MDPI

Article

Prediction Model of Borehole Spontaneous Combustion Based on Machine Learning and Its Application

Yun Qi [1,2,†], Kailong Xue [1,*,†], Wei Wang [1,*,†], Xinchao Cui [1] and Ran Liang [1]

[1] School of Coal Engineering, Shanxi Datong University, Datong 037000, China; qiyun_sx@sxdtdx.edu.cn (Y.Q.);220857002125@sxdtdx.edu.cn (X.C.); 210857002131@sxdtdx.edu.cn (R.L.)
[2] China Safety Science Journal Editorial Department, China Occupational Safety and Health Association, Beijing 100011, China
[*] Correspondence: sdtu-faker@sxdtdx.edu.cn (K.X.); wangwei@sxdtdx.edu.cn (W.W.)
[†] These authors contributed equally to this work.

Abstract: In order to quickly and accurately predict borehole spontaneous combustion danger and avoid borehole spontaneous combustion fires, a borehole spontaneous combustion prediction model combining the Hunger Games search optimization algorithm (HGS) and Random Forest (RF) algorithm was introduced. The number of trees and the minimum number of leaf nodes in RF were optimized by HGS. Based on the data obtained from the temperature rise experiment of spontaneous combustion characteristics in a Shandong mine laboratory, O_2, CO, C_2H_4, $CO/\Delta O_2$ and C_2H_4/C_2H_6 were selected as the input indexes for the prediction of borehole spontaneous combustion, and the spontaneous combustion temperature was selected as the output indexes to train the built model. The prediction performance and accuracy of the model were evaluated using four indexes: the mean absolute error (MAE), mean absolute percentage error (MAPE), root mean square error (RMSE) and coefficient of determination (R2). At the same time, the prediction results of the HGS-RF model were compared with those of the RF model, Sparrow search algorithm (SSA) optimization RF model, particle swarm optimization RF model (PSO) optimization RF model and quantum particle swarm optimization RF model (QPSO) optimization. The results showed that the MAE of the RF, SSA-RF, PSO-RF, QPSO-RF and HGS-RF model samples were 17.541, 15.7752, 12.5903, 6.8594 and 6.6921, respectively. MAPE was 13.81%, 10.9766%, 9.6802%, 4.5731% and 5.1536%, respectively. RMSE values were 21.5646, 15.2017, 17.0091, 11.9879 and 12.1691, respectively. The R2 values were 0.9043, 0.9315, 0.9266, 0.9668, and 0.9717, respectively. At the same time, the reliability of the HGS-RF model was supplemented by taking the test data of the Zhengjia1204 coal mining face as an example. Finally, the model was applied to the prediction of borehole spontaneous combustion in the Jinniu Coal Mine, Shanxi Province. The prediction results show that the HGS-RF model can predict the spontaneous combustion temperature of different boreholes quickly and accurately. The results show that the HGS-RF model is more universal and stable than other models, indicating that the HGS-RF model is more suitable for the prediction of borehole spontaneous combustion.

Keywords: gas extraction; borehole spontaneous combustion; Hunger Games search algorithm (HGS); Random Forest (RF); prediction model

1. Introduction

Natural combustion in gas extraction boreholes is a type of coal mine interior fire that is affected by multiple factors that severely limit the high production, efficiency and safety of the mine [1]. As the mining depth and intensity increase, the initial temperature of the coal seams increases, posing new challenges for the prevention and control of coal spontaneous combustion disasters. Especially for highly gas-prone coal seams, the risk of spontaneous combustion during drilling is growing. However, spontaneous combustion often occurs at a distance from the exposed surface of the coal, making it difficult to determine the location

of the fire source [2]. Once spontaneous combustion occurs in an extraction borehole, it can lead to the suspension or scrapping of the borehole, and in severe cases, it can cause disasters such as the explosion of an extraction pipe. Therefore, prevention and control of spontaneous combustion in extraction boreholes is an urgent and significant issue [3,4]. Determining the risk of spontaneous combustion from drilling is a fundamental basis for adopting fire prevention measures. How to adopt scientific and reasonable methods to improve the prediction accuracy is of great guidance for the control of spontaneous combustion in boreholes.

In recent years, scholars from home and abroad have conducted extensive experimental studies on the spontaneous combustion process of coal [5]. Among them, the indicator gas analysis method has been widely used due to its advantages such as low cost, elevated detectability and sensitivity, and strong regularity. It mainly uses methods such as on-site sampling and laboratory analysis to determine the generation of gas products corresponding to a certain coal type at different temperature stages, and then it determines the degree of coal seam danger based on the monitoring of tunnel gas in actual production and corresponding test results [6]. However, due to the complex nonlinear relationship between spontaneous combustion of coal and gas, how to interpret and deal with this nonlinear relationship through a scientific and reliable method is key to the solution. Many scholars have combined coal spontaneous combustion warning indicators with machine learning smart algorithms for coal spontaneous combustion prediction, which have consistently improved the accuracy of the prediction results [7,8]. Qi Yun et al. [9] used the ensemble evaluation method of the set-valued statistics Entropy to model the human decision-making process and treated multidimensional data resulting from spontaneous combustion of coal to avoid bias in quantitative evaluation. Wang Wei et al. [10] modified the G2 weighting method by improving the CRITIC method and combined it with the TOPSIS method to construct a risk assessment model for spontaneous combustion in drilling, effectively evaluating the natural hazards of drilling. Zhou Xu et al. [11] constructed a predictive model for the coal ignition temperature based on the PSO-XGBost algorithm and compared it with the RF model and the GBRT model, noting that the constructed model is better in accuracy and robustness. Zhai Xiaowei et al. [12] comprehensively considered the factors influencing the risk of coal spontaneous combustion and established a coal spontaneous combustion evaluation model based on Structural Equation Model (SEM), providing a reliable method for coal mining enterprises to prevent coal spontaneous combustion disasters. Zhao [13] adopted the idea of Ensemble learning to build the PCA Ada Boost coal spontaneous combustion prediction model, which not only reduces the difficulty of data acquisition but also improves the real-time prediction. Deng Jun et al. [14] and Zheng Xuezhao et al. [15] developed predictive models for the spontaneous combustion of coal based on the Random Forest algorithm. Through comparison and analysis with the SVM, BPNN neural network, PSO-SVM, PSO-BPNN, PSO-RF and other algorithmic models, they pointed out that the RF model is not only fast to operate and simple to optimize parameters but also not prone to overfitting. Jia Pengtao et al. [16] proposed a prediction model of coal Autoignition temperature based on the PSO algorithm to optimize the Simple Recurrent Units (SRU) in order to solve the problems of weak generalization and poor robustness of existing prediction models. By mining the non-linear relationship between the Autoignition temperature and the gas index properties, this model reduces the difficulty of parameter selection and structural complexity of the model and improves the efficiency and accuracy of coal spontaneous combustion prediction. Wen Tingxin et al. [17] used the Kernel principal component analysis method to extract the nonlinear characteristics of the characteristic indicators with elevated correlation, took the extracted principal components as the discriminant factors of Fisher Discriminative model and proposed a prediction model of coal spontaneous combustion based on KPCA Fisher discriminant analysis. Shukla et al. [18] used five machine learning tools to predict the spontaneous combustion sensitivity (WOP). The results showed that XGBoost algorithm had the highest prediction accuracy, and WOP could be used as the prediction parameter of spontaneous combustion risk. Abiodun I L

