. However, the volatility and uncertainty of wind energy as a natural resource bring great challenges to wind power generation. In the following, we use the optimal model to predict the test set based on the parameter optimization of the random forest algorithm.

Figure 11 draws the result diagram of parameter optimization based on the parameter combination of n_estimators and max_depth. The results shown in the road show that the model's prediction accuracy is relatively high and stable at more than 90 percent. The optimal parameter combination is {max_depth=None, n_estimators 200cm}, and the accuracy is 97.67 percent. The result of training the test set using the model of the optimal parameter is shown in Figure 12. It can be seen from Figure 12 that the real and predicted values of solar power are concentrated in the middle, indicating that this model achieves an excellent effect in predicting power generation.

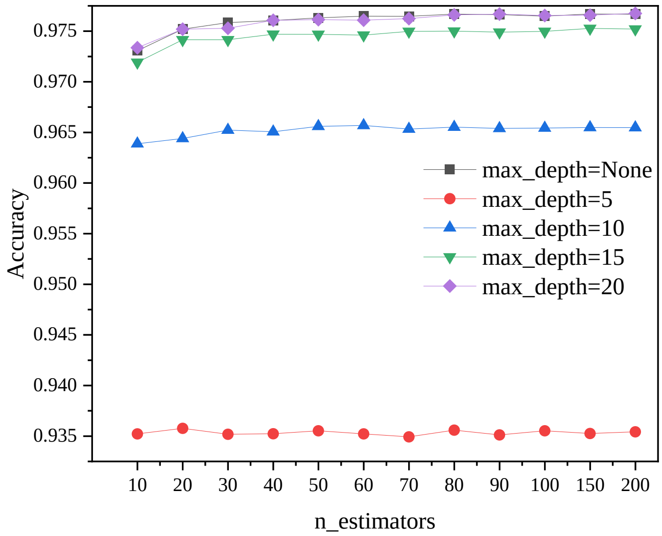


Fig. 11 Optimization of Wind Power parameters based on Random Forest

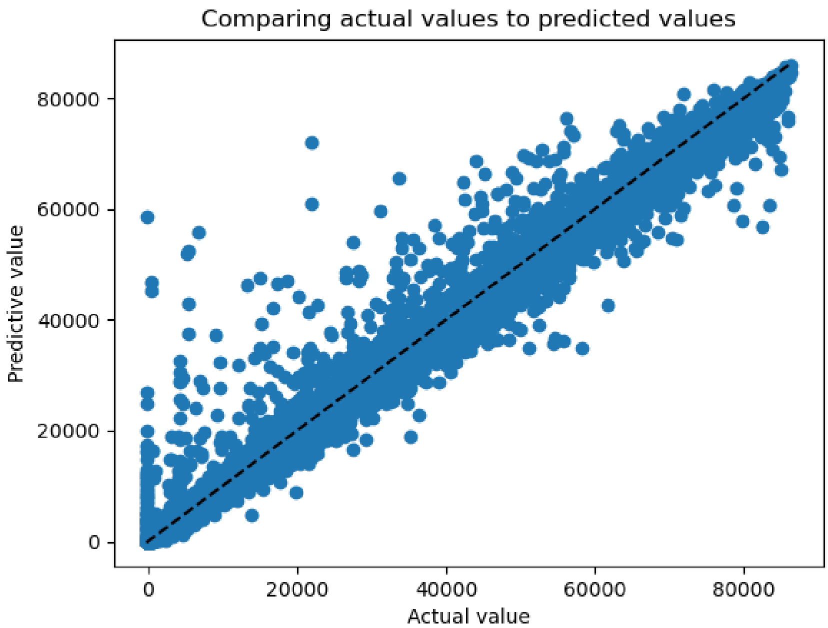


Fig. 12 Comparison between real value and predicted value of wind power based on random forest

