MDPI

Article

Comparison and Determination of Optimal Machine Learning Model for Predicting Generation of Coal Fly Ash

Chongchong Qi [1,2,*], Mengting Wu [1], Xiang Lu [2], Qinli Zhang [1] and Qiusong Chen [1]

[1] School of Resources and Safety Engineering, Central South University, Changsha 410083, China; qinli.zhang@csu.edu.cn (Q.Z.); qiusong.chen@csu.edu.cn (Q.C.)

[2] State Key Laboratory of Coal Resources and Safe Mining, China University of Mining and Technology, Xuzhou 221116, China; xiang.lu@cumt.edu.cn

[*] Correspondence: chongchong.qi@csu.edu.cn

Abstract: The rapid development of industry keeps increasing the demand for energy. Coal, as the main energy source, has a huge level of consumption, resulting in the continuous generation of its combustion byproduct coal fly ash (CFA). The accumulated CFA will occupy a large amount of land, but also cause serious environmental pollution and personal injury, which makes the resource utilization of CFA gradually to be attached importance. However, given the variability of the amount of CFA generation, predicting it in advance is the basis to ensure effective disposal and rational utilization. In this study, CFA generation was taken as the target variable, three machine learning (ML) algorithms were used to construct the model, and four evaluation indices were used to evaluate its performance. The results showed that the DNN model with the $R = 0.89$, $R^2 = 0.77$ on the testing set performed better than the traditional multiple linear regression equation and other ML algorithms, and the feasibility of DNN as the optimal model framework was demonstrated. Applying this model framework to the engineering field enables managers to identify the next step of the disposal method in advance, so as to rationally allocate ways of recycling and utilization to maximize the use and sales benefits of CFA while minimizing its disposal costs. In addition, sensitivity analysis further explains ML's internal decisions and verifies that coal consumption is more important than installed capacity, which provides a certain reference for ensuring the rational utilization of CFA.

Keywords: CFA; generation; machine learning; multiple linear regression; sensitivity analysis; utilization

1. Introduction

The acceleration of industrial processes has led to the rapid development of the energy industry as a large player. As the main energy supplier, power stations based on coal and lignite provide massive amounts of energy, and coal consumption has soared [1]. Although electricity demand and coal emissions experienced a small decline in 2020, as the COVID-19 outbreak depressed energy demand, the economic stimulus package and the rollout of vaccines promoted the economic rebound, leading to a 9% increase in coal-fired power generation in 2021, its highest level ever [2]. In the first half of 2021, coal market consumption showed an 11% year-on-year growth. Coal consumption in the European Union is expected to increase by 4% by the end of the year [3], and coal may remain a mainstay of international energy in the short term.

The huge consumption of coal makes the generation of coal fly ash (CFA), as a byproduct of coal combustion [4], continue to increase [5], particularly in India. Over a 10-year span (2009–2010 to 2018–2019), CFA in the power sector increased by nearly 76% and is now producing around 217 million tones [6]. A large amount of deposition of CFA takes up land resources and directly pollutes the soil [7], in addition to causing serious impacts on water and air due to inappropriate treatment. On the one hand, a large amount of rainfall makes CFA landfills a potentially dangerous place, producing toxic leachate that seeps into

groundwater and pollutes water resources [8]; on the other hand, toxic elements may be discharged into the air with the flue gas produced, endangering air quality. Li et al. have pointed out that solid Hg waste produced by fly ash from coal burning is the main source of Hg in the environment [9]. Moreover, human health is also at risk from long-term exposure to CFA diffusing into the air or from drinking contaminated groundwater [10]. As a result, academia and industry are paying increasing attention to the resource utilization of CFA.

In recent years, CFA has gradually been effectively utilized in various fields, among which the construction industry is the most widely used. CFA is used as the supplementary cementitious material to partially replace cement in concrete or to prepare geopolymers [11] and is also used as coarse and fine aggregate in asphalt pavement [12]. Due to its potential for soil improvement and heavy metal adsorption [13], CFA has a good development prospect in the agricultural field. Moreover, CFA also has a large presence in the manufacturing of ceramic glass [14], metal matrix composites, and metal coatings.

However, the amount of generation of CFA is variable, so predicting the generation of CFA in advance is the basis for ensuring its effective disposal and rational utilization. Some scholars used neural networks in MATLAB and linear regression statistical analysis in IBM SPSS to predict the generation of CFA in power plants in five or ten years [15]. However, the description of this method is too simple and general, and the accuracy of the prediction has not been verified. Others predicted the average annual output of hazardous wastes by multiplying the amount of industrial hazardous wastes generated in the base year by the average annual growth rate index [16], which is too complicated and time-consuming, and the accuracy cannot be guaranteed either.

In view of the limitations of the above prediction methods, this paper used advanced machine learning (ML) algorithm [17] to predict the generation of CFA by constructing three different regression models, of which installed capacity and coal consumption were input variables, and the generation of CFA was output variable. The established model framework can be applied to the engineering site after thorough evaluation and comparison, which can quickly and accurately predict the amount of CFA generation, thus saving time for further planning of CFA disposal, and is the basis for reasonable recycling of CFA.

2. Dataset

Data are the basis for the development of machine learning algorithms, and any ML algorithms need data to evaluate their effects. How to collect a comprehensive and appropriate dataset and analyze it was key to this research.

2.1. Data Collection

In this study, domestic and foreign databases and related academic websites were searched, a large number of studies in the literature and academic reports related to CFA were consulted, the relevant data were sorted out and recorded, and finally, the dataset used in this paper was obtained through screening. This dataset was extracted from a report documenting CFA generation and utilization in coal-fired power plants across India in 2019–2020 and contained data from 183 power plants across 17 states (outliers with a coal consumption value of 0 were removed) [18], as shown in Figure 1. Chhattisgarh was the most sampled state, with 27 power plants, accounting for 14.8%, but only one power plant was sampled in Assam. From a holistic perspective, the distribution of sampling sites was relatively uniform.

2.2. Data Analysis

The ultimate purpose of the algorithm is to fit the distribution of the data and predict the trend of change. Different datasets have different feature distributions, so the statistical distribution and correlation analysis of data serve as sources of reference for establishing the optimal algorithm model.

In this paper, the dataset with a distribution characteristic presented in the form of a bubble chart included two features: installed capacity and coal consumption, and the target