et al. [19] established a spontaneous combustion prediction model optimized by an artificial neural network (ANN) based on the SHO algorithm and applied the model to the actual production mine. The results showed that the prediction ability of the model was very good and the error was close to 0, and it is pointed out that volatile substances (VM) and oxygen (02) had the greatest influence on Wits-Ehac and FCC. Oxygen and nitrogen (N2) have the greatest influence on cross point temperature (XPT), which provides an effective method for spontaneous combustion prediction. Although the above studies have played a role in predicting spontaneous combustion disasters in mines, there are still several shortcomings in the following: The BPNN neural network model suffers from local overfitting and slow convergence rate; the SVM model is only suitable for small samples, and the model itself does not provide a method to compute the penalty factor and the kernel parameter, which has certain limitations; the PSO model is sensitive to parameter choices. Although the algorithm has a fast rate of convergence, it is prone to get stuck in a local optimum; The prediction accuracy of the RVM model is not very high, making it difficult to achieve accurate predictions of spontaneous combustion, and its practical value is low; RF models have limited processing power for low-dimensional data and strong stochasticity. From this, it can be seen that the current methods for predicting the spontaneous combustion of coal have many limitations and do not effectively meet the requirements of the spontaneous combustion prediction of coal. Currently, there is a lack of research on the factors that influence and prevent spontaneous combustion in mining boreholes, where spontaneous combustion disasters often occur. The coal seams in the Shandong mines are deeply buried and have high gas content, causing severe gas disasters. During the gas extraction process, some boreholes experienced natural combustion due to loose sealing of boreholes and air leaks, leading to the suspension of drilling and even the abandonment of boreholes.

In view of this, based on the background of the study of spontaneous combustion of gas drainage holes in a mine in Shandong province, the authors first used mathematical prediction models to study the risk of spontaneous combustion of drainage holes. The authors then proposed to combine the Hunger Games search (HGS) algorithm with the RF algorithm, which has strong optimization ability and a rapid rate of convergence, and introduce it into the prediction of the risk of spontaneous combustion of holes [20]. Using the HGS optimization algorithm to optimize the number of trees and the minimum number of leaves in the RF, we developed a HGS-RF based predictive model for drilling spontaneous combustion risk. Meanwhile, the RF, SSA-RF and PSO-RF models were compared and analyzed to verify the prediction accuracy of the HGS-RF model. The proposed model was applied to the prediction of spontaneous combustion from drilling in the Jinniu Coal Mine to further investigate the generality and stability of the HGS-RF model. In order to improve the accuracy of predicting spontaneous combustion in gas extraction boreholes and to lay the foundation for the adoption of scientifically sound measures to prevent and control spontaneous combustion in boreholes, the results can provide theoretical support for other mines to address the problem of spontaneous combustion in extraction boreholes.

2. Basic Principles

2.1. The Index Gas Is Coupled with the Spontaneous Combustion Temperature of Coal

The process of borehole spontaneous combustion is very complicated, and it involves the interaction and reaction between various gases. From the point of view of the oxidation and spontaneous combustion characteristics of coal, various gases will be released in the natural oxidation process of coal with the increase of temperature, and the production rate of these gases will show regular changes with the increase of coal temperature. Therefore, the spontaneous combustion process of coal can be comprehensively evaluated and predicted by the multi-component gas coupling temperature.

In the process of coal spontaneous ignition, CO will run through the whole time. Usually, above 50 °C, it can be measured, and the concentration is relatively high. Alkanes (ethane, propane) appear almost synchronously with CO throughout the whole process, but their concentration is lower than CO, and there are different development rules in

different coal types. Olefins appear later than CO and alkanes, ethylene can be measured at about 110 °C, which is a symbol gas in the accelerated oxidation stage of coal spontaneous ignition, and its initial concentration is slightly higher than that of alkyne gas. Alkynes appear in the high-temperature section, and there is a significant difference in temperature and time difference with the former, which is the product of coal spontaneous ignition into the intense oxidation stage (i.e., the combustion stage). Therefore, some gases are selected from these gases as indicator gases and accurately detected, and the signs and states of spontaneous ignition can be judged. The spontaneous combustion characteristics of different coal types may be different, so it is necessary to select the appropriate gas combination according to the actual situation of the mining area to evaluate. It is generally believed that the selected gas should have the following basic characteristics:

(1) Sensitivity: In the process of natural oxidation reaction of coal, there will inevitably be a certain determination index gas, and with the intensification of coal oxygen reaction and the rise of coal temperature, the change trend of the determination index shows monotonicity.

(2) Uniqueness: In the case that the coal has not undergone a natural oxidation reaction, there is no gas, and the gas will only be generated in the process of natural oxidation of coal, indicating that the gas has uniqueness.

