energies

Article

Calorific Value Forecasting of Coal Gangue with Hybrid Kernel Function–Support Vector Regression and Genetic Algorithm

Xiangbing Gao [1], Bo Jia [2], Gen Li [3] and Xiaojing Ma [4,*]

[1] Xinjiang Xinneng Group Company Limited, Urumqi Electric Power Construction and Commissioning Institute, Urumqi 830000, China
[2] State Grid XinJiang Company Limited Electric Power Research Institute, Urumqi 830000, China
[3] School of Electrical Power Engineering, South China University of Technology, Guangzhou 510641, China
[4] School of Electrical Engineering, Xinjiang University, Urumqi 830017, China
* Correspondence: maxiaojing1983@xju.edu.cn

Abstract: The calorific value of coal gangue is a critical index for coal waste recycling and the energy industry. To establish an accurate and efficient calorific value forecasting model, a method based on hybrid kernel function–support vector regression and genetic algorithms is presented in this paper. Firstly, key features of coal gangue gathered from major coal mines are measured and used to build a sample set. Then, the forecasting performance of single kernel function-based models is established, and linear kernel and Gaussian kernel functions are chosen according to forecasting results. Next, a hybrid kernel combined with the two kernel functions mentioned above is used to establish a calorific value forecasting model. In addition, a genetic algorithm is introduced to optimize critical parameters of SVR and the adjustable weight. Finally, the forecasting model based on hybrid kernel function–support vector regression and genetic algorithms is built to predict the calorific value of new coal gangue samples. The experimental results indicate that the hybrid kernel function is more suitable for forecasting the calorific value of coal gangue than that of a single kernel function. Moreover, the forecasting performance of the proposed method is better than other conventional forecasting methods.

Keywords: coal gangue; calorific value forecasting; support vector regression; hybrid kernel function; genetic algorithm

Citation: Gao, X.; Jia, B.; Li, G.; Ma, X. Calorific Value Forecasting of Coal Gangue with Hybrid Kernel Function–Support Vector Regression and Genetic Algorithm. *Energies* **2022**, *15*, 6718. https://doi.org/10.3390/en15186718

Academic Editor: Marco Marengo

Received: 16 August 2022
Accepted: 5 September 2022
Published: 14 September 2022

1. Introduction

The energy structure characteristics of "rich coal, poor oil and less gas" in China make coal resources the main energy resource [1]. In fact, coal is acknowledged worldwide as one of the most important resources for human beings and chemical raw materials for industry, economy, medicine and daily life [2]. In coal production, a considerable amount of coal gangue is inevitably mixed into the raw coal, which generally accounts for 15~20% of the raw coal output. Coal gangue is a kind of solid waste rich in sulfur and a large number of heavy metals, including arsenic, cadmium, chromium, copper, tribute, etc. It is mainly composed of Al_2O_3 and SiO_2 [3]. Coal gangue has the characteristics of high density and low combustion value. As a result, the mixed combustion of coal gangue and coal will reduce its combustion efficiency and increase the emission of exhaust gas. Hence, it is necessary and crucial to explore scientific and proper ways to make good use of coal gangue [4–11].

Nowadays, coal gangue is no longer considered a useless byproduct due to the development of carbon-based waste recycling technologies and the energy industry. Coal gangue can be used in many fields and in various ways, such as building materials, functional filler, power generation, etc. The possible application fields of coal gangue are generally determined by its calorific value. Coal gangue with a high calorific value, which is set between 6270 and 12550 kJ/kg, is used to generate electrical power with a

fluidized-bed roaster or recirculating fluidized-bed boiler. Coal gangue with medium calorific value (2090~6270 kJ/kg) and low calorific value (less than 2090 kJ/kg) can be used for making brick and cement and for land rehabilitation and functional fillers, respectively. Therefore, it is necessary and important to identify the calorific value of coal gangue before determining which fields it should be used in [12–14].

Generally, the calorific value of coal or gangue is measured by following the Standard GBT 213-2008 in China, which is expensive, complicated and low efficiency [15]. Hence, it is necessary to establish a reliable, convenient and efficient approach to provide the calorific value of coal gangue. According to the investigation of published literature, it was found that previous research on coal gangue mainly focused on the identification of coal and the separation of coal from gangue. Furthermore, there is rarely literature related to how to obtain the calorific value of coal gangue conveniently and rapidly. Thus, the research status on coal gangue in China has not kept up with carbon peaking and carbon neutrality goals [16,17].

Therefore, this paper proposes a novel calorific value forecasting model combining support vector regression with a genetic algorithm. Considering the nonlinear mapping relationship between features of coal gangue and calorific value, a hybrid kernel function is adopted to promote the exploration and exploitation of the forecasting model, while a genetic algorithm is adopted to tune critical parameters to obtain reliable and accurate results. This method can output the calorific value of coal gangue simply and correctly without implementing any measurements.

The remainder of the paper is organized as follows: methodologies employed in this paper are briefly described in Section 2. Section 3 presents the calorific value forecasting procedure. Section 4 presents the experimental results and comparative analysis. Finally, Section 5 summarizes the overall contributions and discusses the future direction of the research.

2. Methodology

2.1. Support Vector Regression

In order to further solve the problem of regression and function estimation, Vapnik proposed support vector regression (SVR) on the basis of SVM in 1997 [18]. Compared with the commonly used neural network methods, SVR has the characteristics of simple structure and easy implementation and has been applied to many different fields such as data classification, regression estimation, function approximation and pattern recognition [19]. Moreover, compared with other artificial intelligent methods, SVR is a better fit for small sample solutions and nonlinear and high dimensional pattern recognition. The key characteristic of SVR is that the number of support vectors, rather than input data dimensions, determines computational complexity, which is a benefit of reducing computational cost and preventing the "dimensional curse".

Given dataset D = $\{(x_i, y_i)\}$, ($x_i \in R^n$, $y_i \in R$, $i = 1, 2, \ldots n$), x and y are n-dimensional input variables and the corresponding output values, respectively. n denotes the number of samples. The regression estimation function is represented as follows:

$$f(x) = \langle \omega, \Phi(x) \rangle + b \tag{1}$$

where ω and b denote weight vector and offset vector, respectively; kernel function $\Phi(x)$ is able to map the linearly inseparable data in the lower dimensional space to the linearly separable data in the high dimensional space. *f(x)* is the forecasting value.