variable was the generation of CFA. As shown in Figure 2, the size of bubbles represents the CFA generation. Data points were mainly distributed in the lower-left corner, meaning that when the installed capacity was between 0 and 2000 MW, and the coal consumption varied from 0 to 5 MT, the amount of generation of CFA was small, less than 2.95 MT. With the increase in installed capacity and coal consumption, the CFA generation also increased. The maximum generation of 8.85 MT was realized when the installed capacity was 4760 MW, and the coal consumption was about 25 MT.
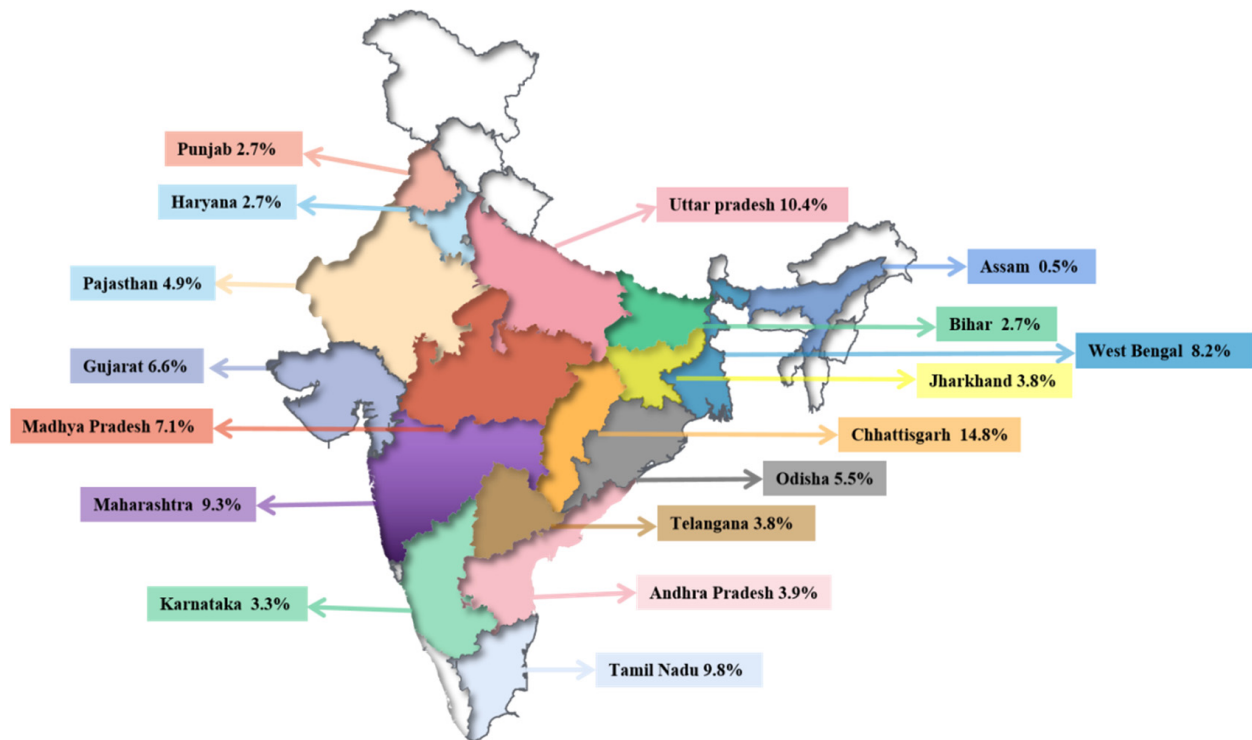

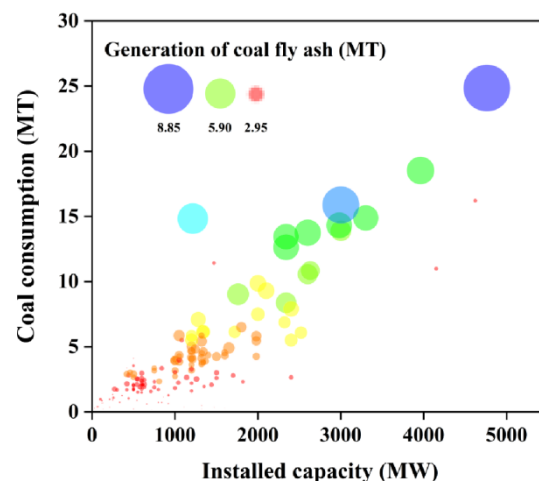**Figure 1.** Distribution of data sampling in India.

Figure 2. Bubble chart of data distribution. Note that the size of the bubble represents the generation of CFA. Purple bubbles represent CFA maximum generation and red represents minimum generation.

Correlations between features or between features and target variables were measured by Pearson correlation coefficients (R), which were between -1 and 1, as shown in Figure 3. The correlation degree between coal consumption and generation of CFA was the highest, and R was 0.9. Meanwhile, R between installed capacity and target variable was 0.73, indicating a strong correlation between installed capacity and CFA generation.
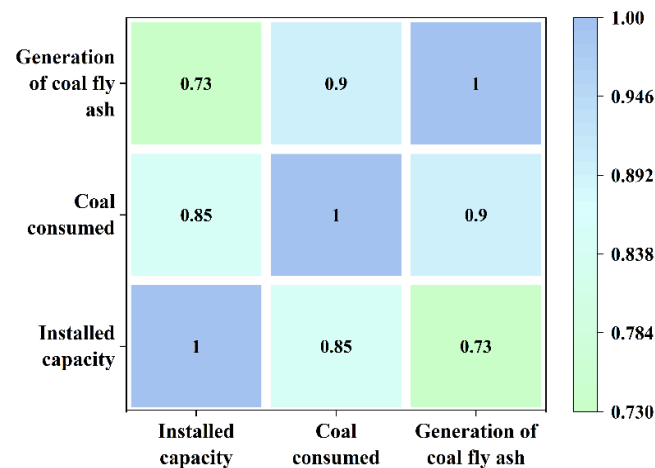
Figure 3. Correlation heat map.

3. Methodology

To achieve rapid and accurate prediction of CFA generation, Python 3.8 programming language was used in this paper, and three machine learning algorithms were selected to construct the model framework using the scikit-learn library [19]. Four evaluation indices were used to measure the performance of the model [20]. Finally, the optimal model was determined according to the evaluation results and compared with traditional methods to verify the feasibility and superiority of the prediction framework. The specific methodology is shown in Figure 4.
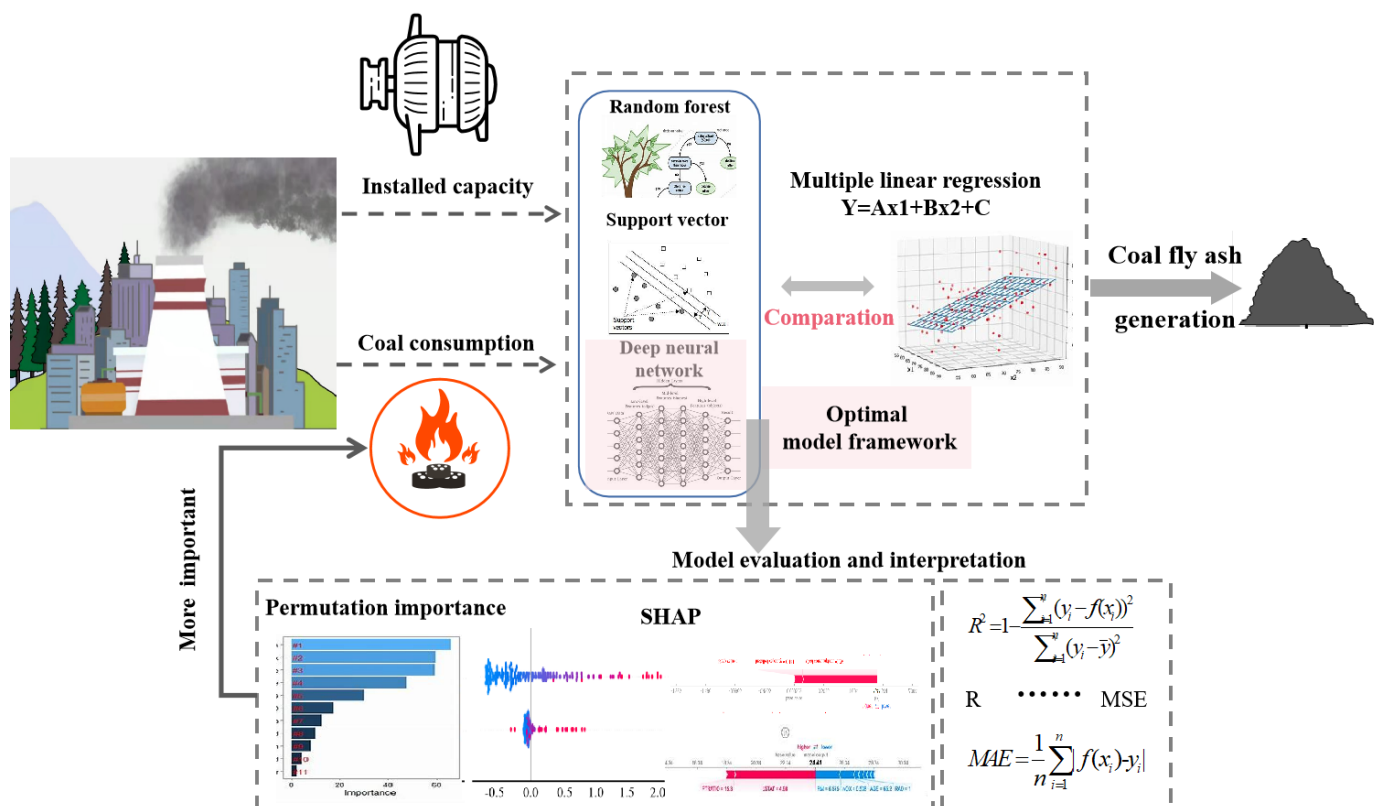

Figure 4. Complete diagram of methodology.