(3) Regularity: When coal is undergoing natural oxidation reaction, a certain indicator gas appears, and all coal samples on the working face are generated during natural oxidation. However, the temperature point at which this gas is generated at the earliest does not change much, and there is a good corresponding relationship between the concentration or generation rate of this gas and coal temperature.

(4) Testability: When coal is a natural oxidation reaction, a gas generated can be measured by the existing detection instrument, or the amount of a gas generated can be measured by the existing detection instrument, which indicates that the gas has testability.

In summary, the temperature coupling of spontaneous combustion of coal can be realized by analyzing the content and change of various gases and combining with the temperature information of coal. This coupling relationship provides more comprehensive and accurate spontaneous combustion characteristics, which is helpful to predict and monitor the spontaneous combustion risk of coal, as well as take corresponding prevention and control measures of borehole spontaneous combustion.

2.2. Hunger Games Search Algorithm

The Hunger Games search algorithm is a novel global population class optimization algorithm proposed by Yang et al. [21] in 2021 to model the behavior of hungry animals seeking food. Compared with traditional algorithms, HGS has better robustness, strong optimization capabilities, fewer optimization parameters and significant advantages in terms of convergence accuracy, speed, and stability [22].

First, the initial hunger level of each individual in the population is set, and the hunger weight is calculated based on the hunger level. Individuals with stronger hunger levels are more active in searching for food. In the beginning of the search for food phase, each individual acts independently to search for food. As the foraging process continues, foraging behavior may shift from individual foraging to group cooperative foraging. If an individual in the cooperative group finds food in the current region, it informs other individuals in the current population that food can be found here, causing the cooperative foraging individuals to move continuously toward the optimal solution. In summary, throughout the algorithm, both cooperative foraging between groups and foraging by individuals are allowed. Hence, this strategy ensures to some extent that the population can explore all possible regions within the specified space, ensuring the diversity of the algorithm.

In a population, any individual has different hunger levels, so the hunger weights W_1 and W_2 are defined by the following equation:

$$W_1^i = \begin{cases} hungry(i)\frac{N}{SHungry}r_4, r_3 < l \\ 1, r_3 > l \end{cases} \tag{1}$$

$$W_2^i = (1 - \exp(-\left|hungry(i) - SHungry\right|)) \times r_2 \times 2 \tag{2}$$

In the equation, *N* represents the total number of populations, and *SHungry* represents the sum of individual hunger levels. Sum (*hungy*), r_1, r_2 and r_3 are all random numbers between [0, 1]. *hungy* (*i*) represents the hunger level of each individual and is calculated using Equation (3).

$$hungry(i) = \begin{cases} 0, AllFitness(i) == BF \\ hungry(i) + H, else \end{cases} \tag{3}$$

In the equation, *AllFitness* (*i*) represents the set of fitness values for all individuals *i*, and *H* is calculated as follows:

$$TH = \frac{F(i) - BF}{WF - BF} \times r_6 \times 2 \times (UB - LB) \tag{4}$$

$$H = \begin{cases} LH \times (1 + r), TH < LH \\ TH, TH \geq LH \end{cases} \tag{5}$$

In the formula, r_4 is a random number between [0, 1]; *BF* is the current optimal fitness value; *WF* is the current worst fitness value; *UB* and *LB* are the upper and lower bounds of the search area, respectively; *LH* is the minimum value that can be obtained by the set *H*; and the general value is 100.

During foraging, natural organisms typically adopt a cooperative approach, but some individuals may not participate in cooperative foraging alone. When an individual is in a state of starvation, they search randomly for food, and when they find enough food to stop hunger, they stop searching. Based on the above rules, a model of cooperative communication and foraging between individuals is developed as follows:

$$X(t+1) = \begin{cases} X(t)(1 + \text{rand } n(1)), r_1 < l \\ W_1 X_b + RW_2|X_b - X(t)|, r_1 > l, r_2 > E \\ W_1 X_b - RW_2|X_b - X(t)|, r_1 > l, r_2 < E \end{cases} \tag{6}$$

where rand *n* (1) is the normal distribution with mean of 0 and variance of 1; r_1 and r_2 are both random numbers within [0, 1]; W_1 and W_2 are weight values calculated based on the characteristics of the hungry individuals; *Xb* represents the global optimal solution; *X* (*t*) represents the position of the current individual; *l* is a self-set constant; *R* is a random number at [−*a*, *a*]; and the size of a is related to time *t*.

The calculation formulas for *H* and *E* are shown in Equations (7) and (8):

$$H = \begin{cases} LH \times (1 + r), TH < LH \\ TH, TH \geq LH \end{cases} \tag{7}$$

In the formula, rand is a random number within the [0, 1] interval, and *T* is the maximum iteration algebra.

$$E = \sec h(|F(i) - BF|)$$
$$\sec h(x) = \frac{2}{e^x + e^{-x}} \tag{8}$$

where $i \in (1, 2, 3, ..., n)$, *F* (*i*) is the fitness value of the ith individual, *BF* is the current optimal fitness value, and sech is a hyperbolic function.

2.3. Random Forest Algorithm

2.3.1. Decision Tree

Decision tree (DT) algorithms are able to summarize decision rules from a series of information with features and labels and present these rules as a tree structure to solve classification and regression problems. A decision tree is used for decision making, which proceeds as follows: Starting from the root node of the decision tree, search for the corresponding characteristic attribute in the item to be classified and select the corresponding branch according to the characteristic attribute. Then, expand from the branch of the decision tree to the leaf node, and finally take the class saved by the leaf node as the final decision result so as to establish a classification system based on the decision tree. Common decision tree algorithms include ID3, C4.5 and CART.

2.3.2. Bagging Though

The Bagging idea is an ensemble approach in the field of machine learning that combines multiple weak learning models to obtain a strong learning model. The principle is to randomly draw from the original sample set and randomly select a total of K rounds. Each round selects n training samples from the original sample set, resulting in K training sets. One model is trained using one training set at a time, and a total of K base models are obtained through K-round training sets. The K-basis model is used to make predictions on the test set, and the K-prediction results are summarized. The mean of the above model is calculated as the final result.