4.2.1. LSTM

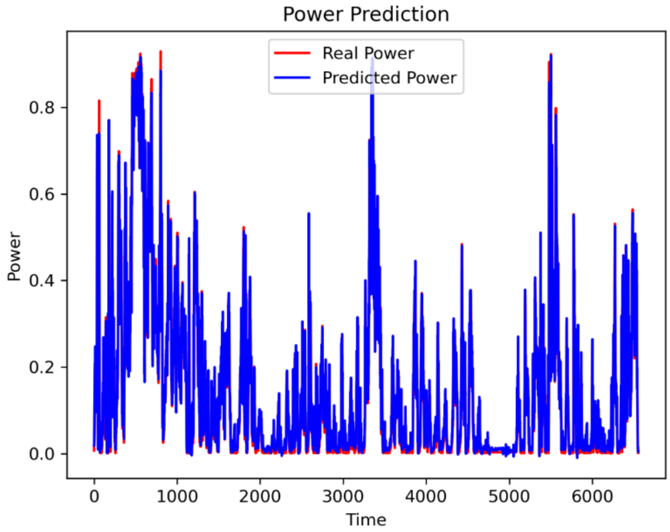


Fig. 13 Real and predicted solar power based on LSTM

As the stochastic forest model has some limitations in dealing with the long-term dependence of time series data, we also introduce the LSTM model, and we will use LSTM to predict the power generation power of the wind data.

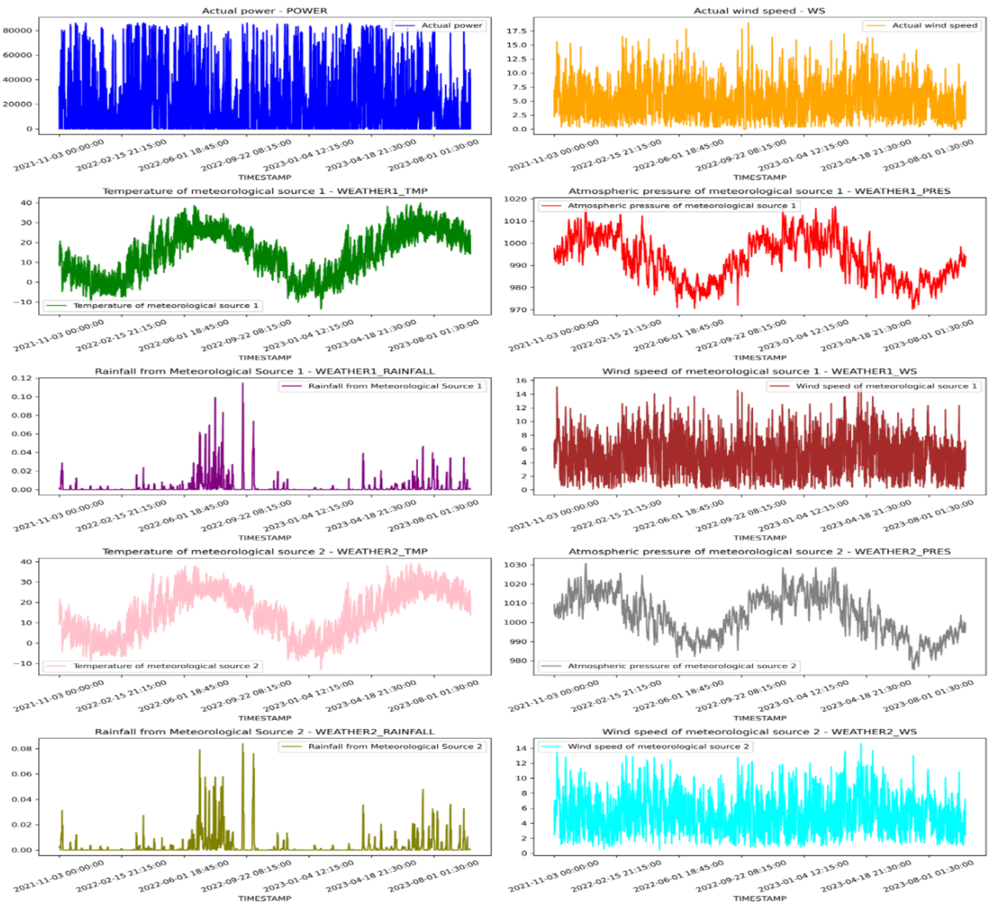


Fig. 14 Visualization of wind data

First, we process the dataset's data and facilitate visualization, as shown in Figure 14 below. Figure 14 shows how each data feature changes over time so that we can better understand the characteristics of the data. At the same time, according to the heat map shown in Figure 15, we further select the more important features as input. The selected features include 'POWER', 'WS', 'WEATHER1_PRES', 'WEATHER1_RAINFALL', 'WEATHER1_WS', 'WEATHER2_PRES', 'WEATHER2_RAINFALL', 'WEATHER2_WS'.