The parameters ω and b can be assessed by minimizing the structured risk function:

$$\begin{aligned} &\min \tfrac{1}{2}\|\omega\|^2 + \tfrac{C}{n}\sum_{i=1}^{n} \varepsilon(y_i - f(x_i)) \\ &s.t.\ \varepsilon(y_i - f(x_i)) = \begin{cases} 0 & |y_i - f(x)| \le \varepsilon \\ |y_i - f(x)| - \varepsilon & otherwise \end{cases} \end{aligned} \tag{2}$$

where C is the error penalty factor for balancing empirical risk and confidence degree. $\varepsilon(\cdot)$ stands for the ε-non-sensitive loss function, and ε denotes the ε-intensive loss parameter. The optimization problem can be transformed into the following forms shown below by introducing slack variables ξ_i and ξ_i*.

$$\begin{aligned} &\min_{\omega,\xi,\xi^*} \tfrac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) \\ &s.t.\begin{cases} y_i - \langle \omega, \Phi(x)\rangle - b \le \varepsilon + \xi_i \\ -y_i + \langle \omega, \Phi(x)\rangle + b \le \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \ge 0 \end{cases} \end{aligned} \tag{3}$$

A Lagrange function is introduced to tackle the quadratic programming problem above, and the dual optimization problem is shown below:

$$\begin{aligned} &\min \tfrac{1}{2}\sum_{i,j=1}^{n}(a_i^* - a_i)\left(a_j^* - a_j\right)K(x_i \cdot x_j) + \varepsilon\sum_{i=1}^{n}(a_i^* + a_i) - \sum_{i=1}^{n} y_i(a_i^* - a_i) \\ &s.t.\begin{cases} \sum_{i=1}^{n}(a_i^* - a_i) = 0 \qquad i = 1, 2, \cdots n \\ 0 \le a_i^*, a_i \le C \end{cases} \end{aligned} \tag{4}$$

where a_i, a_i^* represents the Lagrangian multipliers, $K(x_i, x_j)$ is the preset kernel function. Then, the support vector regression is described as follows:

$$f(x) = \sum_{i=1}^{n}(a_i - a_i^*)K(x_i, x) + b \tag{5}$$

2.2. Hybrid-Kernel Function

In general, SVR adopts kernel functions to project a low-dimensional sample set to a higher-dimensional feature space and turn nonlinear problems into linear problems. Since different kernel functions possess different mapping capabilities and supply different prediction accuracies, great efforts have been made to select the proper kernel function for SVR [20–22]. Four popular kernel functions are listed below:

(1) Linear kernel function

$$k(x_i, x) = x_i \cdot x \tag{6}$$

(2) Polynomial kernel function

$$k(x_i, x) = (x_i \cdot x + 1)^d \tag{7}$$

(3) Gaussian kernel function

$$k(x_i, x) = \exp(-\gamma\|x_i - x\|)^2 \tag{8}$$

(4) Sigmoid kernel function

$$k(x_i, x) = \tanh(\gamma x_i \cdot x + \theta) \tag{9}$$

Usually, those kernel functions aforementioned can be separated into two classes: local kernel and global kernel. Global kernel function, such as linear kernel function and polynomial kernel function, emphasizes the influence of the data as a whole on the kernel function and has better interpolation ability. Hence, global kernel functions have strong generalization performance but weak learning ability. In contrast, the local kernel function, including a Gaussian kernel function and Sigmoid kernel function, emphasize the influence of the data near the key point on the kernel function, which result in the local kernel function possessing a desirable learning ability but mediocre generalization performance.

Figures 1 and 2 display the data distribution characteristics of a polynomial kernel and Gaussian kernel, respectively.

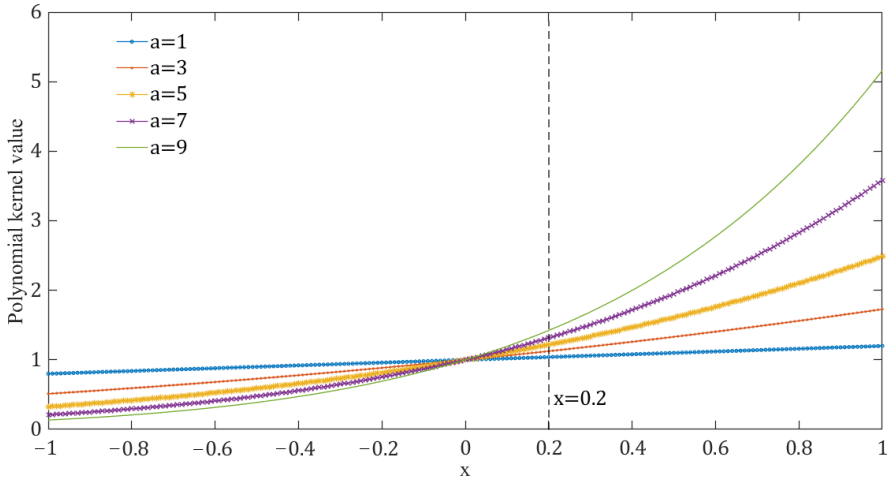


Figure 1. Data distribution of polynomial kernel (test point x = 0.2).

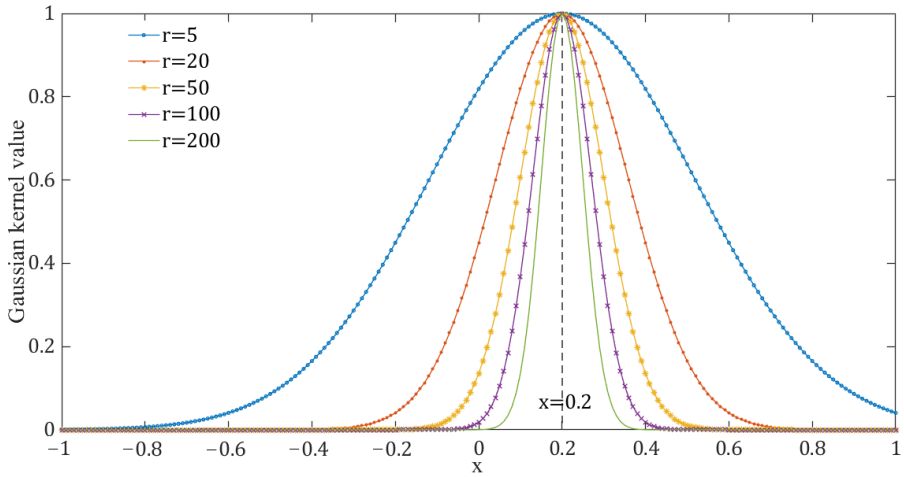


Figure 2. Data distribution of Gaussian kernel (test point x = 0.2).