2.3.3. RF Algorithm

The Random Forest (RF) is a generalization of Bagging's idea on the basis of the decision tree algorithm and Ensemble learning algorithm [23,24]. Using the CRAT decision tree as the weak learner and building on the use of the decision tree, the following improvements are made to the decision tree: for a general decision tree, the best feature will be selected from the n sample features at a node as the left and right subtrees of the decision tree. In contrast, the Random Forest randomly selects some sample features from the nodes and computes the best feature to complete the split of the left and right subtrees of the decision tree. This can avoid certain overfitting features and eliminate any pruning of the decision tree, resulting in high accuracy and generalization performance of the overall model results. The RF structure is illustrated in Figure 1.

The construction steps of the RF are as follows:

1. The Bagging sampling method is used to extract k data subsets (S_i, i = 1, 2, …, k) from the original data set S, and in this k times of extraction, the data not extracted each time constitute k out of pocket data sets, and the extracted data sets are called in-pocket data sets.
2. Randomly select m * attributes from m features as a sub dataset, and then select the optimal feature from that subset for partitioning to construct a CART decision tree.
3. Each CART decision tree grows to its maximum degree without any pruning operation, and the value of m remains constant.
4. In total, k CART decision trees are generated for each of the k extractions, and each tree does not influence each other and exists independently.
5. The generated decision trees are integrated to form a Random Forest, and the average of the output values of all decision trees is taken as the final prediction value of the Random Forest.
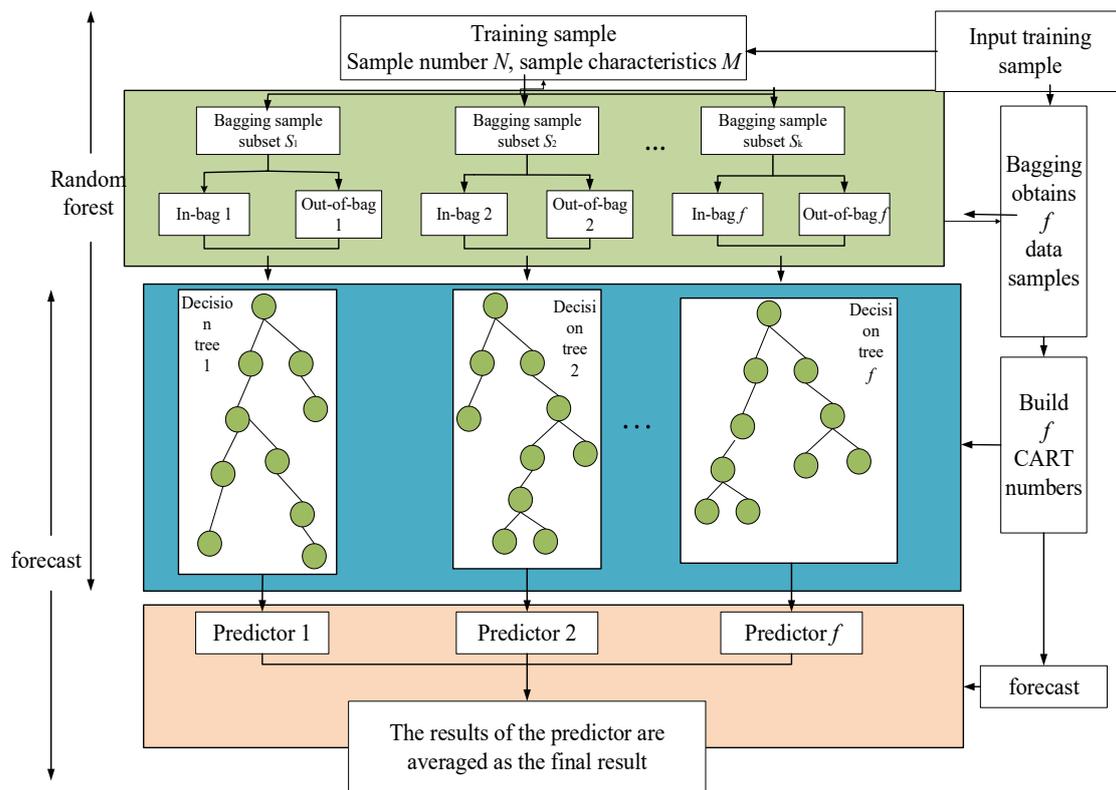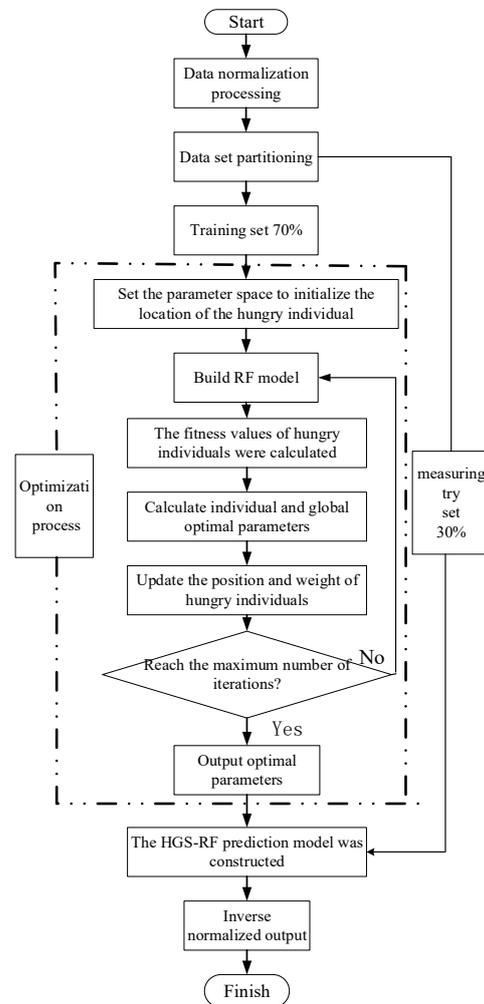
Figure 1. Random Forest training process.