The loss of wind power function can be visualized as shown in Figure 16, which shows that the current model is in the learning process and has no problems, such as fitting. Similarly, we make a single-step prediction of wind power, and the prediction results are shown in Figure 17 and the effect needs to be further optimized.

Grid search is used to optimize the parameters, and the parameter setting is consistent with the solar power prediction. The best superparameter of the model is {'batch_size': 16,' epochs': 10, 'optimizer':' Adadelta'}, and the prediction accuracy is 97.65 percent. As shown in Figure 13, the predicted value is close to the actual value. The prediction accuracy is close to that of the stochastic forest model.

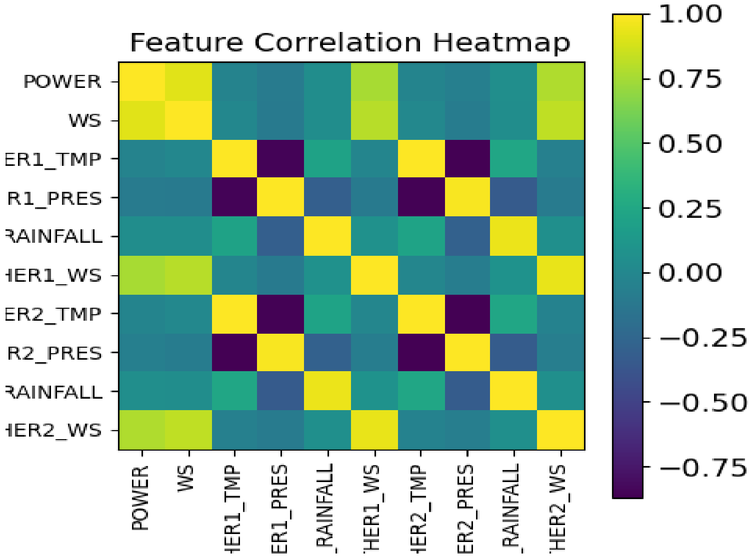


Fig. 15 Wind data heat map

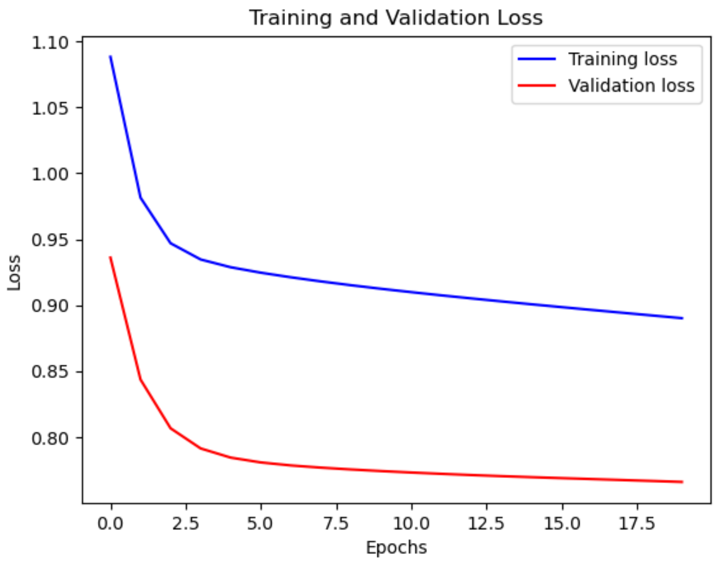


Fig. 16 Visualization of Wind Power function loss based on LSTM

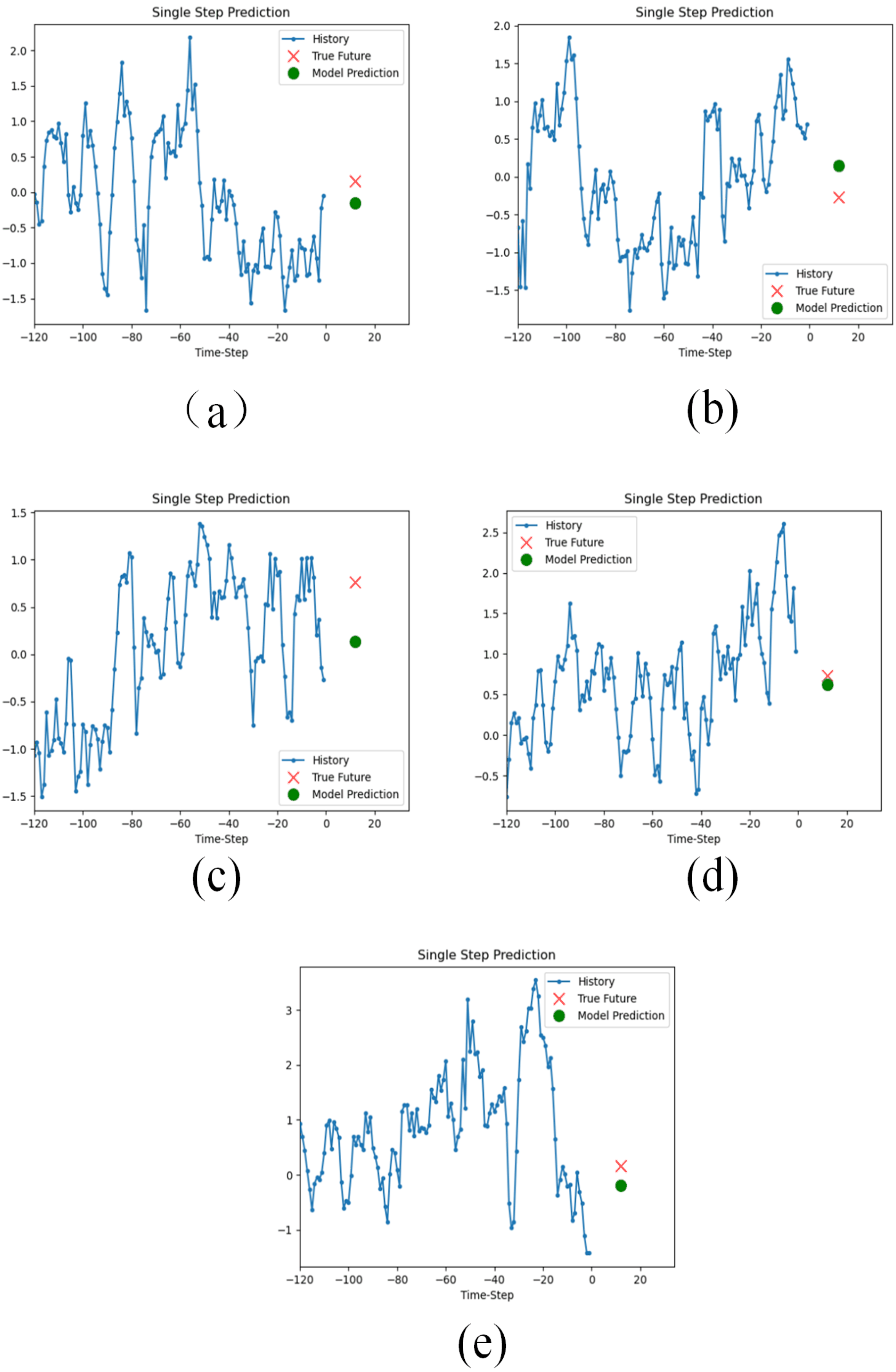


Fig. 17 Single-step Forecast of Wind Power based on LSTM

5. Main Conclusions and Deficiencies

5.1. Conclusion

The purpose of this study is to predict the power of solar and wind power and to compare the prediction results of different machine learning models. The results show that both the random forest model and the LSTM model excel in predicting power generation, but the LSTM model performs better.

In the process of solar power prediction, through parameter optimization, the optimal parameters of the random forest model are combined as '{max_depth=None, n_estimators=150}', and the prediction accuracy reaches 96.72 percent. However, because the stochastic forest model has some limitations in dealing with the long-term dependence of time series data, we further introduce the LSTM model.

LSTM model can capture the long-term dependence of data in time series, is suitable for various series lengths, and has a strong ability to express. After parameter optimization, the best parameter combination of the LSTM model is {'batch_size': 16,' epochs': 8, 'optimizer':' Adadelta'}, and the prediction accuracy is improved to 97.37 percent, which is significantly better than that of random forest model.

The introduction and optimization of the LSTM model significantly improve the prediction accuracy of solar power generation. Although the prediction on some specific data values still needs to be further optimized, it shows high prediction accuracy and overall stability. Using the same method to predict the power of wind power has also achieved good results. Considering this comprehensively, the advantage of the LSTM model in time series data prediction is fully reflected, which provides an effective method for predicting the power generation rate of new energy.

To sum up, this study verifies the superior performance of the LSTM model in solar and wind power prediction. It provides a valuable reference for future applications in the field of new energy.