3.1. Modeling Methods

Machine learning algorithms used in this study had a general modeling process. Firstly, the original data were preprocessed, including the removal of outliers or normalization (this

part is explained in detail in Section 3.2). Then, coal consumption and installed capacity after treatment were taken as input variables, and CFA generation was taken as the output variable. Then, training, evaluation, and prediction were carried out by random forest, support vector machine, and neural network. The specific principles and steps of the three algorithms are as follows:

3.1.1. Random Forest

Random forest (RF) is an integration algorithm that combines the outputs of multiple decision trees into one result to deal with classification and regression problems [21]. It has the characteristics of ease of use and flexibility [22]. The construction of RF includes the following four main steps:

1. Random sampling and training decision tree: The original data population with sample size N is randomly sampled N times, and each time, the samples need to be put back [23]. N samples formed at last are used to train a decision tree;
2. Randomly selected attributes as node-splitting attributes: When the nodes of the decision tree are split, m attributes ($m << M$) should be randomly selected from the M attributes of each sample, and then some strategies (such as information gain) should be adopted to select one attribute as the final split attribute of the node;
3. Step 2 is repeated until the tree cannot be split, noting that no pruning occurs during the entire decision tree formation process;
4. A large number of decision trees are established according to steps 1~3 to form an RF.

3.1.2. Support Vector Regression

Support vector regression (SVR) is an important branch of support vector machine (SVM) [24]. SVR has only one type of sample point in the end. The optimal hyperplane it seeks is to minimize the total deviation of all sample points from the hyperplane.

Different from traditional regression methods, SVR indicates that, as long as the deviation degree of $f(x) = \omega^T \Phi(x) + b$ and y is not too large, the prediction can be considered correct without calculating the loss. SVR can obtain a regression model in the form of $f(x) = \omega^T \Phi(x) + b$ by inputting the training sample set $X = (x_i, y_i)_{i = 1 \sim N, y_i \in R}$ [25], where $\Phi(x)$ is the vector mapped to X, $\omega = (\omega_1, \omega_2, \ldots \omega_n)$ is the normal vector, and b is the intercept. Then, an interval band with a distance of ε is created on both sides of the linear function (tolerance deviation) [26]. The loss is not calculated when all samples fall into the interval band but is calculated only when the absolute value of the gap between $f(x)$, and y is greater than ε. Finally, the optimized model is obtained by minimizing the total loss and maximizing the interval.

3.1.3. Deep Neural Network

A deep neural network (DNN) is an extension based on perceptron. The internal structure of DNN has only one input layer and one output layer, but there are multiple hidden layers in the middle [27]. Each layer of the neural network has several neurons. The neurons between layers are connected to each other but are not within a layer, and the neurons in the next layer are connected to all the neurons in the previous layer [28,29].

Generally speaking, the steps of constructing a DNN structure include the following three points: (1) network construction, (2) assignment parameters, and (3) iterative calculation. The main principles of iterative calculation include forward-propagation (FP) and back-propagation (BP) algorithms [30].

The FP algorithm uses several weighted coefficient matrices *W* and bias vector *B* to carry out a series of linear operations and activation operations with input vector *X*. Starting from the input layer, the output of the previous layer is used to calculate the output of the next layer, and then one layer after another is calculated until it reaches the output layer, and the predicted value *Y* is obtained. In comparison, the BP algorithm uses the gradient descent method to iteratively optimize the loss function to obtain the minimum value [31]. Additionally, it then seeks the appropriate linear coefficient matrix *W* and bias vector *B*

corresponding to the hidden layer and output layer, so that the output calculated by all the input of training samples is equal or close to the sample label as far as possible [32].

In short, FP is the recognition process of the predicted value Y, while BP is the reverse adjustment of parameters W and B according to the difference between the target y and the predicted Y. After repeated forward- and back-propagation training, the neural network model with high accuracy is finally formed.

3.2. Dataset Preprocessing and Splitting

The dimensionality and its unit of evaluation index (feature) affect the result of data analysis. To eliminate the influence of dimension between indicators, standardizing or normalizing data to achieve comparability between data indicators are generally adopted [33].

The variance ratio between features and target variables of the dataset in this paper was 200:4:2. There are several orders of magnitude differences between the variances, which leads to features with large variances dominating the algorithm, resulting in poor modeling performance [34]. Therefore, the "processing" module in the sklearn was used to standardize data (sklearn.preprocessing.scale) whose outliers has been removed.

The preprocessed dataset was divided into training and testing sets. Among them, the training set was used to train the model, whereas the testing set was used to verify the final effect of the model. In view of the impact of the division ratio of the dataset on model performance, the size of the testing set in this paper varied from 10% to 45% with an interval of 5%, and R was used as the evaluation index to determine the optimal division ratio [22].

3.3. Model Evaluation

After the model was constructed, it was necessary to evaluate its effect and then select the optimal model by comparison. In this paper, four common indicators—namely, R, R squared (R^2), mean-squared error (MSE), and mean absolute error (MAE)—were used to evaluate the model. The calculation formulas are as follows:

$$R = \frac{\sum\limits_{i=1}^{n} (f(x_i) - \overline{f(x_i)})(y_i - \overline{y})}{\sqrt{\sum\limits_{i=1}^{n} (f(x_i) - \overline{f(x_i)})^2} \sqrt{\sum\limits_{i=1}^{n} (y_i - \overline{y})^2}} \tag{1}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - f(x_i))^2}{\sum_{i=1}^{n} (y_i - \overline{y})^2} \tag{2}$$

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (f(x_i) - y_i)^2 \tag{3}$$

$$MAE = \frac{1}{n} \sum_{i=1}^{n} | f(x_i) - y_i| \tag{4}$$

where n represented the number of samples, y_i was the real observed value, $\overline{y}$ represented the average of the real value, and $f(x_i)$ was the predicted value, with a mean value of $\overline{f(x_i)}$.

As introduced in Section 2.2, R is used to reflect the degree of linear correlation between two variables; in addition, R^2 is used to judge the degree of fit between the prediction model and the real data [35]. The best value of R^2 is 1 and can be negative. MSE calculates the mean of the sum of squares of sample point errors corresponding to the fitting data and original data, and the smaller the value is, the better the fitting effect is [36]. MAE is used to evaluate how close the predicted results are to the real dataset, and the smaller the MAE, the better the model [37].