3. Prediction Models Based on HGS-RF for Spontaneous Combustion in Boreholes

3.1. Construction of HGS-RF Model

The RF model has a strong performance in dealing with regression problems, but because the training accuracy is affected by the number of trees and the parameter setting of the smallest leaf node, the manual adjustment of its parameters is often subjective. In addition, the number of trees is too large to significantly improve the model performance, and the phenomenon of "underfitting" is prone to occur when the number of trees is too small. Therefore, HGS, with strong global search ability, fast convergence speed and good stability, was introduced to optimize its two key parameters so as to improve the model performance and thus establish a borehole spontaneous combustion prediction model based on HGS-RF. The build process is shown in Figure 2, The steps of construction are as follows:

(1) Initialize the number of individuals *N*, the maximum number of iterations Maxiter, the constant l, and the upper and lower bounds and dimension of the parameter space *D*.

(2) The location information of the hungry individual X_i is initialized, and the fitness value is computed based on the fitness function, where the fitness function of the HGS-RF prediction model is the mean squared error of the training set. The fitness value corresponding to the hungry individual with the smallest fitness value is chosen as the global optimum.

(3) According to Equation (1), update the location information and hunger characteristics of hungry individuals, calculate the fitness value of the updated hungry individuals and compare it with the extreme fitness value of the individual. Then, select a better result for iterative updating.

(4) The optimal value of the hungry individual is compared to the global optimum, and a smaller fitness value is chosen as the new global optimum.

(5)	Repeat steps (3) and (4) to determine if the maximum number of iterations Maxiter has been reached. If so, terminate the iteration and select the parameter corresponding to the global optimal value as the optimal parameter.

(6)	The optimal parameters are given to the Random Forest to construct the HGS-RF prediction model.

Figure 2. Flow chart of the HGS-RF model.

3.2. Performance Evaluation Metrics for Models

In order to accurately analyze the accuracy of the HGS-RF borehole spontaneous combustion risk prediction model, four indicators, namely the coefficient of determination (R-Square, R2), mean absolute percentage error (MAPE), root-mean-square deviation (RMSE) and mean absolute error (MAE), were introduced as the evaluation basis [25]. The formula for the calculation is as follows:

$$R^2 = 1 - \frac{\sum\limits_{i=1}^{n}(y_i - f_i)^2}{\sum\limits_{i=1}^{n}(y_i - \overline{y})^2} \tag{9}$$

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{f_i - y_i}{y_i}\right| \times 100\% \tag{10}$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(f_i - y_i)^2} \tag{11}$$

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|f_i - y_i| \times 100\% \tag{12}$$

where n is the number of samples; f_i is the predicted value, °C; y_i is the true value, and °C is the average of the true values, °C. The smaller the MAE, MAPE, and RMSE values are, and the closer the R^2 is to 1, the better the performance is.

3.3. Data Sources

In this paper, the coal samples from a mine in the Shandong Province were tested by a coal spontaneous ignition experiment and temperature-programmed experiment, and the single-indicator gas and compound-indicator gas were analyzed, respectively. Thus, the change law of the concentration of C_2H_4, C_2H_6 and other gases with the increase of temperature was obtained. The characteristics of the spontaneous combustion of coal were determined, and five prediction indexes of spontaneous combustion temperature were obtained [26], namely CO, O_2, CO/ΔO_2, C_2H_4 and C_2H_4/C_2H_6. A total of 337 experimental data of spontaneous combustion characteristics meeting the conditions were obtained for analysis, and the O_2 concentration, CO concentration, C_2H_4 concentration, CO/ΔO_2 ratio and C_2H_4/C_2H_6 ratio were selected as the input indexes and temperature the as output index. Some experimental sample data are shown in Table 1.

Table 1. Partial sample data.

O_2/%	CO/10^{-6}	C_2H_4/10^{-6}	CO/ΔO_2	C_2H_4/C_2H_6	°C
17.06	130	0	0.33	0	47.6
16.84	103	0	0.25	0	47.71
17.43	113	0	0.32	0	47.93
17.92	123	0	0.4	0	48.04
17.48	109	0	0.31	0	48.81
. . .	. . .	. . .	. . .	. . .	. . .
16	1597	0.46	3.19	0.04	114.93
17.72	667	0.23	2.03	0.02	115.28
14.71	1340	0.36	2.13	0.03	115.99
16.94	1582	0.57	3.9	0.02	116.34
19.03	1495	0.4	7.59	0.03	117.05
. . .	. . .	. . .	. . .	. . .	. . .
6.51	12986	685.74	8.96	0.11	405.76
3.52	13370	294.14	7.65	0.11	414.47
1.89	14248	490.07	7.46	0.12	418.83
1	14134	890.72	7.07	0.12	427.54
1.5	13429	291	6.89	0.11	431.9

3.4. Application of the HGS-RF Prediction Model

3.4.1. Determine Model Parameters

When applying the HGS-RF based prediction model for drilling spontaneous combustion hazards, in order to obtain more accurate prediction results, it is necessary to optimize the RF parameters, as shown in Table 2 [27]. Due to the small sample size and to ensure randomness in the sample selection, it is necessary to use the cobb error estimation method for error estimation. Therefore, oob_ sets the sour parameter to true. Based on previous experience, this criterion was set as the default parameter, which is the mean squared error (MSE); max_ set features to "auto"; max_ set the depth to 'none'; min_ samples_ sets the splitting parameter to 2. The n_ estimator and min_ samples_ leaf parameters were optimized by HGS.

Table 2. Required optimization parameters.

