INTRODUCTION

Coal has been the main energy source of China, accounting for nearly 60% of its total energy consumption (see Figure 1). However, coal consumption is also the main cause of greenhouse gas emissions and air pollution.1 According to data from the World Resources Institute (WRI, 2012), China has also been the largest emitter of global greenhouse gas emissions.2 For sustainable development, China has begun to transform its energy structure.3 As shown in Figure 1, the proportions of coal consumption are continuously decreasing and clean energy sources such as hydroelectric, natural gas, and other renewable energy are gradually increasing. However, the coal consumption...
did not decrease continuously; it demonstrated a decreasing trend in 2014, but historically it has steadily increased since 1965. Some researchers identified that Chinese coal consumption peaked in 2014, while coal consumption increased only slightly in 2018. This elusive situation was the basis of this study to predict the energy structure of China.

Previous studies on energy structure usually focused on energy efficiency, energy demand, and air pollution. These studies reported the relationship between energy structure and sulfides, and the impact of energy structures on energy efficiency. In addition, there are also studies that discuss the impact of a single type of energy source on the energy structure. Ji and Zhang reported that financial progress plays an important role in the development of renewable energy and effective policies are crucial for changing energy structures. In recent years, China has promulgated many policies to reduce the consumption of fossil fuels and to encourage the development of new energy. These policies have contributed to the construction of a new national energy demonstration zone in Ningxia, with the development of distributed photovoltaics installed in southern regions, increased numbers of wind turbines in coastal areas, and a new hydropower-based energy generation in the southwestern region.

China experienced rapid economic development that has reached a new normal. Therefore, historical data can provide sufficient support for predicting future energy structures. Many methods have been developed to predict the various end-use energy consumption of China. These methods can be divided into two categories: variable regression methods and time-series prediction.

In the variable regression method, variables such as gross domestic product (GDP) and carbon emissions are used to regress energy consumption, and then, the values of variables in the future are estimated based on the background information. Finally, these values can be input into a regression equation to predict the future energy consumption. Overall, the input of this type of method is the energy consumption-related variables and its output is energy consumption. Machine learning methods such as support vector machine (SVM) and artificial neural network (ANN) are the most commonly used traditional methods in this field because they can work with the nonlinear relationships among variables. Zeng et al. developed a novel method named “ADE-BPNN” to estimate energy consumption. This method is the fusion of the back-propagation neural network (BPNN) model and adaptive differential evolution (ADE) algorithm. Wang et al. proposed a system dynamics model to predict China’s coal consumption in different scenarios. Economic variables are the most commonly used variables to regress energy consumption because they are closely related to energy consumption. Wang et al. applied a self-adaptive multi-verse optimizer (AMVO) to optimize the parameters of SVM. They used the model to predict China’s primary energy consumption using several independent variables such as GDP per capita, population, urbanization rate, share of the industry in GDP, and share of coal in primary energy consumption. Solode reviewed the history, methods, and application of forecasting natural gas consumption. As he mentioned, many algorithms, such as the fuzzy genetic algorithm (FGA) and genetic algorithm (GA), have been applied in this field. Recently, fusion methods have been developed to overcome the drawbacks of each single method. Wang et al. fused the sparse AdaBoost, echo-state network (ESN), and fruit fly optimization algorithm (FOA) to predict industrial electricity consumption. In addition, ESN with differential evolution was also developed to predict electricity energy consumption and has been used widely in predicting energy consumption. Differential evolution was used to improve the precision of bagged-ESN in predicting energy consumption. In summary, there are two drawbacks of this methodology: (a) To predict energy structure for the next few years, knowledge of the related variables in the next few years is required. This leads to two types of errors, namely, the error of related variables and the error of model. These two types of errors will cause much larger faults than linear superposition. (b) The precision of the method would increase by increasing the number of related variables into the model. However, the more number of variables that are added into the model, the more variables that must be estimated for predictions. It is clearly a dilemma to resolve this issue.