When it comes to data samples that contain heterogeneous information, uneven distribution and irregularity, the conventional SVR models based on a single kernel function are not necessarily suitable for prediction. Therefore, by taking advantage of the merits and eliminating the potential disadvantages of all aforementioned kernel functions, a hybrid kernel function (HKF) of a combined local and global kernel is proposed to improve the interpretability and generalization performance. According to Mercer's conditions, the nonnegative linear combination of Mercer nuclei is still a Mercer nucleus. Thus, a hybrid kernel function, which consists of two kernel functions from different categories, is merged by means of viable weights to obtain a strong learning ability and generalization ability [23]. The hybrid kernel function combined with the Gaussian kernel and polynomial kernel is set as follows:

$$k_{hybrid}(x_i, x) = \mu \cdot (x_i \cdot x + 1)^{\alpha} + (1 - \mu) \cdot \exp(-\gamma \|x_i - x\|^2) \quad (10)$$

where α and γ present the power exponent and kernel width. The adjustable weight for the local kernel function μ is set between 0 and 1. Figure 3 shows the data distribution of the hybrid kernel with a polynomial kernel and Gaussian kernel, and the test point in the figure is x = 0.2, α = 1 and γ = 20. It is found that the hybrid kernel function can have

the advantages of both the local kernel and global kernel and can improve learning and generalization ability.

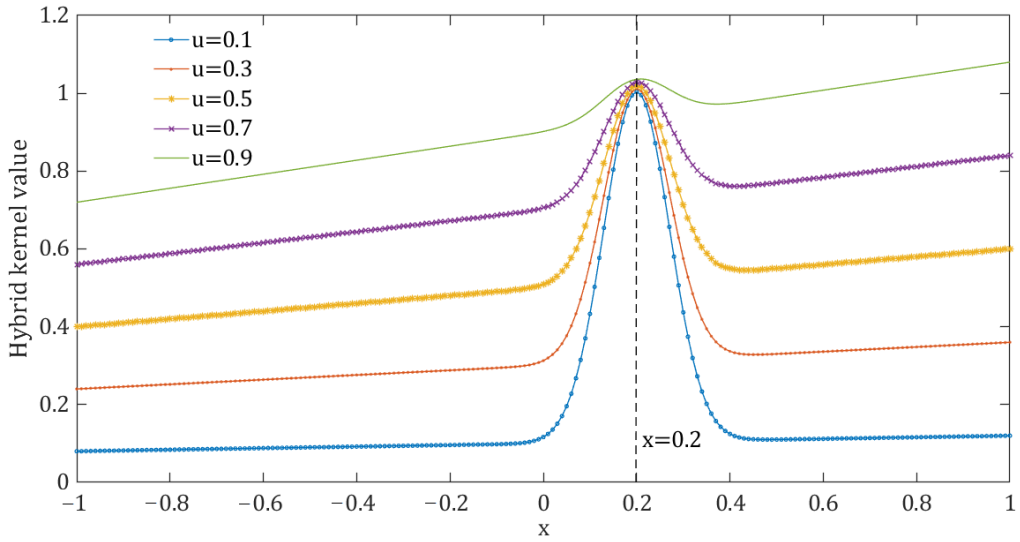


Figure 3. Data distribution of a hybrid kernel function (Gaussian kernel and polynomial kernel; test point x = 0.2).

2.3. Genetic Algorithm

Genetic algorithm (GA), as a global optimization algorithm with high robustness and wide applicability, was first proposed by John Holland in the 1970s [24]. GA is inspired by Darwin's principle of the "survival of the fittest" and the process of natural selection. GA has been adopted in various fields for optimization and search problems due to ease of convergence, high computational efficiency and a better global view of the search space [25].

The GA usually starts with a randomly initiated population amount that is composed of several features. Each set of features is called a chromosome and is considered a candidate solution to a specific problem. Moreover, the solutions are estimated by a fitness function, which is established based on the problem. Bio-inspired operations, including selection, crossover and mutation, are implemented to optimize all features concerned and output the optimal solution. Selection is a process that conducts searches in a predefined area with a higher possibility of finding optimum answers. A chromosome with a high fitness value is selected to reproduce offspring for the next generation. The crossover operator exchanges genes of two selected chromosomes and produces new chromosomes that inherit partial parental genetic information. The popular methods used in this operation are single-point, two-point, multi-point, or uniform way. The mutation process provides the feature of randomness and the ability to avoid becoming stuck in the local minimum. This operator operates on a single sequence and changes the feature of a potential solution with a tiny probability named mutation rate. Those genetic operators will be implemented until the stopping criterion is met [26]. The common optimization procedure of GA is shown in Figure 4.

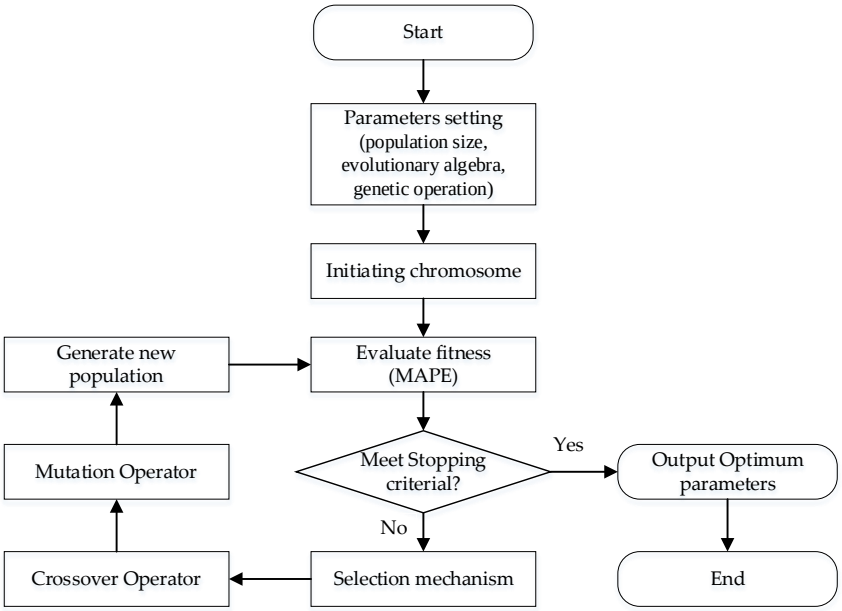


Figure 4. Diagram of the optimization procedure of GA.

3. Procedure for Forecasting Using Proposed Regression

To establish an efficient and accurate forecasting model, it is necessary and important to measure basic characteristics of coal gangue related closely to calorific value. Afterward, a dataset composed of those raw measuring data will be used to train and test the forecasting model. In order to establish a reliable and accurate forecasting model, two steps are needed: (1) data preparing and preprocessing, and (2) model training and verification of the forecasting model.

3.1. Data Preparing and Preprocessing

There is great diversity in characteristics of coal gangue from different coal mines and different regions. Even for coal gangue from an identical coal mine but different coal seam may produce large differences in fundamental characteristics. Therefore, it is impossible to gather all kinds of coal gangues and measure basic characteristics. In this paper, more than 1000 pieces of coal gangue samples collected from several major coal mines in China are investigated, and the fundamental characteristics obtained are applied to build forecasting models.