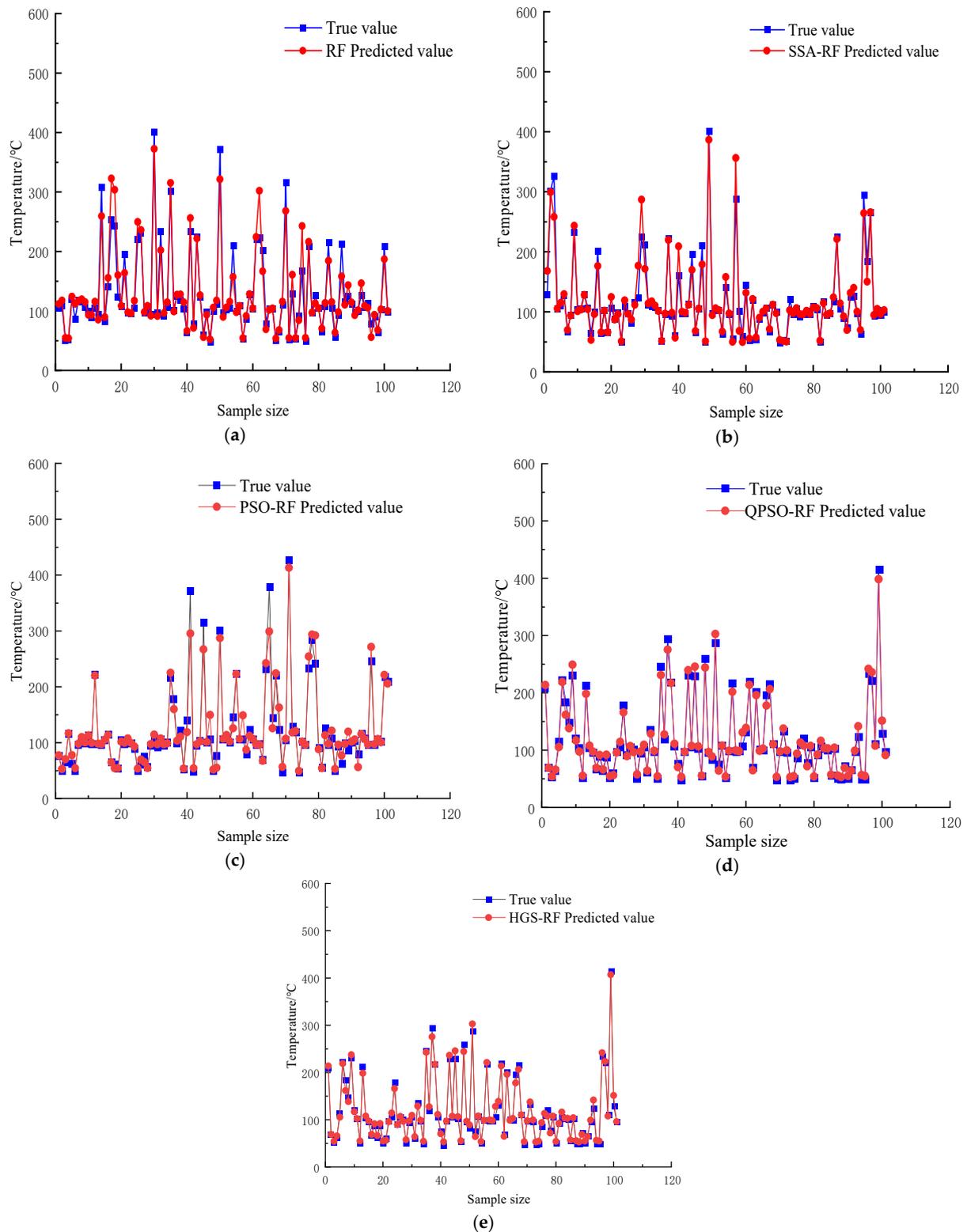| Parameter	Role
n_estimators	Number of decision trees
oob_sore	Whether to use external samples to assess model strengths and weaknesses
criterion	Classification criteria for nodes
max_features	Maximum number of features required to construct an optimal model of a decision tree
max_depth	Limit the maximum depth of the decision tree
min_samples_split	The minimum number of samples that can be divided into nodes is set as 2 in this paper.
min_samples_leaf	Minimum number of samples contained in a leaf node

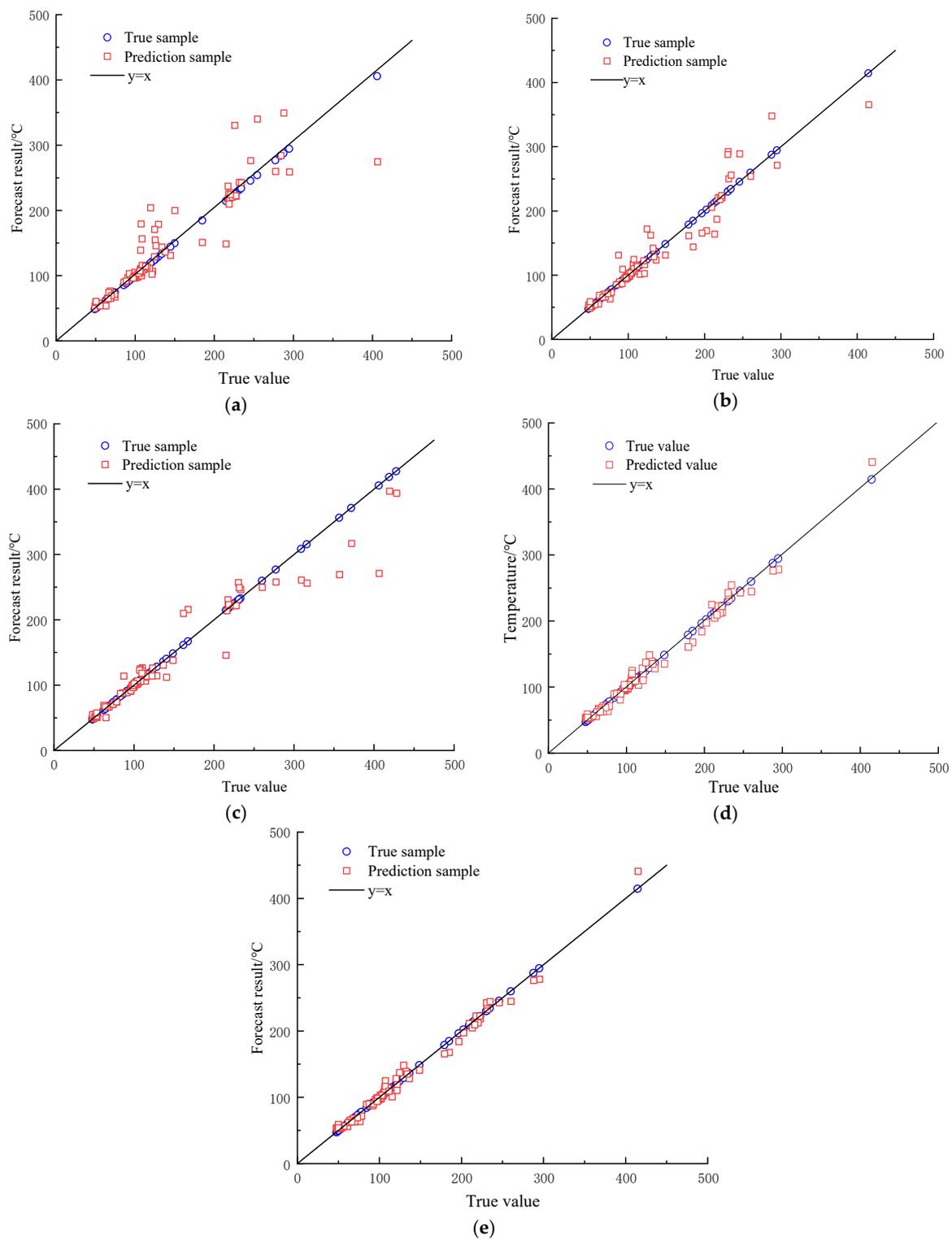3.4.2. Prediction Results and Comparative Analysis