4. Result and Discussion

4.1. Determination of Dataset Division Ratio

To avoid the randomness of the evaluation results, in this paper, we evaluated each division ratio of the dataset 50 times repeatedly and took the mean value of the correlation coefficient R as the final performance of the model under a specific partition. As shown in Figure 5, the RF model was taken as an example. For the training set, the influence of the division ratio on the modeling performance was small, and the R fluctuated, by a small margin, around 0.98. Focusing on the testing set, when it accounted for 10% of the dataset, R was 0.84; when the size of the testing set was 15%, the performance of the model reached the highest, satisfying $R = 0.87$. After that, R generally decreased, with a further increase in division ratio up to 45%. In summary, the RF model performed best when the size of the testing set was 15%, and the analysis results of SVR and DNN were consistent with it. Therefore, the training set:testing set = 0.85:0.15 ratio was determined as the optimal division ratio.
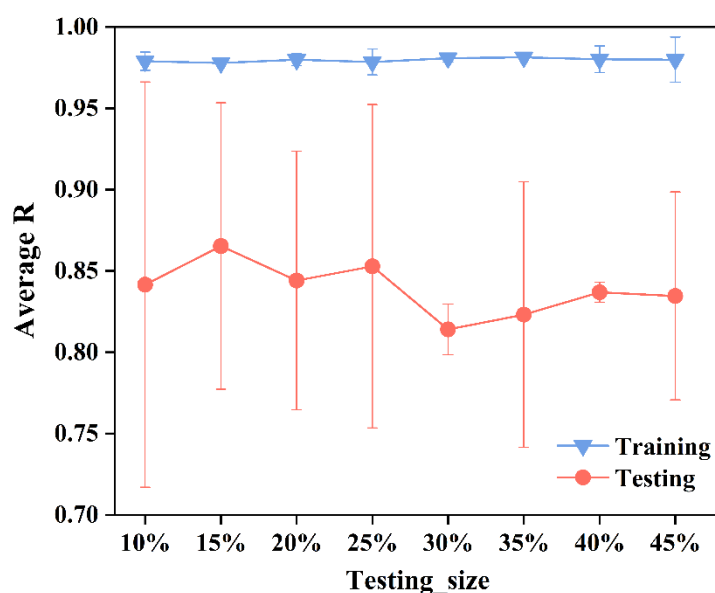

Figure 5. Optimal division ratio of dataset on RF model.

4.2. Parameters of the Model

In this study, RF and SVR, as traditional machine learning regression algorithms, were trained with corresponding default parameters in the ensemble module of sklearn, as shown in Table 1. DNN is a deep learning model in which performance is greatly affected by network structure and parameters [38]. Based on the trial-and-error method and suggestions in references [39,40], the neural network layer, learning rate, activation function, and epoch were constantly changed during the model training process, and 10% of the data were separated for performance verification. As shown in Figure 6, when the loss on the validation set tends to fluctuate stably with the increase in steps, the DNN model that included one input layer, five hidden layers, and one output layer was finally determined. The number of neurons in each layer was 2→8→32→64→16→8→1. To speed up convergence based on the gradient descent method and prevent overfitting, two "Batch normalization" layers and one "dropout" layer were also included. The specific network structure and parameters are shown in Figure 7 and Table 2.

Table 1. Default hyperparameters for RF and SVR models.

RF		SVR	
Parameters	**Default Value**	**Parameters**	**Default Value**
n_ estimators	100	kernel	'rbf'
min_ samples_ split	2	degree	3
min_ samples_ leaf	1	gamma	scale
max_ features	'auto'	C	1
max_ depth	None	epsilon	0.1
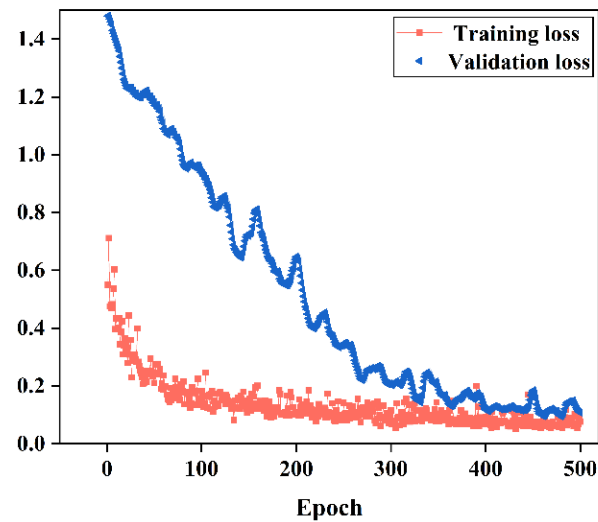
Figure 6. Loss of training set and testing set.
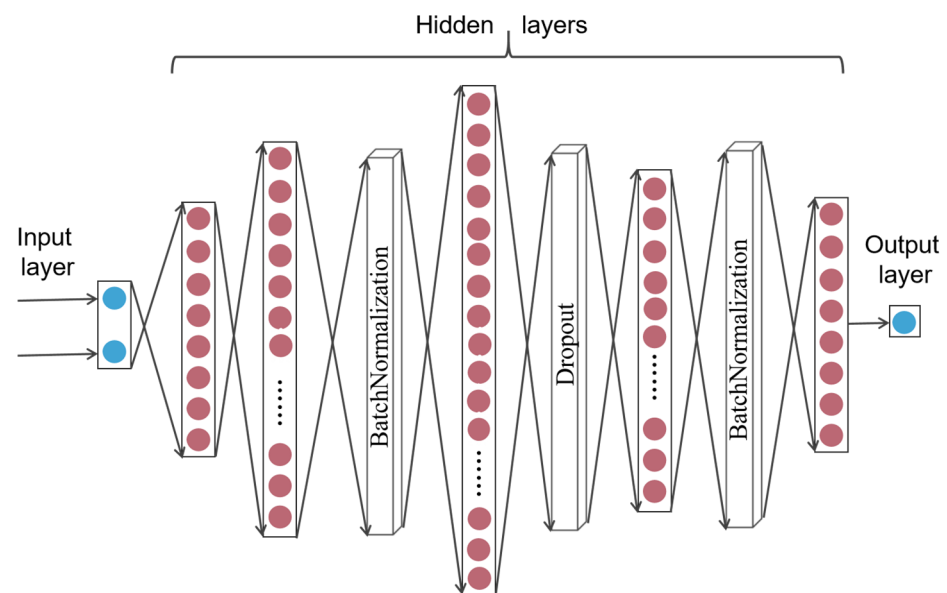

Figure 7. Network structure of optimal DNN model.

Table 2. Specific parameters of DNN structure.

| Parameters	Option or Value	Implication
Activation function	Relu	The output is no longer a linear combination of the inputs and can approximate any function
Optimizer	Adam	A hybrid of momentum gradient descent and RMSprop.
Learning rate	0.0005	The weight of neural network input is adjusted.
Batch size	128	Number of samples is used for training.
Epoch	500	One epoch is equal to training with all the samples in the training set.

4.3. Comparative Analysis of Model Performance

To obtain a reliable model, fivefold-cross-validation was adopted for RF and SVR models (cross_val_predict), while DNN used the parameter "validation_split" to perform simple cross-validation. Moreover, as the result of a simple random partition is accidental, it cannot represent the actual performance of the model. Therefore, modeling and evaluation for three ML models were repeated 50 times on the training and testing sets, respectively, and the evaluation indexes were averaged as the final performance of the model.