For time-series prediction, researchers consider energy consumption as a time series and use historical energy
consumption data to forecast future energy consumption. Because energy consumption is affected by countless variables, this type of method considers that the historical data of energy consumption contain all the information affecting/effecting energy consumption. Therefore, both the input and the output of this type of method are energy consumption data. The gray model (GM) and autoregressive integrated moving average model (ARIMA) are the most commonly used traditional methods in this field. Yuan et al. predicted the energy consumption of China by GM(1,1) and ARIMA. In addition, they compared the results of these two methods and found that the residuals of the two models are the opposite of each other. As the complexity of the problem increases, researchers often improve these traditional methods to achieve higher precision. An optimized hybrid GM(1,1) model was developed by Xu et al. to improve prediction accuracy of electricity energy consumption for short term analysis. Wang et al. used single-linear, hybrid-linear, and nonlinear forecasting techniques to predict energy demand in China and India. In recent years, deep-learning methods have proven to be powerful tools for time-series prediction. LSTM is considered the most suitable deep-learning method to process time series. Solar power generation, nuclear consumption, and coal consumption were predicted by LSTM. Although LSTM has achieved high precision in energy prediction, there are two drawbacks that need to be solved: (a) It is difficult to explain the prediction mechanism because of the application of historical energy data to predict energy consumption in the future, and (b) it cannot make full use of all information, because only historical energy data are used.

To avoid the drawbacks of these two methodologies, CNN-LSTM was developed in this study to fuse the advantages of both methods. First, the GDP and proportions of coal, oil, natural gas, hydroelectric power, nuclear power, and other renewable energy sources in the previous five years to construct features for the sixth year were used. This feature was convoluted by CNN as pixels. Following convolution, LSTM was used to process the features as a time series. Finally, the proportion of each energy source in the next few years was predicted. The major contributions of this research are as follows.

- Time-series prediction and variable regression are fused to predict the future energy structure. To achieve this, CNN and LSTM are integrated to significantly improve the prediction accuracy.
- Experiments on data from 1965 to 2014 demonstrate that the accuracy of the study method is higher than that of the six previous methods.
- The energy structure of China from 2019 to 2030 has been predicted by CNN-LSTM, which could be used as a guide for policy-making.

2 METHODS

2.1 Work frame

The work frame of our method is as Figure 2.

As shown in Figure 2, the workflow of CNN-LSTM can be divided into 3 steps:
The first step is to construct features. For prediction of the energy structure in the \(i\)th year, GDP and proportion of coal, oil, natural gas, hydroelectric power, nuclear power, and other renewable energy in previous five years are needed. Therefore, the dimension of the feature for each year’s energy structure is 5 \(\times\) 7. The unit of GDP is different from the proportion of the energy source; therefore, the feature must be normalized. Then, the final feature of energy structure for each year can be obtained.

The second step is to encode the CNN layers. In this step, the feature of energy structure is considered as pixels. Therefore, CNN was used to extract useful information from the features; two convolution layers and two pooling layers were selected. The parameters of these four layers are discussed in Section C. After encoding, the dimension of final feature for each year’s energy structure is 1 \(\times\) 7.

The last step is to predict the energy structure in \(i\)th year by LSTM. The final feature was entered into the input gate of LSTM. Two LSTM layers were constructed; their parameters are shown in Section D. After adjusting for weights, the LSTM model was built to predict energy structures of the future.

Overall, the study features contain two types of information. The first is a time series and the other comprises energy structure-related variables. Subsequently, CNN fuses the two different types of information to output the final feature to LSTM. Finally, LSTM predicts energy structure using a time-series prediction method.

### 2.2 Construction of feature

In this paper, the previous 5-year data are used to predict the data of 6th year. Therefore, if we want to predict energy structure in \(i\)th year, the feature of energy structure in \(i\)th year should be GDP and proportion of coal, oil, natural gas, hydro-electric power, nuclear, and other renewable energy from \(i\)-5th year to \(i\)-1th year.

\[
\begin{bmatrix}
C_{i-5}, O_{i-5}, H_{i-5}, N_{i-5}, NG_{i-5}, R_{i-5}, GDP_{i-5} \\
\vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \\
C_{i-1}, O_{i-1}, H_{i-1}, N_{i-1}, NG_{i-1}, R_{i-1}, GDP_{i-1} \\
\end{bmatrix}
\Rightarrow
\begin{bmatrix}
C_i \\
O_i \\
H_i \\
N_i \\
NG_i \\
R_i \\
\end{bmatrix}
\]

(1)

As shown in formula (1), \(C\) denotes the proportion of Coal consumption. \(O\) represents the proportion of Oil consumption. \(H\) is the share of Hydro consumption. Nuclear energy consumption is denoted by \(N, NG, R,\) and \(GDP\) denotes Natural Gas, other Renewable energy, and GDP respectively.