To verify the accuracy of the predictions of the HGS-RF model, the RF, SSA-RF, PSO-RF and QPSO-RF models were introduced to predict the degree of spontaneous combustion in boreholes, and the predictions were compared and analyzed. The dimension of the parameter space D in HGS was 2, the population was 30, the maximum number of iterations was 120, $L = 0.03$, the lower bound of H was 100 and the range of values for the upper limit B_U and lower limit B_L of parameter space D was [10, 1] and [200, 50], respectively. The leaf parameters n in RF_ estimator and min_ samples_ were optimized by HGS with values of 100 and 2, respectively. The remaining parameters remained the same as in the previous text. The population size in the PSO was 60, the maximum number of iterations was 200, the learning factor was 1.5, the speed limit was maximum 1 and minimum −1, the inertia weight was 0.8, the population limit was maximum 5 and minimum −5; the number of sparrow populations in the SSA was 30, with a maximum number of iterations of 120; and the proportion of sparrow population discoverers was 20%. The number of the QPSO population was 30, the maximum number of iterations was 200, the parameter search range was [0.001, 100] and the shrinkage expansion coefficient α decreased linearly from 1.0 to 0.5. Based on the model parameter settings and algorithm descriptions described above, a comparison of the actual and predicted values of coal temperature in the test samples of different prediction models was finally obtained, as shown in Figure 3; The fitting effect of the training and testing samples is shown in Figure 4.

From Figures 3 and 4, it can be that the prediction model of borehole spontaneous combustion risk degree based on HGS-RF was almost optimal in terms of the coincidence degree of test samples and the fit degree of training samples and test samples. The performance of RMSE was slightly worse than that of QPSO-RF model, indicating that its prediction result is closest to the true value. In order to further quantify the predictive performance of each model, the model performance evaluation indicators mentioned above were summarized, as shown in Table 3. Table 3 shows that compared with other models, except the QPSO-RF model, the MAPE of the HGS-RF model in the test samples decreased by 3.6564%, 3.823% and 4.5266%, respectively. The MAE decreased by 7.0189, 5.0831 and 4.8982, respectively. The R2 increased by 0.0674, 0.0402 and 0.0451, respectively. Compared with the RF, SSA-RF and PSO-RF, the RMSE decreased by 9.3955, 3.0326 and 4.84, respectively. The RMSE and MAPE of the HGS-RF model were 0.1812 higher than those of the QPSO-RF model, but the R2 was 0.0049 higher, and the MAE was 0.1637 lower. The HGS-RF model not only had high prediction accuracy but also had a low discretization degree and good robustness for QPSO-RF. The MAPE, RMSE and MAE of RF, SSA-RF and PSO-RF models increased significantly in the test stage, while the R2 decreased significantly, indicating that there was overfitting in the test stage, which led to the reduction of model robustness and the increase of error of test samples. Although the prediction results of the HGS-RF model are similar to those of the QPSO-RF model, the prediction process of QPSO-RF model needs to set more

local learning factors and global learning factors, which takes a long time and occupies relatively more resources, which is not conducive to finding the optimal parameters, and the prediction results obtained were also close to the HGS-RF model.

Figure 3. Comparison of the real and predicted values for different model test samples. (**a**) RF model; (**b**) SSA-RF model; (**c**) PSO-RF model; (**d**) QPSO-RF model; (**e**) HGS-RF model.