As shown in Figure 8, the linear fitting functions between the prediction results of RF model on the training and testing sets, and the real values were $y = 0.878x + 0.134$ and $y = 0.861x + 0.094$, while those of the DNN model were $y = 0.790 + 0.375$ and $y = 0.837x + 0.316$. All data points were relatively concentrated on the two curves, and the p values were 4.68E-30, 1.30E-27, 8.35E-73, and 0.00000000000109, respectively, which were less than the significance level of 0.05. In addition, R and R^2 were relatively high, indicating the good performance of the models. Moreover, the linear regression between actual and SVR-estimated generation of CFA was $y = 0.451x + 0.614$ and $y = 0.861x + 0.094$, respectively, on the training and testing sets. Compared with RF and DNN models, the SVR model had relatively discrete data distribution on the training set, and its performance was slightly worse.
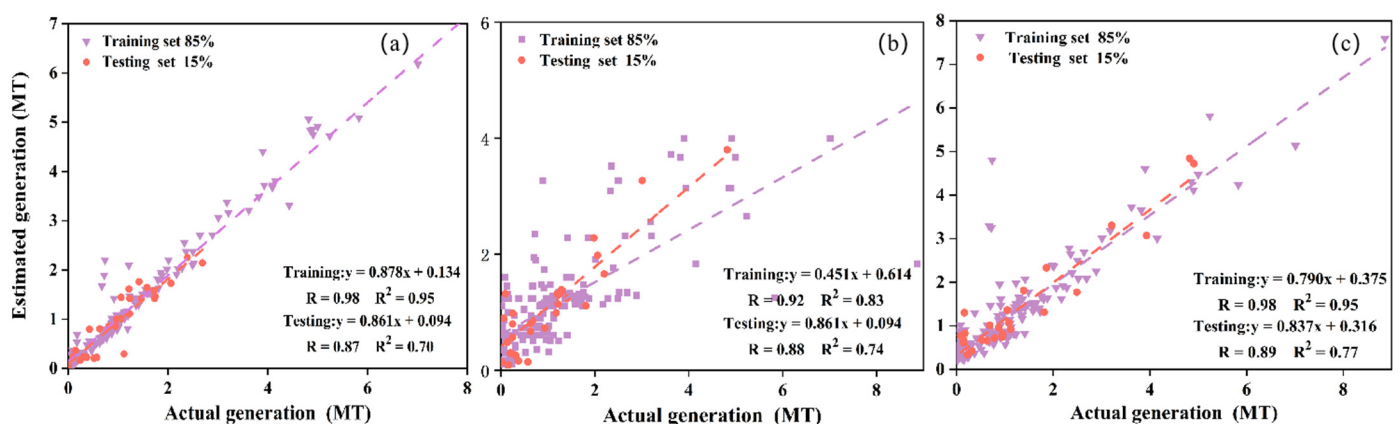

Figure 8. Scatter distribution fitting of generation of CFA under three modeling frameworks: (**a**) RF model, (**b**) SVR model, and (**c**) DNN model.

As can be seen from Figure 9, the difference between the actual and estimated generation of CFA was small in the three models, and the data were mostly concentrated around 0, indicating the good prediction performance of ML models. The probability of data points in RF and DNN models appearing in the small interval [−0.1,0.1] was close to 0.9, while that of SVR was only 0.45. In addition, for the testing set, the data points on the DNN model were more concentrated in the areas with smaller differences, which implied that the DNN model had higher prediction accuracy.
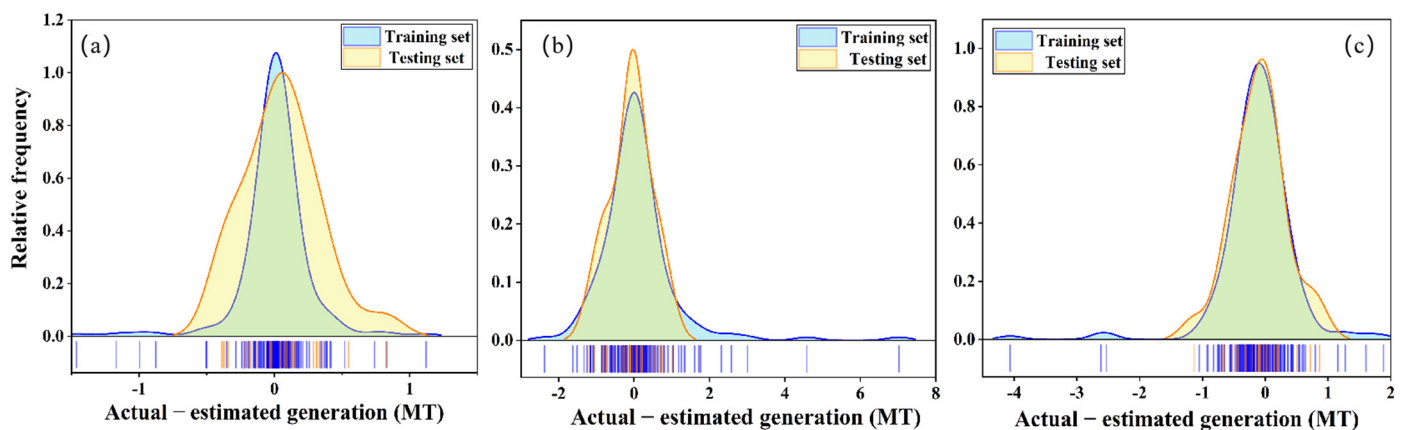
Figure 9. The relative frequency of the difference between actual and estimated CFA generation: (**a**) RF model, (**b**) SVR model, and (**c**) DNN model.

Figure 10 shows a comparison of the performance of the three models more intuitively with four evaluation indices. For the training set in Figure 10a, R and R^2 values of RF and DNN models were the same, which were 0.98 and 0.95, respectively, and slightly higher than those of SVR models, which were 0.92 and 0.83. Meanwhile, the *MSE* and *MAE* values of the RF model were the smallest of the three models. On the contrary, the RF model had the lowest R and R^2 values on the testing set of Figure 10b, which were 0.87 and 0.7. However, R and R^2 values of the DNN model were the highest, which were 0.89 and 0.77, and *MSE* and *MAE* were relatively low. In general, The DNN model was the optimal model framework suitable for the CFA dataset in this study.
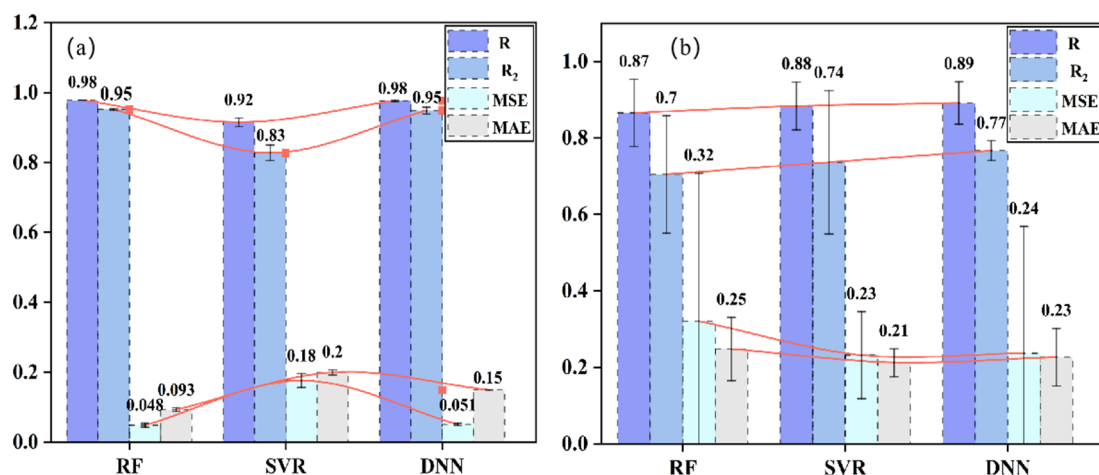
Figure 10. Comparison of four evaluation indicators under three modeling frameworks: (**a**) training set, (**b**) testing set.

4.4. Comparison with Multiple Linear Regression

Multiple linear regression is a conventional data analysis method that uses multiple independent variables to predict or estimate dependent variables [41]. In this method, the dataset was repeatedly divided 50 times according to the same ratio of 0.85:0.15, and the multiple linear regression equation $Y = Ax_1 + Bx_2 + C$ was established. The average results of statistical analysis are shown in Table 3. After 50 evaluations, the mean R^2 and R of the multiple regression training set were 0.82 and 0.90, which were lower than the results of the three ML models using fivefold cross-validation. Then, the data of the testing set were put into the equation for verification, and the mean values of R and *p*-value were 0.86 and 0.0000643805, respectively, indicating a significant correlation between the results. However, the mean value of R^2 was 0.76, which was higher than the RF and SVR models

but lower than that of the DNN models, which once again proved that the DNN model was more suitable for the dataset. The specific results are in the attachment.

Table 3. Statistical parameter analysis of multiple linear regression.