Since basic characteristics of coal gangue, such as air drying base moisture (M_{ad}), air drying base ash (A_{ad}), air drying base volatiles (V_{ad}) and air drying base fixed carbon (FC_{ad}), are recognized as crucial features of the air drying base bomb calorific value ($Q_{b,ad}$), measurement experiments were carried out to gain those parameters. The experimental instruction strictly followed Standard GBT 213-2008 and Standard GBT 212-2008 to guide the measuring process. The typical tools used in this study include balance, drier, Muffle furnace and calorimeter, which are shown below in Figure 5. Among them, 'balance' (Figure 5a) was used to weigh 1 g of coal gangue for the measuring experiment. A muffle furnace (Figure 5b) can be used to measure air drying base ash, air drying base volatiles and air drying base fixed carbon. The drier (Figure 5c) was applied to obtain air drying base moisture. The calorimeter (Figure 5d) was used to measure the calorific value of coal gangue.

(**a**) Balance

(**b**) Muffle furnace

(**c**) Drier

(**d**) Calorimeter

Figure 5. Typical tools used in experiments.

Considering the repeatability and tiny difference among obtained samples, 750 measured sample results out of over one thousand results composed of M_{ad}, A_{ad}, V_{ad}, FC_{ad} and $Q_{b,ad}$ were used to establish the coal gangue dataset. The partially measured samples gained are shown in Table 1 below.

Table 1. Partially measured samples of coal gangue.

Sample No.	M_{ad}/%	A_{ad}/%	V_{ad}/%	FC_{ad}/%	$Q_{b,ad}$/(MJ/kg)
1	3.69	63.46	17.66	15.19	7.68
2	1.08	66.87	16.02	16.03	7.56
3	4.52	64.14	17.2	14.14	6.56
4	2.98	68.61	15.82	12.59	6.11
5	1.37	65.46	17.99	15.18	8.33
6	0.98	74.88	14.72	9.42	4.64
7	1.27	66.06	17.11	15.56	7.43
8	2.72	65.45	17.04	14.79	7.48
9	0.85	70.38	18.47	10.3	5.31
10	3.24	63.69	17.68	15.39	7.97
11	2.09	70.09	15.56	12.26	6.15
12	1.84	72.24	14.88	11.04	5.44
13	3	63.36	17.84	15.8	7.67
14	4.42	67.96	15.36	12.26	5.66
15	1.18	60.63	19.36	18.83	10.14
16	1.11	57.92	20.16	20.81	11.05
17	2.06	73.44	14.22	10.28	4.79
18	1.68	73.5	14.55	10.27	4.73
19	1.8	71.33	15.54	11.33	5.69
20	3.3	64.5	16.79	15.41	7.72
... ...	... ...	... ...	... ...	... ...	... ...
731	1.83	63.72	17.92	16.53	8.7
732	4.2	62.9	16.74	16.16	8.01
734	2.58	62.1	18.25	17.07	9.21
735	2.36	73.98	13.96	9.7	4.86

Table 1. *Cont.*

Sample No.	M_{ad}/%	A_{ad}/%	V_{ad}/%	FC_{ad}/%	$Q_{b,ad}$/(MJ/kg)
736	1.26	70.26	15.6	12.88	6.06
737	2.27	66.2	16.68	14.85	7.48
738	3.18	62.81	17.61	16.4	8.36
739	1.12	68.72	16.39	13.77	6.5
740	3.46	66.7	16.34	13.5	6.36
741	15.06	55.82	20.18	8.94	5.53
742	6.66	63	16.64	13.7	6.15
743	5.26	65.42	14.98	14.34	5.95
744	2.43	72.7	14.34	10.53	5.39
745	8.8	53.26	17.47	20.47	9.23
746	7.46	52.04	20.42	20.08	8.28
747	4.36	63	16.3	16.34	6.98
748	5.52	63.79	14.99	15.7	6.84
749	0.8	58.56	21.47	19.17	10.25
750	7.46	52.04	20.42	20.08	8.28

It can be seen from Table 1 that there is a large range of changes among different features. Thus, these feature data are normalized in the range of [0,1] to reduce estimation error, promote calculation speed and improve generalization. The raw data were normalized with Equation (11)

$$x_n = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \tag{11}$$

where x_i and x_n represent the data before and after normalization, respectively; x_{max} and x_{min} are the maximum and minimum of the raw testing results.

3.2. Training and Testing of the Forecasting Model

The normalized coal gangue samples were separated into a training set and a testing set. Then, hybrid kernel functions and support vector regression were used to establish the calorific value prediction model. Then, GA was implemented in this paper to tune the critical parameters of SVR and kernels to promote forecasting accuracy and generalization ability. The details of establishing a calorific value forecasting model with GA are briefly described below:

- Initialization of GA and encoding parameters

Critical parameters of GA, including the size of the population, crossover possibility, mutation possibility and maximum iteration number, are predefined at the beginning. Then, the initial value of the chromosome, which is composed of penalty factor C, kernel bandwidth σ, intensive loss parameter ε, adjustable weight μ, etc., is set randomly. In addition, the real code method was adopted in this study to encode all parameters as it is ideal for solving complicated issues and convenient for deploying genetic operators to individuals. The ranges of all free parameters are shown in Table 2.

Table 2. Parameters of SVR and GA.

Algorithms	Parameter	Value
SVR	Mixing coefficient μ	(0,1)
	Penalty factor C	(0.01,100)
	RBF bandwidth σ	(0.01,100)
	Epsilon ε	(0.001,0.1)
GA	Population size	100
	Iterations	100
	Crossover probability	0.8
	Mutation probability	0.02

- Fitness function computation

The fitness function is generally applied to evaluate the performance of each individual. The 5-fold cross-validation method was adopted to assess the forecasting accuracy in this study. Mean absolute percentage error (MAPE) and squared correlation coefficient (r^2) were employed as the fitness function to estimate the quality of each individual and assess prediction performance. In general, the smaller value of the MAPE, the higher the forecasting accuracy, while the greater the value of r^2, the better the prediction performance. In addition, absolute percentage error (APE) is also used to estimate the quality of forecasting results. APE, MAPE and r^2 can be computed as follows:

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{f(x_i) - y_i}{y_i}\right| \tag{12}$$

$$r^2 = \frac{\left(n\sum_{i=1}^{n} f(x_i)y_i - \sum_{i=1}^{n} f(x_i)\sum_{i=1}^{n} y_i\right)^2}{\left(n\sum_{i=1}^{n} f(x_i)^2 - \left(\sum_{i=1}^{n} f(x_i)\right)^2\right) - \left(n\sum_{i=1}^{n} {y_i}^2 - \left(\sum_{i=1}^{n} y_i\right)^2\right)} \tag{13}$$

$$APE = \left|\frac{f(x_i) - y_i}{y_i}\right| \tag{14}$$

where x_i is the training data, and y_i and $f(x_i)$ represent the actual value and forecasting value provided by the established model. n stands for the size of the sample set.