Since \(C, O, H, N, NG,\) and \(R\) are all lower than 1, whereas \(GDP\) are much larger than these variables. Therefore, we need normalize these variables. Here, we used Z-Score normalization to preprocess the feature.

\[
x' = \frac{x - \mu}{\sigma}
\]

\(\mu\) is the mean of \(x\), \(\sigma\) is the standard deviation of \(x\). We should put each column of feature into the formula (2).

### 2.3 Encoding by CNN

CNN is good method for image processing; it is inspired by the human visual nervous system. CNN has two main characteristics: (a) it can effectively reduce large amounts of data into small amounts of data, and (b) it can effectively retain image features and conform to the principle of image processing. At present, CNNs have been widely used, such as face recognition, automatic driving, and bioinformatics.\(^9,24,25\)

CNN preserves neighborhood relations of the input and spatial locality in the corresponding latent higher-level feature representations. Even though it is common, fully connected deep architectures do not scale well to realistic-sized high-dimensional features in terms of computational complexity; CNNs do because the number of free parameters describing their shared weights does not depend on the input dimensionality.\(^26\) Considering the background, the dimension of feature for each year’s energy structure is 5 \(\times\) 7, so CNN can automatically extract effective information from this pixel-like feature.

A typical CNN contains three parts: a convolutional layer, a pooling layer, and a fully connected layer. The convolutional layer is responsible for extracting local features in an image; the pooling layer is used to significantly reduce the parameter magnitude (dimensionality reduction); and the fully connected layer is similar to the part of a traditional neural network and is used to output the desired result. In this study, the fully connected layer of CNN was not used because LSTM was used to output the prediction results.

The convolution layer can be discerned as using a filter (a convolution kernel) to filter small areas of an image, to obtain the feature values of such small areas. In specific applications, multiple convolution kernels are required. Each convolution kernel can be considered to represent an image mode. The pooling layer can reduce the data dimensions more effectively than the convolution layer. It can significantly reduce the amount of computation, and also effectively avoid overfitting.

As mentioned in Section A, two convolution layers and two pooling layers were used in this study. The parameters are set as Table 1.

Activation functions Tanh and ReLu are as following:

\[
\text{tanh}(x) = \frac{\sinh(x)}{\cosh(x)} = \frac{e^x - e^{-x}}{e^x + e^{-x}}
\]

(3)
2.4 | Predicting by LSTM

LSTM is a special recurrent neural network (RNN), which is used mainly to solve the problem of gradient disappearance and gradient explosion during long sequence training. LSTM is a time-dependent method and requires global processing. Because global processing requires that only a complete input can contain all the information, nonglobal inputs may lose information, which can lead to modeling failure. In addition, LSTM contains memory (hidden) cells that transfer the information of each step to adjust the next step dynamically. LSTM is a commonly used method in time-series prediction. Liu et al.27 fused wavelet packet decomposition, CNN, and LSTM to predict wind speed. Peng et al.28 used differential evolution to improve LSTM and predicted electricity costs.

To put it simply, LSTM can perform better in longer sequences than ordinary RNNs. The forget gate of LSTM makes LSTM better than traditional RNN. There are three gates of LSTM. They are input gate, forget gate, and output gate.

The input gate is as follows:

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i)$$  \hspace{1cm} (5)

where $x_t$ denotes input data; $h_{t-1}$ denotes the output of previous step; $b_i$ and $W_i$ denote the bias and weight of input gate, respectively; and $\sigma()$ is a sigmoid function.

Then, the forget gate ($f_t$) is used to process the data further:

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

where $W_f$ is the weight of the forget gate, $h_{t-1}$ denotes output of the previous step, and $b_f$ denotes the bias of the forget gate.