	$Y = A \times 1 + B \times 2 + C$					
	Regression Coefficient	**95% LCL**	**95% UCL**	**SE**	**T**	***p*-Value**
Installed capacity (A)	−0.221283602	−4.6257086	−2.89796	2.5330302	−2.4167858	0.05265861
Coal consumption (B)	0.365218825	0.310336	0.404628	0.0228592	15.6097245	3.49436E-29
Constant (C)	0.173085366	0.1951878	0.318641	0.0759174	2.188839	0.041769567

4.5. Feature Analysis

In this section, the analysis of the sensitivities of two features that affect the generation of CFA is presented using the permutation importance provided by sklearn and eli5, and "TreeExplainer" and "KernelExplainer" in the Shapley Additive Interpretation (SHAP) library.

4.5.1. Permutation Importance

The evaluation of the sensitivity of the feature depends on the degree of degradation of the model performance score after the feature is randomly rearranged [42]. As shown in Figure 11, after the values of coal consumption were randomly shuffled, the decrease in *MSE* of RF, SVR, and DNN algorithms were 1.73, 1.07, and 1.89, respectively, which were generally higher than those in the case of installed capacity randomly disturbed. This proved that the three models reached a consensus on the view that coal consumption had a greater impact on the generation of CFA.
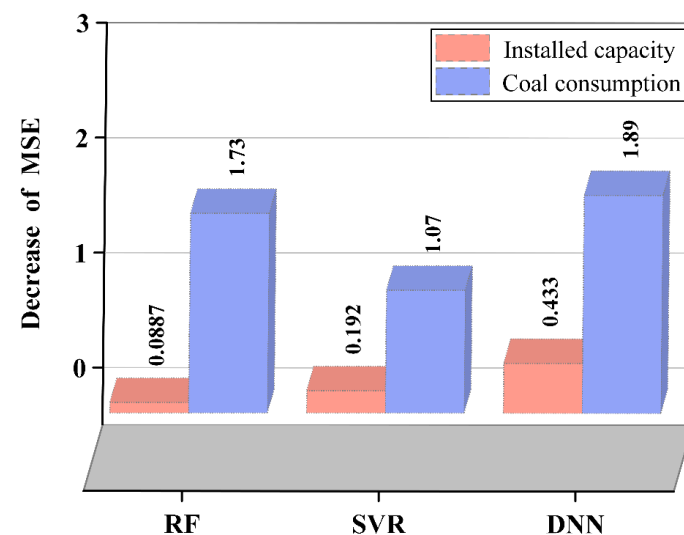

Figure 11. Permutation importance under three machine learning algorithms.

4.5.2. SHAP

SHAP is a model agnostic interpretation method that can be used for both global and individual applications. SHAP can judge which feature is more important, as well as reflect the positive and negative influences of features on the target variable [43]. The model generates a predictive value for each sample, and the SHAP value is the contribution value assigned to each feature in the sample [44].

To better understand the overall pattern, Figure 12a shows the results of calculated SHAP values for each feature of each sample. Among them, features were arranged from top to bottom in order of importance on the *y* axis [45], which indicated that coal consumption had a greater influence on the model, consistent with permutation importance.

In addition, the color represented the feature value (red was high, blue was low [46]). It can be seen that, under the three algorithm models, higher coal consumption increased the predicted generation of CFA. However, for installed capacity, the results were different. In RF and SVR models, larger installed capacity increased the predicted generation of CFA, but in DNN, the result was completely opposite. As shown in Figure 12b, the first sample for which preprocessed feature values were 0.8059 and 1.363 was used as an example to explain the generation details of a single prediction. In the figure, the red bar represents the range in which a feature played a positive role in the prediction of the model [47]; the base value was the mean value of the target variables of all samples, and $f(x)$ was the final predicted value for this sample, which satisfies $f(x)$ = base value + $\sum$SHAP value. The analysis showed that the prediction results of the three algorithm models were slightly different for the same sample, which may be affected by the algorithm principle and disrupted data. However, coal consumption and installed capacity both played positive, driving roles in the prediction of the model, but coal consumption had a greater impact.
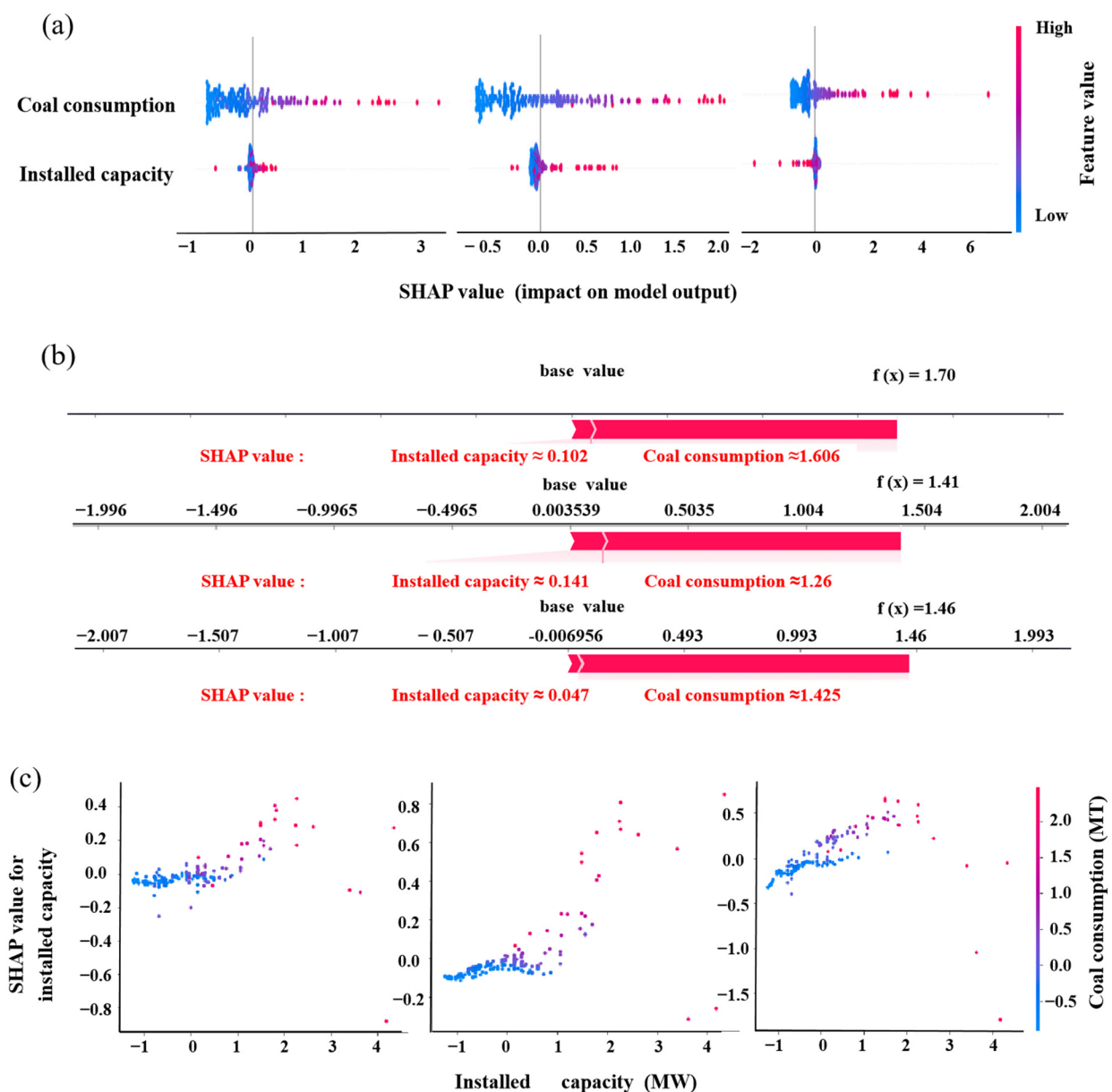
Figure 12. Features interpretation from three perspectives using SHAP: (**a**) global explanation, (**b**) local explanation, and (**c**) feature interaction. It is worth noting that the RF, SVR, and DNN models are from top to bottom or left to right in the three subgraphs.