- Genetic operation

A potential solution with better fitness has a higher probability of being chosen to reproduce offspring by genetic operators. Roulette wheel selection, arithmetical crossover and uniform mutation methods were adopted to generate new offspring. A flowchart of the optimization process is shown in Figure 6. At last, the optimal solution was applied to build the optimal forecasting model, and the index MAPE and r^2 were used to estimate the forecasting performance of the proposed model.

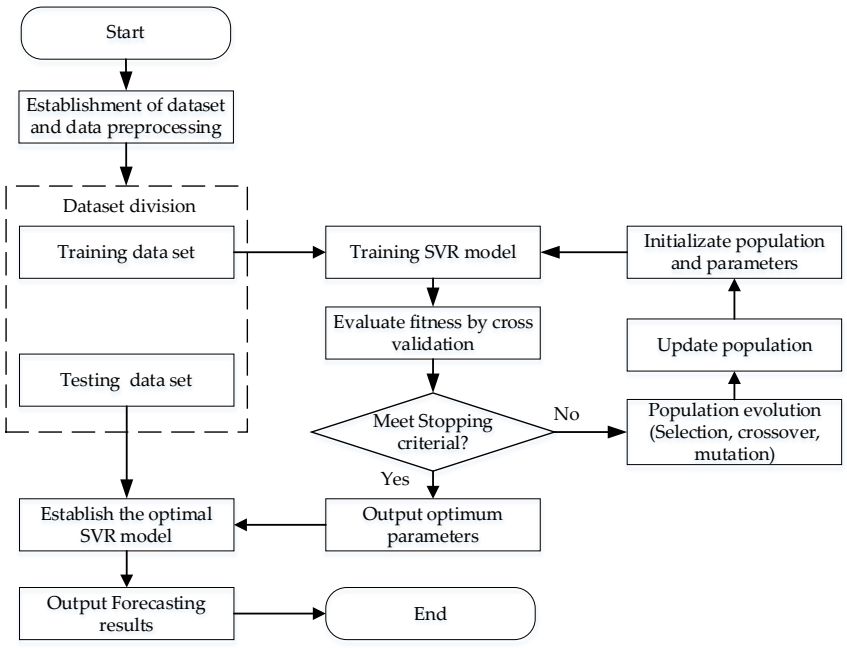


Figure 6. Flowchart of parameter optimization of HKF-SVR.

4. Discussion

The 750 coal gangue samples established in Section 3.1 were used to create a forecasting model. The data distributions of those features are displayed in Figure 7. It can be observed from Figure 7 that features of coal gangue vary within a large range, and hence, it is necessary to implement normalization to improve generalization performance and decrease computational error. After normalization, 80% of all samples were employed for the training set, and the rest of the samples (or 150 samples) were used to test the predicting performance of the established model. To be specific, features including M_{ad}, A_{ad}, V_{ad} and FC_{ad} were set as inputs, and the air drying base bomb calorific value ($Q_{b,ad}$) was employed as the output. Then, the forecasting performances of SVR models established on the single kernel function were compared to obtain proper kernel functions to establish the forecasting model with hybrid kernel functions. It should be noted that the proposed approaches were tested experimentally in the MATLAB (R2016) environment with the help of the LIBSVM toolbox.

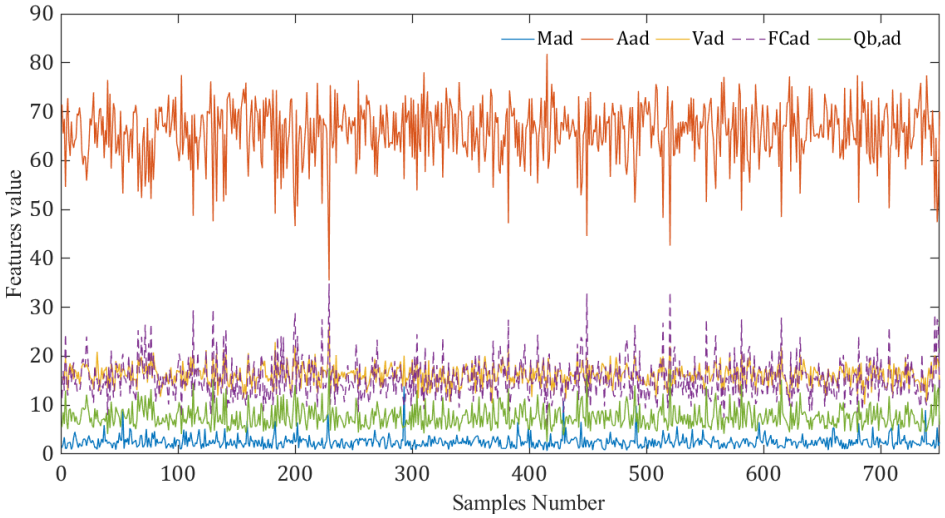


Figure 7. Data distributions of all features.

All local kernel functions and global kernel functions have pros and cons. Therefore, there is no obligation to use specific types of kernel functions to establish forecasting models. Firstly, the forecasting performance of each kernel function was tested and compared. The default value for each kernel was adopted in this experiment. The experimental results of the training set are shown in Table 3.

Table 3. Comparison of forecasting performance on the training set.

Kernel Function Type	APE_{max}	APE_{min}	$MAPE_{avg}$	r^2
Linear	0.3163	0.0000	0.0854	0.9635
Polynomial	0.9566	0.0000	0.1160	0.9290
Gaussian	0.6035	0.0000	0.0814	0.9629
Sigmoid	0.3550	0.0000	0.0892	0.9590

It can be found from Table 3 that between global kernels, the linear kernel function has better forecasting accuracy (MAPE) and depiction ability (r^2) than that of the polynomial kernel function. For local kernels, the Gaussian kernel function is able to offer better forecasting performance (MAPE and r^2) than that of the Sigmoid kernel function. In addition, the liner kernel and Gaussian kernel-based forecasting model can predict the calorific value of coal gangue better than that of other models based on other kernels, which can be seen in Table 4 and Figures 8 and 9.

Table 4. The average forecasting performance of the test set.