The next step is to update the previous state.

$$\overline{C_t} = \tanh(W_C[h_{t-1}, x_t] + b_C)$$  \hspace{1cm} (7)

where $\overline{C_t}$ denotes the state of the previous step, $W_C$ denotes its weight, and $b_C$ denotes its bias.

Then, the cell state of current moment can be obtained by the previous state. Because $i_t$, $f_t$, and $\overline{C_t}$ can be obtained by (5), (6), and (7), they can be inserted into (7) to obtain the cell state of the current moment.

$$C_t = f_t * C_{t-1} + i_t * \overline{C_t}$$  \hspace{1cm} (8)

Finally, the cell state is input into output gate:

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o)$$  \hspace{1cm} (9)

where $o_t$ denotes the output gate and $W_o$ and $b_o$ are its weight and bias, respectively.

$$h_t = o_t * \tanh(C_t)$$  \hspace{1cm} (10)

The last step is to generate the output for the next iteration. The parameters of LSTM are set as Table 2.

### TABLE 1 Parameters in CNN

| Layers                      | Details                                   |
|-----------------------------|-------------------------------------------|
| First convolution layer     | Kernel size: 2 x 2                        |
|                             | Activation function: Tanh                 |
|                             | Filters: 32                               |
| First pooling layer         | Kernel size: 2 x 1                        |
|                             | Activation function: ReLu                 |
| Drop out                    | Rate: 0.25                                |
| Second convolution layer    | Kernel size: 2 x 2                        |
|                             | Activation function: Tanh                 |
|                             | Filters: 64                               |
| Second pooling layer        | Kernel size: 2 x 1                        |
|                             | Activation function: ReLu                 |

### TABLE 2 Parameters in LSTM

| Parameters                   | Details                                   |
|------------------------------|-------------------------------------------|
| The number of layers         | 2                                         |
| First layer                  | LSTM layer                                |
|                             | Nodes: 64                                 |
|                             | Activation function: ReLu                 |
| Second layer                 | Dense layer                               |
|                             | Nodes: 6                                 |
|                             | Activation function: ReLu                 |
| Loss function                | Mean Squared Error Loss                   |
| Optimizer                    | Adaptive Moment Estimation (Adam)         |

The energy consumption data from 1965 to 2018 were downloaded from BP Statistical Review of World Energy.29
Various energy sources are included: coal, oil, natural gas, nuclear, hydroelectric, and other renewable energy. The GDP data of China were obtained from “China Statistical Yearbook”. The proportion of each energy source could be calculated by the ratio of the energy consumption to the total energy consumption.

### 3.2 Validity of CNN-LSTM

Because data from 1965 to 2018 were obtained, the data from 1965 to 2008 were used to train the model, and the data from 2009 to 2018 were used to test it. Six previous methods were selected for comparison with CNN-LSTM to show its effectiveness: CNN, LSTM, BPNN, SVM, kernel-based nonlinear multivariate gray model (KGM), and ARIMA.

Because the study method CNN-LSTM is the fusion of CNN and LSTM, it was compared with CNN and LSTM to test the improvement from this fusion. The study approach is a deep-learning method, while BPNN is the representative method of shallow neural networks. Zeng et al. showed its power in predicting energy consumption in 2017, and therefore, it is selected to show the difference between deep-learning methods and shallow neural networks. All the methods mentioned above are based on neural networks; however, SVM has principles different from that of neural networks and has continuously been a rival of various neural network methods in the past few years. The gray model is the most popular method for time-series prediction and has been improved by many researchers. The KGM was developed in 2018 and has been widely used in predicting energy consumption recently. For this study, KGM was used as the representative of the gray model methods to compare with. ARIMA is the most traditional method in time-series prediction, and therefore, it is selected to show the performance difference between the fusion method and the time-series prediction method.

The parameters of these methods were set as Table 3.

Firstly, Figure 3 shows the prediction results of coal consumption.

As shown in Figure 3, CNN-LSTM performed best among the seven methods. The results show that CNN-LSTM is better than using CNN and LSTM alone. In addition, the results are more stable than the other methods.

As the aim of this study is to predict energy structure, tests were conducted not only on the coal consumption performance, but also on all the energy sources. Therefore, the study predicted the proportion of the six energy sources from 2009 to 2018. The error is calculated as the MAE of these six energy sources.