5.2. The Deficiency of the Model

In the process of parameter tuning, the parameter settings of random forest and LSTM, as well as their interaction and influence, should be considered simultaneously. This may make the parameter tuning process more difficult and require more time and experience. If the wind and solar environment continue to change, the tracking and prediction ability of the deep learning algorithm is insufficient. Future research will combine a variety of optimization algorithms to improve the prediction performance of deep learning and random forest algorithms.

References

[1] Yong Wang, Rui Yang, and Lang Sun, "A Novel Structure Adaptive Discrete Grey Bernoulli Model and Its Application in Renewable Energy Power Generation Prediction," *Expert Systems with Applications*, vol. 255, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[2] Quan Lu, Wenxuan Huang, and Linfei Yin, "Decomposition Prediction Fractional-Order Active Disturbance Rejection Control Deep Q Network for Generation Control of Integrated Energy Systems," *Applied Energy*, vol. 377, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[3] Yue Yu et al., "Dynamic Time-Delay Discrete Grey Model Based on GOWA Operator for Renewable Energy Generation Cost Prediction," *Renewable Energy*, vol. 242, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[4] Seyyed Reza Ghaffari-Razin, Amir Reza Moradi, and Navid Hooshangi, "Modeling and Forecasting of Ionosphere TEC Using Least Squares SVM in Central Europe," *Advances in Space Research*, vol. 70, no. 7, pp. 2035-2046, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[5] Ruddy Blonbou, "Very Short-Term Wind Power Forecasting with Neural Networks and Adaptive Bayesian Learning," *Renewable Energy*, vol. 36, no. 3, pp. 1118-1124, 2011. [CrossRef] [Google Scholar] [Publisher Link]

[6] Zhe Song, Yu Jiang, and Zijun Zhang, "Short-Term Wind Speed Forecasting with Markov-Switching Model," *Applied Energy*, vol. 130, pp. 103-112, 2014. [CrossRef] [Google Scholar] [Publisher Link]

[7] Lei Yang et al., "Suport-Vector-Machine-Enhanced Markov Model for Short-Term Wind Power Forecast," *IEEE Transactions on Sustainable Energy*, vol. 6, no. 3, pp. 791-799, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[8] Zehong Zhou, Xiaohui Li, and Huaren Wu, "Wind Power Prediction Based on Random Forests," *Proceeding of 2016 4th International Conference on Electrical & Electronics Engineering and Computer Science*, pp. 352-356, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[9] Vladimir Simankov et al., "A Solar and Wind Energy Evaluation Methodology Using Artificial Intelligence Technologies," *Energies*, vol. 17, no. 2, pp. 1-23, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[10] Yuzhuo Hu et al., "Polynomial Surface-Fitting Evaluation of New Energy Maximum Power Generation Capacity Based on Random Forest Association Analysis and Support Vector Regression," *Frontiers in Energy Research*, vol. 11, pp. 1-13, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[11] Chen Wang et al., "A New Transfer Evolutionary Multi-Task Optimization Algorithm for Bi-Level Optimal Configuration of Distributed Generations and Energy Storage Systems Considering Uncertainties," *Journal of Energy Storage*, vol. 113, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[12] Tirunagaru V. Sarathkumar et al., "Forecasting of Virtual Power Plant Generating and Energy Arbitrage Economics in the Electricity Market Using Machine Learning Approach," *Scientific Reports*, vol. 15, pp. 1-13, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[13] Ha Eun Lee et al., "Comparative Life Cycle Assessment of Carbon-Free Ammonia as Fuel for Power Generation Based on the Perspective of Supply Chains," *Energy*, vol. 312, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[14] Alain K. Chaaban, and Najd Alfadl, "A Comparative Study of Machine Learning Approaches for an Accurate Predictive Modeling of Solar Energy Generation," *Energy Report*, vol. 12, pp. 1293-1302, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[15] Zhijian Qu et al., "A D-Stacking Dual-Fusion, Spatio-Temporal Graph Deep Neural Network Based on a Multi-Integrated Overlay for Short-Term Wind-Farm Cluster Power Multi-Step Prediction," *Energy*, vol. 281, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[16] Huajian Yang, Chuqiang Chen, and Xinhua Xue, "Coupled Convolutional Neural Network with Long Short-Term Memory Network for Predicting Lake Water Temperature," *Journal of Hydrology*, vol. 655, 2025. [CrossRef] [Google Scholar] [Publisher Link]