Kernel Function Type	APE_{max}	APE_{min}	$MAPE_{avg}$	r^2
Linear	0.4338	0.0001	0.0801	0.9485
Polynomial	0.4695	0.0001	0.1086	0.9320
Gaussian	0.4102	0.0000	0.0755	0.9484
Sigmoid	0.4348	0.0001	0.0824	0.9408

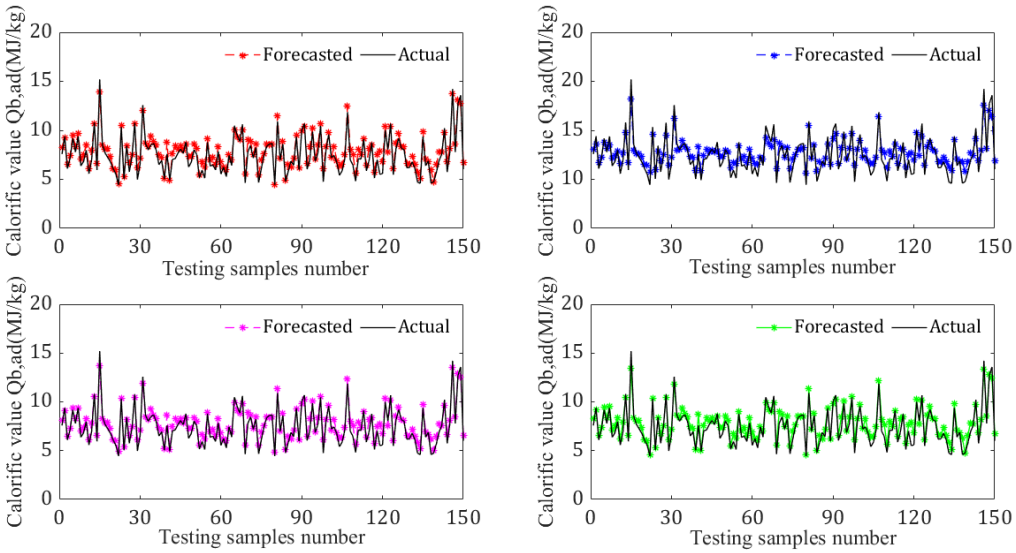
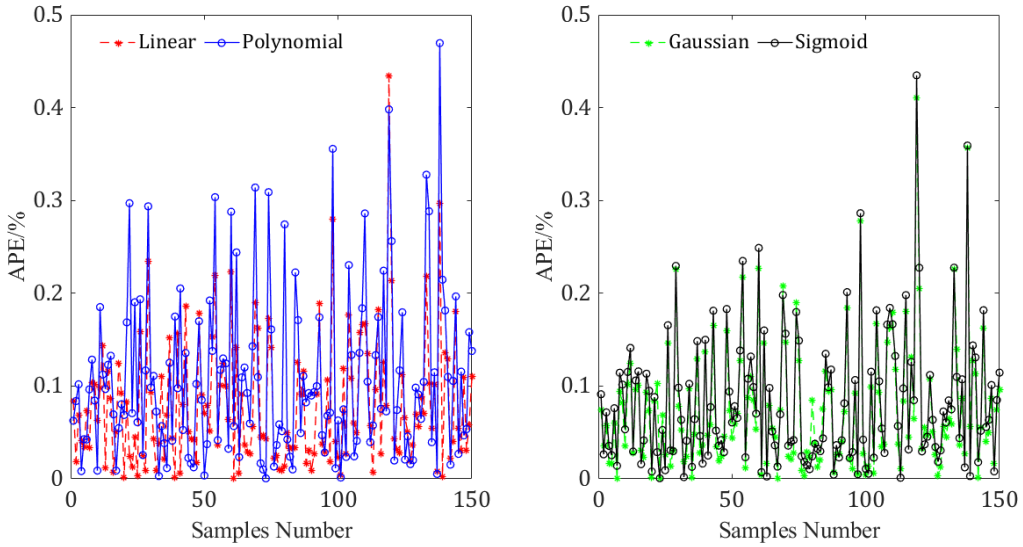


Figure 8. Comparison of forecasted calorific value ($Q_{b,ad}$) on different single kernel functions.



(**a**) APE of global kernel functions (**b**) APE of local kernel functions

Figure 9. APE of forecasted calorific value ($Q_{b,ad}$) based on a single kernel function.

According to the obtained results based on the training set and testing set, which are shown in Table 4 and Figures 8 and 9, it can be observed that all indexes of the linear kernel and Gaussian kernel used for assessing the forecasting performance are better than that of polynomial and Sigmoid kernel functions. Thus, the linear kernel function and Gaussian kernel function were selected to build the hybrid kernel function to further improve forecasting accuracy. Afterward, GA was introduced to tune the critical kernel parameters of ε-SVR, including penalty factor C, kernel bandwidth σ, intensive loss parameter ε

and adjustable weight μ. The changing range for each parameter is shown in Table 2. In addition, five-fold cross-validation was adopted to assess the fitness for selecting the optimum choice among the candidate solutions. In order to reduce the randomness of the final results, the numerical experiments of each model were conducted 20 times. The optimal parameters were set as C = 16.98, σ = 1.35, ε = 0.02 and μ = 0.79; the actual values and forecasted calorific values ($Q_{b,ad}$) for testing set are shown as Figure 10.

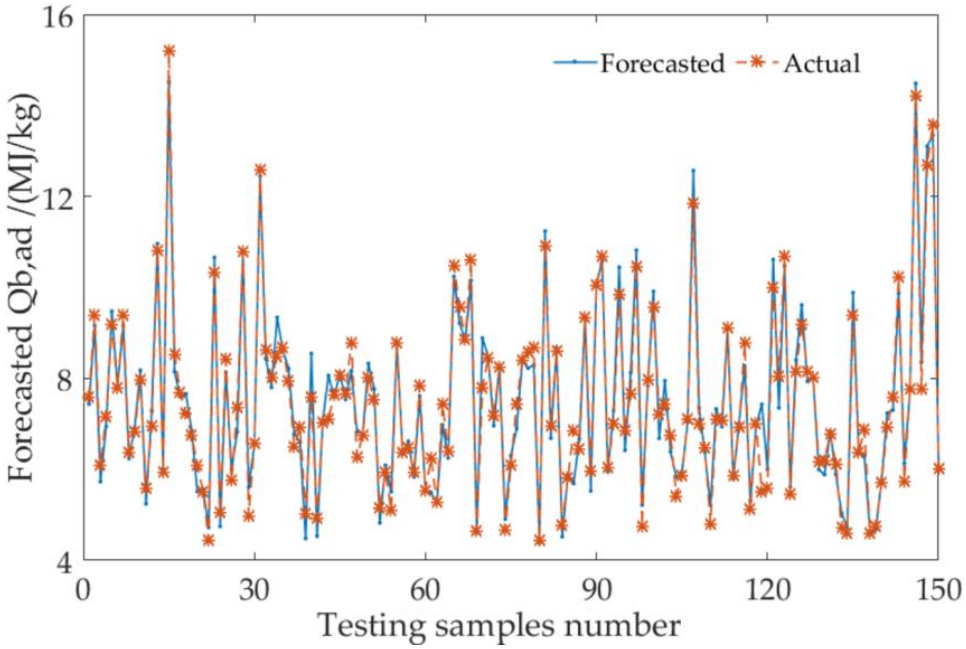


Figure 10. Comparison of actual values and forecasted calorific values based on the hybrid kernel function (where C = 16.98, σ = 1.35, ε = 0.02 and μ = 0.79).