As shown in Table 4, CNN-LSTM performed best in MAE and MAXE, whereas LSTM's MSE is minimal. Compared
with other methods, LSTM is the second best method to predict proportion of coal consumption.

Since our aim is to predict energy structure, we not only tested the performance on coal consumption, but also all the energy source. Therefore, we predicted the proportion of six energy sources from 2009 to 2018. The error is calculated as the MAE of these six energy sources.

The study predicted the proportions of six energy sources for each year. The MAE of six energy sources for each year are shown in Figure 4. Although CNN-LSTM is not the best method to predict energy structure in all years, it performed in a very stable manner; if the results of 2009-2018 are considered as a whole, it performed the best.

As shown in Table 5, CNN-LSTM performed best among these seven methods. The performances of CNN and LSTM were close. Overall, we have proven the effectiveness of CNN-LSTM on predicting energy structure by the data from 1965 to 2019. The next step is to use CNN-LSTM to predict the energy structure in next decade.

### Table 4: Error comparison of seven methods

| Method   | MSE  | MAE  | MAXE |
|----------|------|------|------|
| CNN-LSTM | 0.037| 0.033| 0.067|
| CNN      | 0.029| 0.055| 0.101|
| LSTM     | 0.028| 0.048| 0.089|
| SVM      | 0.054| 0.061| 0.124|
| BPNN     | 0.059| 0.076| 0.167|
| ARIMA    | 0.067| 0.084| 0.162|
| KGM      | 0.071| 0.084| 0.154|

### Table 5: Error comparison

| Method   | MSE  | MAE  | MAXE |
|----------|------|------|------|
| CNN-LSTM | 0.0176| 0.0464| 0.0664|
| CNN      | 0.0323| 0.0617| 0.1089|
| LSTM     | 0.0268| 0.0681| 0.1035|
| SVM      | 0.04  | 0.1060| 0.1634|
| BPNN     | 0.074 | 0.1039| 0.2488|
| KGM      | 0.0699| 0.1100| 0.2750|
| ARIMA    | 0.0729| 0.1337| 0.2433|

were used to predict the energy structure for 2019. Then, it was possible to obtain all the variables that were needed to predict the energy structure in 2020, except GDP. Therefore, to start the iteration, the GDP from 2019 to 2030 was estimated first.

As shown in Figure 5, China’s economy has changed from a rapid growth phase to a stable growth mode. This situation is called the “new normal”. Recently, the growth rate was approximately 10% each year. Therefore, the study predicted China GDP with a 10% growth rate in the next decade.

The study obtained all the variables needed to predict the energy structure and results are shown in Figure 6.

As seen in Figure 6, the proportion of coal consumption will continue to decrease and remain at approximately 47% after 2025. Demand for gasoline will continue to increase. In addition, there will be no major changes in hydroelectric and natural gas consumption. However, considering the 5% annual increase in China’s total energy consumption, coal consumption will continue to increase.

### 3.3 Energy structure in next decade

In this section, all the data from 1965 to 2018 were used to train the CNN-LSTM model, and data from 2014 to 2018 were used to predict the energy structure for 2019. Then, it was possible to obtain all the variables that were needed to predict the energy structure in 2020, except GDP. Therefore, to start the iteration, the GDP from 2019 to 2030 was estimated first.

As shown in Figure 5, China’s economy has changed from a rapid growth phase to a stable growth mode. This situation is called the “new normal”. Recently, the growth rate was approximately 10% each year. Therefore, the study predicted China GDP with a 10% growth rate in the next decade.

The study obtained all the variables needed to predict the energy structure and results are shown in Figure 6.

As seen in Figure 6, the proportion of coal consumption will continue to decrease and remain at approximately 47% after 2025. Demand for gasoline will continue to increase. In addition, there will be no major changes in hydroelectric and natural gas consumption. However, considering the 5% annual increase in China’s total energy consumption, coal consumption will continue to increase.

### 4 CONCLUSIONS

China has gradually replaced coal by developing clean energy alternatives. Although the proportion of coal consumption

![Figure 4: Error of energy structure](image-url)
declines every year, coal consumption showed an increasing trend in 2018. Forecasting China’s energy structure can show the energy demand and environmental impact in advance, which is conducive to the improvement of sustainable development policies.