In addition to explaining the model globally and locally, Figure 12c revealed hidden relationships among features through quick, precise interactions. The analysis showed that the interaction between coal consumption and installed capacity had positively correlated influences on CFA generation prediction. Specifically, when both coal consumption and installed capacity were high, the installed capacity had a great influence on the generation of CFA, except for some outliers. On the contrary, when the coal consumption and installed capacity were relatively small, the installed capacity contributed little to the variation in the model output and even hindered the prediction.

As indicated above, the effect of installed capacity on the CFA generation was not always positive, compared with that of coal consumed. In real life, the installed capacity is the designed capacity for one specific powder station. The actual capacity is influenced by many external factors, such as coal production, the market, policies, etc. Therefore, the correlation between installed capacity and CFA generation was not as close as the correlation between coal consumed and CFA generation. Moreover, it is possible that a power station with a large installed capacity produced a relatively small amount of power, and thus CFA, due to the influence of the above-mentioned factors. ML models based on datasets with such special cases might indicate the negative influence of installed capacity for some data samples.

5. Significance and Outlook

High energy consumption leads to increased generation of solid wastes such as CFA, posing a potential threat to the environment and human health. Meanwhile, more CFA byproducts are gradually being recycled and utilized to achieve sustainability [48]. However, the uncertainty of CFA generation poses difficulties to the rational planning and design of its disposal and utilization. The optimal model framework constructed in this study can quickly and accurately predict the generation of CFA only by inputting coal consumption and installed capacity, which is feasible and efficient. Applying this model framework to the engineering field enables managers to identify the next step of the disposal method in advance, so as to rationally allocate ways of recycling and utilization to maximize the use and sales benefits of CFA while minimizing its disposal costs. However, due to the small size of the dataset and few input variables, the results of this model framework lack further validation, and its general application needs to be improved. Subsequent studies can expand the search scope and consider various factors affecting CFA generation.

6. Conclusions

DNN was determined as the optimal ML model through comparative evaluation, which can accurately predict the generation of CFA. In addition, the sensitivity analysis of the features also provided a certain point of reference for ensuring the rational utilization of CFA. The specific conclusions are as follows:

(1) Among the three model algorithms, the DNN model had the best performance. R and R^2 on the training set were 0.98 and 0.95, whereas these on the testing set were 0.89 and 0.77, respectively;

(2) The R^2 of the traditional multiple linear regression equation on the testing set was 0.76, higher than those of RF and SVR models, but lower than that of the DNN model;

(3) Permutation importance and SHAP both indicated that coal consumption had a greater positive effect on the generation of CFA. As influenced by other factors, the influence of installed capacity on CFA generation was as significant as coal consumed and could be negative for some special data samples.

Author Contributions: Conceptualization, C.Q.; methodology, M.W. and C.Q.; validation, M.W.; investigation, M.W. and C.Q.; data curation, M.W. and C.Q.; writing—original draft preparation, M.W.; writing—review and editing, All authors; visualization, All authors; supervision, C.Q. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

References

1. Naqvi, S.Z.; Ramkumar, J.; Kar, K.K. Coal-based fly ash. In *Handbook of Fly Ash*; Kar, K.K., Ed.; Butterworth-Heinemann: Oxford, UK, 2022; pp. 3–33. [CrossRef]
2. IEA. *Electricity Market Report*; IEA: Paris, France, 2022; Available online: https://www.iea.org/reports/electricity-market-report-january-2022 (accessed on 25 February 2022).
3. IEA. *Global Energy Review*; IEA: Paris, France, 2021; Available online: https://www.iea.org/reports/global-energy-review-2021 (accessed on 25 February 2022).
4. Wang, X.Y. Evaluation of the hydration heat and strength progress of cement-fly ash binary composite. *J. Ceram. Process. Res.* **2020**, *21*, 622–631. [CrossRef]
5. Mathapati, M.; Amate, K.; Durga Prasad, C.; Jayavardhana, M.L.; Hemanth Raju, T. A review on fly ash utilization. *Materials* **2022**, *50*, 1535–1540. [CrossRef]
6. Arora, S. *An Ashen Legacy: India's Thermal Power Ash Mismanagement*; Centre for Science and Environment: New Delhi, India, 2020.
7. Blaha, U.; Sapkota, B.; Appel, E.; Stanjek, H.; Rösler, W. Micro-scale grain-size analysis and magnetic properties of coal-fired power plant fly ash and its relevance for environmental magnetic pollution studies. *Atmos. Environ.* **2008**, *42*, 8359–8370. [CrossRef]
8. Chowdhury, A.; Naz, A.; Chowdhury, A. Waste to resource: Applicability of fly ash as landfill geoliner to control ground water pollution. *Materials* **2021**, *20*, 897. [CrossRef]
9. Chen, Q.; Chen, L.; Li, J.; Guo, Y.; Wang, Y.; Wei, W.; Liu, C.; Wu, J.; Tou, F.; Wang, X.; et al. Increasing mercury risk of fly ash generated from coal-fired power plants in China. *J. Hazard. Mater.* **2022**, *429*, 128296. [CrossRef]
10. Jiang, A.; Zhao, J. Experimental Study of Desulfurized Fly Ash Used for Cement Admixture. In *Proceedings of Civil Engineering in China–Current Practice and Research Report*; Hindawi: Hebei, China, 2010; pp. 1038–1042.
11. Ragipani, R.; Escobar, E.; Prentice, D.; Bustillos, S.; Simonetti, D.; Sant, G.; Wang, B. Selective sulfur removal from semi-dry flue gas desulfurization coal fly ash for concrete and carbon dioxide capture applications. *Waste Manag.* **2021**, *121*, 117–126. [CrossRef]
12. Shanmugan, S.; Deepak, V.; Nagaraj, J.; Jangir, D.; Jegan, S.V.; Palani, S. Enhancing the use of coal-fly ash in coarse aggregates concrete. *Mater. Today Proc.* **2020**, *30*, 174–182. [CrossRef]
13. Kotelnikova, A.D.; Rogova, O.B.; Karpukhina, E.A.; Solopov, A.B.; Levin, I.S.; Levkina, V.V.; Proskurnin, M.A.; Volkov, D.S. Assessment of the structure, composition, and agrochemical properties of fly ash and ash-and-slug waste from coal-fired power plants for their possible use as soil ameliorants. *J. Clean. Prod.* **2022**, *333*, 130088. [CrossRef]
14. Zhu, M.; Ji, R.; Li, Z.; Wang, H.; Liu, L.; Zhang, Z. Preparation of glass ceramic foams for thermal insulation applications from coal fly ash and waste glass. *Constr. Build. Mater.* **2016**, *112*, 398–405. [CrossRef]
15. Zahari, N.M.; Mohamad, D.; Arenandan, V.; Beddu, S.; Nadhirah, A. Study on prediction fly ash generation using statistical method. In Proceedings of the 3rd International Sciences, Technology and Engineering Conference (ISTEC), Penang, Malaysia, 17–18 April 2018.
16. Widyarsana, I.; Tambunan, S.A.; Mulyadi, A.A. Identification of Fly Ash and Bottom Ash (FABA) Hazardous Waste Genera-tion From the Industrial Sector and Its Reduc-tion Management in Indonesia. *Res. Sq.* **2022**. [CrossRef]
17. Cakir, M.; Guvenc, M.A.; Mistikoglu, S. The experimental application of popular machine learning algorithms on predictive maintenance and the design of IIoT based condition monitoring system. *Comput. Ind. Eng.* **2021**, *151*, 106948. [CrossRef]
18. Prakash, M. *Report on Fly Ash Generation at Coal/Lignite Based Thermal Power Stations and its Utilization in The Country for The Year 2019–2020*; Central Electricity Authority Government of India Ministry of Power: New Delhi, India, 2020.
19. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
20. Meiyazhagan, J.; Sudharsan, S.; Venkatesan, A.; Senthilvelan, M. Prediction of occurrence of extreme events using machine learning. *Eur. Phys. J. Plus* **2022**, *137*, 16. [CrossRef]
21. Li, H.; Lin, J.; Lei, X.; Wei, T. Compressive strength prediction of basalt fiber reinforced concrete via random forest algorithm. *Mater. Today Commun.* **2022**, *30*, 103117. [CrossRef]
22. Qi, C.; Wu, M.; Zheng, J.; Chen, Q.; Chai, L. Rapid identification of reactivity for the efficient recycling of coal fly ash: Hybrid machine learning modeling and interpretation. *J. Clean. Prod.* **2022**, *343*, 130958. [CrossRef]