According to Table 5 and Figures 10 and 11, it can be found that the forecasting model based on the hybrid kernel function has higher accuracy on both the training set and testing set than that of models based on a single kernel function. To be specific, the average MAPE decreased by 57.37% and 44.64% for the training set and testing set (compared with Gaussian kernel) when the hybrid kernel function was adopted, while the average squared correlation coefficient increased by 1.47% and 1.31% for the training set and testing set (compared with Gaussian kernel), respectively. It can be observed from Figure 11 that only 5.33% of all APE (only 8 samples) are higher than 0.1, and 30.67% of all APE (46 samples) are higher than 0.05, which suggests that the proposed method is capable of providing accurate forecasting values with minor errors. In summary, the support vector regression model with a hybrid kernel function and genetic algorithms is able to supply a desirable performance and impressive depiction ability on calorific value forecasting of coal gangue.

Table 5. The average forecasting performance of the HKF-SVR-GA model.

Data Set	APE_{max}	APE_{min}	$MAPE_{avg}$	r^2
Training set	0.3328	0.0000	0.0347 ± 0.0021	0.9771 ± 0.0021
Testing set	0.3695	0.0000	0.0418 ± 0.0007	0.9608 ± 0.0019

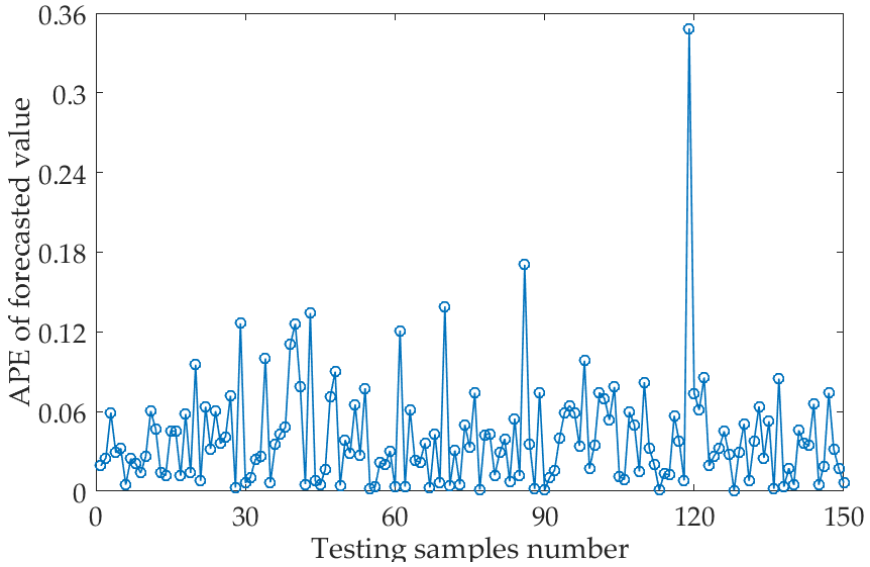


Figure 11. APE distribution of forecasted calorific value ($Q_{b,ad}$) by a hybrid kernel function.

Moreover, the forecasting performance of the HKF-SVR with other conventional methods (including generalized regression neural network and radial basis function neural network) are compared to further demonstrate the superiority of the proposed approach. The forecasted calorific values ($Q_{b,ad}$) by RBFNN, GRNN and HKF-SVR are displayed in Figure 12 and Table 6.

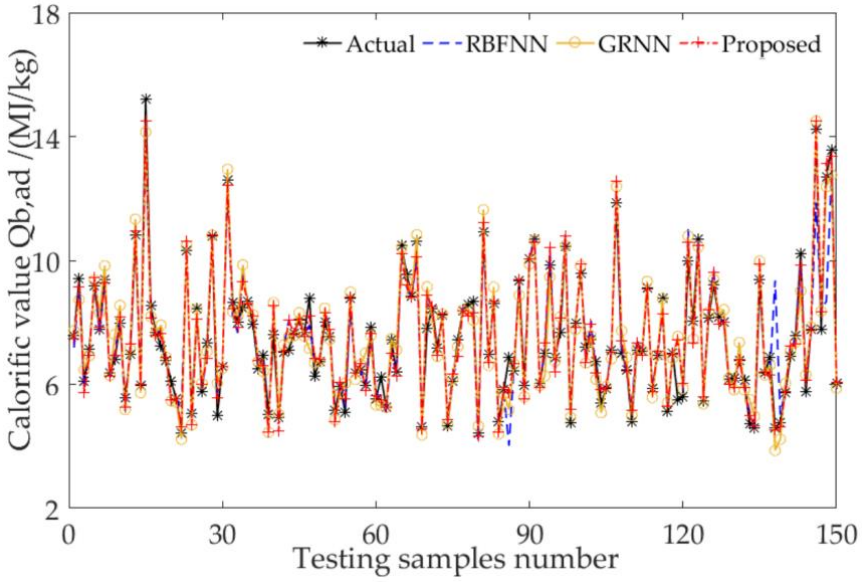


Figure 12. Comparison of forecasted results with different approaches.

Table 6. Forecasting performance comparison of different approaches.

Methods	Training Set		Testing Set	
	MAPE	r^2	**MAPE**	r^2
RBFNN	0.0319	0.9801	0.0542	0.8708
GRNN	0.0299	0.9817	0.0526	0.9386
Proposed	0.0054	0.9857	0.0408	0.9635

Compared with RBFNN and GRNN, it can be seen from Table 6 that the proposed approach has the lowest MAPE and the best r^2 on both the training set and testing set, which verifies the validity and superiority of the hybrid kernel function on forecasting. In summary, the calorific value forecasting model based on HKF-SVR is much more accurate and reliable than other conventional methods, which is beneficial for recycling coal gangue and reducing the environmental pollution.