A large number of forecasting methods have been developed to predict consumption of various energy sources. However, none of these methods focus on energy structure. In this study, the consumption of various energy sources was considered as a whole and used to predict the proportion allocated to each energy source. To avoid drawbacks of the time-series prediction and variable regression, these two methods were combined by using CNN and LSTM. CNN was used to extract features from energy structure-related variables and LSTM was used to process the features obtained from CNN to predict energy structure as a time-series prediction.

To verify the effectiveness of CNN-LSTM, it was compared with six previous methods: LSTM, CNN, BPNN, SVM, KGM, and ARIMA. Historical data were used to test the performance of these methods. The results showed that CNN-LSTM was the best among these methods. After demonstrating the superiority of CNN-LSTM, it was used to predict the energy structure of China for the next decade. The study assumed an annual growth rate of GDP at 10% and input all the variables into the model. Finally, it was found that the proportion of coal consumption would continue to decrease for the next few years and remain at 47% after 2025. Oil and nuclear energy would gradually increase, and other energy sources will not experience significant changes. However, coal consumption is expected to increase continuously in the next decade, as the annual growth rate of total energy consumption was approximately five percent.
Although the study method performed best among the six previous methods, CNN-LSTM still has limitations in terms of parameter selection and data dependence. Setting proper parameters for deep-learning methods is a core problem for researchers. Most scholars adjust parameters manually through experience and iterations. The authors anticipate that parameter optimization methods such as genetic algorithms and simulated annealing algorithms can be used to set parameters more scientifically. In addition, the study approach is a data-driven method, which results in accuracy depending on the size and quality of the data. However, some variables that cannot be quantified, such as policies, also have an impact on energy consumption. Therefore, predicting energy consumption by combining policy is the next step.

ACKNOWLEDGMENT
This work was supported by the China Scholarship Council (Grant Numbers: 201708140198), Program for the Philosophy and Social Sciences Research of Shanxi: Research on the Moderate Industrial Structure Based on Energy Consumption and Carbon Emission, Program for the Philosophy and Social Sciences Research of Higher Learning Institutions of Shanxi (Grant Number: 2017313), and the Shanxi Soft Science Research Foundation, China (Grant Numbers: 2018041057-2).