23. Pi, J.; Jiang, D.; Liu, Q. Random Forest Algorithm for Power System Load Situation Awareness Technology. In *Application of Intelligent Systems in Multi-modal Information Analytics*; Sugumaran, V., Xu, Z., Zhou, H., Eds.; Springer International Publishing: Cham, Switzerland, 2021; pp. 925–929.

24. Wang, L. *Support Vector Machines: Theory and Applications. In Proceedings of Machine Learning and Its Applications*; Advanced Lectures; Springer: Berlin/Heidelberg, Germany, 2001.

25. Thalib, R.; Bakar, M.A.; Ibrahim, N.F. Application of support vector regression in krylov solvers. *Ann. Emerg. Technol. Comput.* **2021**, *123*, 178–186. [CrossRef]

26. Xia, Z.; Mao, K.; Wei, S.; Wang, X.; Fang, Y.; Yang, S. Application of genetic algorithm-support vector regression model to predict damping of cantilever beam with particle damper. *J. Low Freq. Noise Vib. Act. Control* **2017**, *36*, 138–147. [CrossRef]

27. Phapatanaburi, K.; Wang, L.; Oo, Z.; Li, W.; Nakagawa, S.; Iwahashi, M. Noise robust voice activity detection using joint phase and magnitude based feature enhancement. *J. Ambient. Intell. Humaniz. Comput.* **2017**, *8*, 845–859. [CrossRef]

28. Feng, C. Robustness Verification Boosting for Deep Neural Networks. In Proceedings of the 6th International Conference on Information Science and Control Engineering (ICISCE), Shanghai, China, 20–22 December 2019; pp. 531–535.

29. Liu, L.; Chen, J.; Xu, L. Realization and application research of BP neural network based on MATLAB. In Proceedings of the International Seminar on Future Biomedical Information Engineering, Wuhan, China, 18 December 2008; pp. 130–133.

30. Silaban, H.; Zarlis, M. Sawaluddin Analysis of Accuracy and Epoch on Back-propagation BFGS Quasi-Newton. In Proceedings of the International Conference on Information and Communication Technology (ICONICT), Singapore, 27–29 December 2017.

31. Han, T.; Lu, Y.; Zhu, S.-C.; Wu, Y.N. Alternating Back-Propagation for Generator Network. In Proceedings of the Thirty-First Aaai Conference on Artificial Intelligence, San Francisco, CA, USA, 4–9 February 2017; pp. 1976–1984.

32. Yan, Z. Research and Application on BP Neural Network Algorithm. In Proceedings of the 2015 International Industrial Informatics and Computer Engineering Conference, Shaanxi, China, 10–11 January 2015; pp. 1444–1447.

33. Upadhyay, A.; Singh, M.; Yadav, V.K. Improvised number identification using SVM and random forest classifiers. *J. Inf. Optim. Sci.* **2020**, *41*, 387–394. [CrossRef]

34. Mo, C.; Cui, H.; Cheng, X.; Yao, H. Cross-Scale Registration Method Based on Fractal Dimension Characterization. *Acta Opt. Sin.* **2018**, *38*, 1215001. [CrossRef]

35. Mittlböck, M.; Heinzl, H. A note on R2 measures for Poisson and logistic regression models when both models are applicable. *J. Clin. Epidemiol.* **2001**, *54*, 99–103. [CrossRef]

36. Patnana, A.K.; Vanga, N.R.V.; Chandrabhatla, S.K.; Vabbalareddy, R. Dental Age Estimation Using Percentile Curves and Regression Analysis Methods–A Test of Accuracy and Reliability. *J. Clin. Diagn. Res.* **2018**, *12*, ZC1–ZC4. [CrossRef]

37. Qi, J.; Du, J.; Siniscalchi, S.M.; Ma, X.; Lee, C.-H. On Mean Absolute Error for Deep Neural Network Based Vector-to-Vector Regression. *IEEE Signal Process. Lett.* **2020**, *27*, 1485–1489. [CrossRef]

38. Shinozaki, T.; Watanabe, S. Structure Discovery of Deep Neural Network Based on Evolutionary Algorithms. In Proceedings of the IEEE International Conference on Acoustics, Speech, And Signal Processing (ICASSP), Queensland, Australia, 19–24 April 2015; pp. 4979–4983.

39. Panchagnula, K.K.; Jasti, N.V.K.; Panchagnula, J.S. Prediction of drilling induced delamination and circularity deviation in GFRP nanocomposites using deep neural network. *Materials* **2022**, in press. [CrossRef]

40. Beniaguev, D.; Segev, I.; London, M. Single cortical neurons as deep artificial neural networks. *Neuron* **2021**, *109*, 2727–2739.e2723. [CrossRef] [PubMed]

41. Wang, G.; Wu, J.; Yin, S.; Yu, L.; Wang, J. Comparison between BP Neural Network and Multiple Linear Regression Method. *Inf. Comput. Appl.* **2010**, *6377*, 365–370.

42. Afanador, N.L.; Tran, T.N.; Buydens, L.M.C. Use of the bootstrap and permutation methods for a more robust variable importance in the projection metric for partial least squares regression. *Anal. Chim. Acta* **2013**, *768*, 49–56. [CrossRef]

43. Rodriguez-Perez, R.; Bajorath, J. Interpretation of machine learning models using shapley values: Application to compound potency and multi-target activity predictions. *J. Comput.-Aided Mol. Des.* **2020**, *34*, 1013–1026. [CrossRef]

44. Rodríguez-Pérez, R.; Bajorath, J. Interpretation of Compound Activity Predictions from Complex Machine Learning Models Using Local Approximations and Shapley Values. *J. Med. Chem.* **2020**, *63*, 8761–8777. [CrossRef]

45. Peng, J.; Zou, K.; Zhou, M.; Teng, Y.; Zhu, X.; Zhang, F.; Xu, J. An Explainable Artificial Intelligence Framework for the Deterioration Risk Prediction of Hepatitis Patients. *J. Med. Syst.* **2021**, *45*, 61. [CrossRef]

46. Wang, F.; Wang, Y.; Zhang, K.; Hu, M.; Weng, Q.; Zhang, H. Spatial heterogeneity modeling of water quality based on random forest regression and model interpretation. *Environ. Res.* **2021**, *202*, 111660. [CrossRef] [PubMed]

47. Chen, Y.; Cheng, A.; Zhang, C.; Chen, S.; Ren, Z. Rapid mechanical evaluation of the engine hood based on machine learning. *J. Braz. Soc. Mech. Sci. Eng.* **2021**, *43*, 345. [CrossRef]

48. Qi, C.; Xu, X.; Chen, Q. Hydration reactivity difference between dicalcium silicate and tricalcium silicate revealed from structural and Bader charge analysis. *Int. J. Miner. Metall. Mater.* **2022**, *29*, 335–344. [CrossRef]