5. Conclusions

In order to handle the issue that there is a lack of methods for forecasting the calorific value of coal gangue conveniently and accurately, a novel approach based on hybrid kernel function–support vector regression and genetic algorithms is presented in this paper. Firstly, the key characteristics of coal gangue gathered from major coal mines are measured and employed to build a sample set. Then, the forecasting performances of single kernel function-based models are compared, and the results suggest that the linear kernel and Gaussian kernel are capable of providing better accuracy and trend description ability. Next, a hybrid kernel linearly combining two kernel functions was used to create a calorific value forecasting model. Moreover, a genetic algorithm is introduced to optimize the critical parameters of SVR and the adjustable weight. The experimental results indicate that the hybrid kernel function-based model results in lower MAPE and a higher squared correlation coefficient, and the HKF-SVR model is more suitable for forecasting the calorific value of coal gangue than that of the single kernel function. Moreover, the forecasting performance of the presented method is superior to other conventional forecasting methods.

Author Contributions: This paper is the collaborative work of all authors. Conceptualization, X.G. and X.M.; formal analysis, B.J.; investigation, X.G. and G.L.; methodology, G.L. and X.M.; software and validation, B.J. and X.M.; writing—original draft preparation, X.M.; writing—reviewing and editing, G.L. and X.M.; supervision, X.G. and B.J.; project administration, G.L. and X.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Science and Technology Project of Xinjiang Xinneng Group Company Limited Urumqi Electric Power Construction and Commissioning Institute (grant No. SGXJXN00TSJS2200141).

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: Thanks to TuSongjiang Kari of Xinjiang University for his guidance on the procedure.

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

References

1. Feng, H.; Kai, B. Accurate Identification Strategy of Coal and Gangue Using Infrared Imaging Technology Combined with Convolutional Neural Network. *IEEE Access* **2022**, *10*, 8758–8766.
2. Xie, H.P.; Wu, L.X.; Zheng, D.Z. Energy Consumption and Coal Demand Forecasting in 2025 in China. *J. China Coal Soc.* **2019**, *44*, 1949–1960.
3. Han, L.; Ren, W.; Wang, B. Extraction of SiO2 and Al2O3 from coal gangue activated by supercritical water. *Fuel* **2019**, *253*, 1184–1192. [CrossRef]
4. Guo, Y.C.; Wang, X.Q.; Wang, S. Identification Method of Coal and Coal Gangue Based on Dielectric Characteristics. *IEEE Access* **2022**, *9*, 9845–9854. [CrossRef]
5. Liu, Q.; Li, J.G.; Li, Y.S. Recognition Methods for Coal and Coal Gangue Based on Deep Learning. *IEEE Access* **2021**, *9*, 77599–77610. [CrossRef]
6. Zhang, B.; Zhang, H.B. Coal Gangue Detection Method Based on Improved SSD Algorithm. In Proceedings of the 2021 International Conference on Intelligent Transportation, Big Data & Smart City, Frankfurt, Germany, 17–19 September 2021.
7. Cai, X.J.; Zhang, J.J.; Ning, Z.H. A Many-Objective Multistage Optimization-Based Fuzzy Decision-Making Model for Coal Production Prediction. *IEEE Trans. Fuzzy Syst.* **2021**, *20*, 3665–3675. [CrossRef]
8. Yan, P.C.; Sun, Q.S.; Yin, N.N. Detection of coal and gangue based on improved YOLOv5.1 which embedded scSE module. *Measurement* **2022**, *118*, 110530. [CrossRef]

9. Wu, Y.G.; Yu, X.Y.; Hu, S.Y. Experimental study of the effects of stacking modes on the spontaneous combustion of coal gangue. *Process Saf. Environ. Prot.* **2019**, *123*, 39–47. [CrossRef]
10. Wang, L.L.; Wu, X.C.; Gao, X. A method for in-situ measurement of calorific value of coal: A numerical study. *Thermochim. Acta* **2021**, *703*, 179011. [CrossRef]
11. Chelgain, S.C. Estimation of gross calorific value based on coal analysis using an explainable artificial intelligence. *Mach. Learn. Appl.* **2021**, *6*, 100116. [CrossRef]
12. Cao, X.G.; Fei, J.H.; Wang, P. Research on coal and gangue sorting method based on multi-manipulator cooperation. *Coal Sci. Technol.* **2019**, *47*, 7–12.
13. Ma, J.; Yu, Z.M.; Shu, S.H. Environmental hazard of coal gangue in mining area and its control measures. *Coal Eng.* **2015**, *47*, 70–73.
14. Liu, P.; Yan, F. The harm of coal gangue to environment and its comprehensive control. *China Min. Ind.* **2008**, *8*, 49–51.
15. *GBT-213*; Determination of Calorific Value of Coal. Standardization Administration of China: Beijing, China, 2008.
16. Su, L.L. *Research on Coal Gangue Recognition Method Based on Deep Learning*; Xi'an University of Science and Technology: Xi'an, China, 2018.
17. Bi, H.; Wang, C.; Lin, Q. Combustion behavior, kinetics, gas emission characteristics and artificial neural network modeling of coal gangue and biomass via TG-FTIR. *Energy* **2020**, *213*, 118790. [CrossRef]
18. Vapnik, V.; Golowic, S.; Smola, A. Support vector method for function approximation, regression estimation, and signal processing. *Proc. Adv. Neural Inf. Process Syst.* **2008**, *9*, 281–287.
19. Xie, X.J.; Luo, K.Y.; Wang, G.Q. A New L1 Multi-Kernel Learning Support Vector Regression Ensemble Algorithm with AdaBoost. *IEEE Access* **2022**, *10*, 20375–20384. [CrossRef]
20. Pillonetto, G.; Dinuzzo, F.; Chen, T. Kernel methods in system identification, machine learning and function estimation: A survey. *Automatica* **2014**, *50*, 657–682.
21. Xiao, J.; Wei, C.; Liu, Y. Speed estimation of traffic flow using multiple kernel support vector regression. *Phys. A Stat. Mech. Appl.* **2018**, *509*, 989–997. [CrossRef]
22. Kari, T.; Gao, W.S.; Tuluhong, A. Mixed Kernel Function Support Vector Regression with Genetic Algorithm for Forecasting Dissolved Gas Content in Power Transformer. *Energies* **2018**, *11*, 2437. [CrossRef]
23. Hiroki, K.; Toshihiro, Y. Mitigation of Kernel Memory Corruption Using Multiple Kernel Memory Mechanism. *IEEE Access* **2021**, *9*, 111651–111665.
24. Holland, J. *Adaptation in Natural and Artificial Systems*; University of Michigan Press: Ann Arbor, MI, USA, 1975; pp. 21–24.
25. Jafari, A.; Khalili, T.; Babaei, E. A Hybrid Optimization Technique Using Exchange Market and Genetic Algorithms. *IEEE Access* **2020**, *8*, 2417–2427. [CrossRef]
26. Haupt, R.L.; Haupt, S.E. *Practical Genetic Algorithms*, 2nd ed.; Wiley: Hoboken, NJ, USA, 2004.