ORCID
Yan Li https://orcid.org/0000-0003-4064-4532

REFERENCES
1. Qi Y, Stern N, Wu T, Lu J, Green F. China’s post-coal growth. Nat Geosci. 2016;9:564.
2. Miller SM, Michalak AM, Dssetmers RG, et al. China’s coal mine methane regulations have not curbed growing emissions. Nat Commun. 2019;10:1-8.
3. Beeson M. Geoeconomics with Chinese characteristics: the BRI and China’s evolving grand strategy. Econ Polit Stud. 2018;6:240-256.
4. Wang Q, Li R. Decline in China’s coal consumption: an evidence of peak coal or a temporary blip? Energy Pol. 2017;108:696-701.
5. Han Z-Y, Fan Y, Jiao J-L, Yan J-S, Wei Y-M. Energy structure, marginal efficiency and substitution rate: an empirical study of China. Energy. 2007;32:935-942.
6. Ji Q, Zhang D. How much does financial development contribute to renewable energy growth and upgrading of energy structure in China? Energy Pol. 2019;128:114-124.
7. Yang X, Wang S, Zhang W, Li J, Zou Y. Impacts of energy consumption, energy structure, and treatment technology on SO2 emissions: a multi-scale LMDI decomposition analysis in China. Appl Energy. 2016;184:714-726.
8. Zhao T, Hu Y, Zang T, Cheng L. MRTFB regulates the expression of NOMO1 in colon. Proc Natl Acad Sci. 2020;202000499:202000499.
9. Peng J, Zhao T. Reduction in TOM1 expression exacerbates Alzheimer’s disease. Proc Natl Acad Sci. 2020;117:3915-3916.
10. Zhao T, Hu Y, Zang T, Cheng L. Identifying Alzheimer’s disease-related proteins by LRRGD. BMC Bioinform. 2019;20:570.
11. Zeng Y-R, Zeng Y, Choi B, Wang L. Multifactor-influenced energy consumption forecasting using enhanced back-propagation neural network. Energy. 2017;127:381-396.
12. Wang D, Nie R, Long R, Zhao Y. Scenario prediction of China’s coal production capacity based on system dynamics model. Resour Conserv Recycl. 2018;129:432-442.
13. Wang X, Luo D, Zhao X, Sun Z. Estimates of energy consumption in China using a self-adaptive multi-verse optimizer-based support vector machine with rolling cross-validation. Energy. 2018;152:539-548.
14. Soldo B. Forecasting natural gas consumption. Appl Energy. 2012;92:26-37.
15. Wang L, Lv S-X, Zeng Y-R. Effective sparse adaboost method with ESN and FOA for industrial electricity consumption forecasting in China. Energy. 2018;155:1013-1031.
16. Wang L, Hu H, Ai X-Y, Liu H. Effective electricity energy consumption forecasting using echo state network improved by differential evolution algorithm. Energy. 2018;153:801-815.
17. Hu H, Wang L, Peng L, Zeng Y-R. Effective energy consumption forecasting using enhanced bagged echo state network. Energy. 2020;193:116778.
18. Yuan C, Liu S, Fang Z. Comparison of China’s primary energy consumption forecasting by using ARIMA (the autoregressive integrated moving average) model and GM (1, 1) model. Energy. 2016;100:384-390.
19. Xu N, Dang Y, Gong Y. Novel grey prediction model with nonlinear optimized time response method for forecasting of electricity consumption in China. Energy. 2017;118:473-480.
20. Wang Q, Li S, Li R. Forecasting energy demand in China and India: using single-linear, hybrid-linear, and non-linear time series forecast techniques. Energy. 2018;161:821-831.
21. Zhang J, Chi Y, Xiao L. Solar power generation forecast based on LSTM. In: 2018 IEEE 9th International Conference on Software Engineering and Service Science (ICSCESS). IEEE; 2018:869-872.
22. Tang L, Yu L, Wang S, Li J, Wang S. A novel hybrid ensemble learning paradigm for nuclear energy consumption forecasting. Appl Energy. 2012;93:432-443.
23. Chen P, Niu A, Jiang W, Liu D. Air pollutant prediction: comparisons between LSTM, Light GBM and random forest. In: Geophysical Research Abstracts; 2019.
24. Zhao T, Cheng L, Zang T, Hu Y. Peptide-major histocompatibility complex class i binding prediction based on deep learning with novel feature. Front Genet. 2019;10:1191.
25. Zhao T, Hu Y, Zang T, Wang Y. Integrate GWAS, eQTL, and mQTL data to identify Alzheimer’s disease-related genes. Front Genet. 2019;10:1021.
26. Masi J, Meier U, Cireşan D, Schmidhuber J. Stacked convolutional auto-encoders for hierarchical feature extraction. In: International Conference on Artificial Neural Networks. Springer; 2011:52-59.
27. Liu H, Mi X, Li Y. Smart deep learning based wind speed prediction model using wavelet packet decomposition, convolutional neural network and convolutional long short term memory network. Energy Convers Manage. 2018;166:120-131.
28. Peng L, Liu S, Liu R, Wang L. Effective long-term memory with differential evolution algorithm for electricity price prediction. Energy. 2018;162:1301-1314.
29. Dudley B. BP statistical review of world energy. BP Stat Rev, London, UK. 2018:6.
30. Ma X, Liu Z-B. The kernel-based nonlinear multivariate grey model. Appl Math Model. 2018;56:217-238.
31. Moraes R, Valiati JF, Neto WPG. Document-level sentiment classification: an empirical comparison between SVM and ANN. *Expert Syst Appl*. 2013;40:621-633.

32. Ahmad A, Hassan M, Abdullah M, et al. A review on applications of ANN and SVM for building electrical energy consumption forecasting. *Renew Sustain Energy Rev*. 2014;33:102-109.

---

**How to cite this article:** Li Y, He Y, Zhang M. Prediction of Chinese energy structure based on Convolutional Neural Network-Long Short-Term Memory (CNN-LSTM). *Energy Sci Eng*. 2020;8:2680–2689. [https://doi.org/10.1002/ese3.698](https://doi.org/10.1002/ese3.698)