Figure 4. Fitting plots of the training and test samples for different models. (**a**) RF model; (**b**) SSA-RF model; (**c**) PSO-RF model; (**d**) QPSO-RF model; (**e**) HGS-RF model.
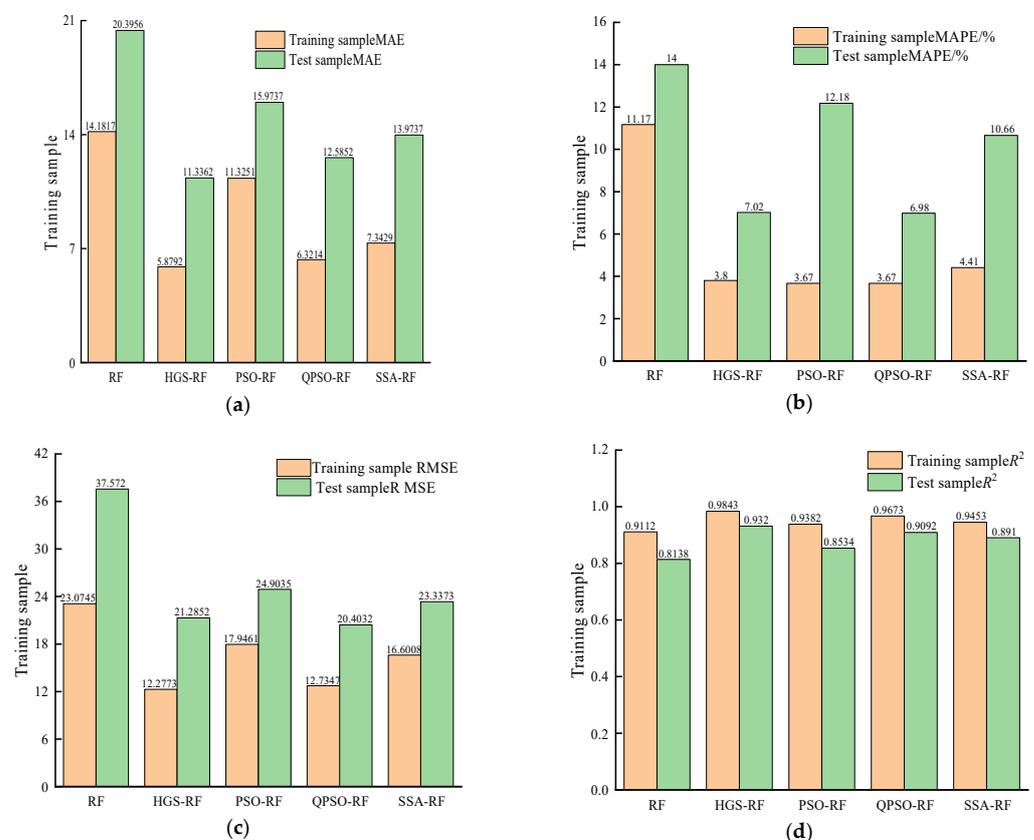
3.4.3. Model Reliability Verification

In order to illustrate the reliability of the HGS-RF model and make it more convincing, 96 sets of data from the 1204 coal face of the Zhengjia Coal Industry in Fenxi, Shanxi Province were used for verification. According to the actual situation, O_2, CO, CO_2, CH_4, C_2H_4, C_2H_6, $CO/\Delta O_2$ and C_2H_4/C_2H_6 were selected as input indicators, and temperature

was selected as the output indicator. Among them, the n_estimators and min_samples_leaf parameters in the RF were obtained by HGS optimization, and the values were 150 and 5, respectively. The other parameters were consistent with those in Section 3.4.2 of this paper. The comparison of evaluation indicators of each model is shown in Figure 4. It can be seen more intuitively from Figure 5 that the HGS-RF model had the best effect, and the MAE of each model test sample was 20.3956, 11.3362, 15.9737, 12.5852 and 13.9737, respectively. The MAPE was 14%, 7.02%, 12.18%, 6.98% and 10.66%, respectively. The RMSE was 37.572, 21.2852, 24.9035, 20.4032 and 23.3373, respectively. The R2 was 0.8138, 0.932, 0.8534, 0.9092 and 0.891, respectively. Combined with the above analysis results, it was finally determined that the regression analysis model of coal spontaneous combustion temperature based on HGS-RF model is a simple and reliable method, and the prediction results were closer to the actual situation.

Table 3. Comparison of the evaluation metrics for the different prediction models for a mine in Shandong Province.

Model	Model Performance							
	R^2		MAPE/%		RMSE		MAE	
	Train	Test	Train	Test	Train	Test	Train	Test
RF	0.9519	0.9043	5.46	13.81	15.7439	21.5646	7.9857	17.541
SSA-RF	0.9616	0.9315	4.3563	10.9766	12.4116	15.2017	7.3011	15.7752
PSO-RF	0.9654	0.9266	5.1076	9.6802	11.1453	17.0091	6.872	12.5903
QPSO-RF	0.9817	0.9668	3.892	4.5731	7.421	11.9879	6.751	6.8594
HGS-RF	0.9851	0.9717	4.87	5.1536	8.327	12.1691	5.3669	6.6921

Figure 5. Comparison of the evaluation indexes of the Zhengjia Coal Industry models. (**a**) Mean absolute error comparison; (**b**) average absolute percentage error comparison; (**c**) root-mean-square error comparison; (**d**) coefficient of determination comparison.

4. Analysis of Analysis of Engineering Examples

In order to further demonstrate the universality and stability of the borehole spontaneous combustion risk warning model based on HGS-RF, 170 groups of borehole samples from three coal seams in 1303 fully integrated caving face of the Jinniu Coal Mine in Shanxi Province were taken as the research objects, and the performance of the HGS-RF model in other mine applications was compared and analyzed. In a previous study [14], 60 sets of data of 7162 fully mechanized mining faces in the Longdong Coal Mine were used for engineering practice. In [17], 30 sets of data from Xuandong No. 2 Coal Mine were used for testing, and good results were obtained. Therefore, 170 sets of data can be representative. The 1303 fully mechanized caving face of the Jinniu Coal Mine is located in the first mining area of the No. 9 coal seam at the 1030 level. The thickness of the coal seam is 5.24~7.30 m with an average thickness of 6.17 m. The inclination of the coal seam is 8~14° with an average of 10°. The mining coal seam belongs to the high gas I type coal seam prone to spontaneous combustion, and the coal dust explosion index is 45.79%, which is explosive. To sum up, there is a risk of borehole spontaneous combustion in the Jinniu Coal Mine during the mining process, so it is necessary to predict the degree of spontaneous combustion risk to ensure the safe production of the mine. The basic parameters of each model are consistent with the above. The sample data of the three extraction boreholes in the coal seam were, respectively, brought into the RF model, PSO-RF model, SSA-RF model, QPSO-RF model and HGS-RF model for comparative analysis. The prediction performance of each prediction model in the application of other mining extraction boreholes was analyzed, and the evaluation index results of the prediction model are shown in Table 4.

Table 4. Comparison of the evaluation indexes of the different prediction models in the Jinniu Coal Mine.

Model	Model Performance							
	R^2		MAPE/%		RMSE		MAE	
	Train	Test	Train	Test	Train	Test	Train	Test
RF	0.9247	0.8815	14.67	17.16	14.3257	18.5114	9.7461	15.3247
SSA-RF	0.9395	0.9267	12.19	14.31	11.5645	15.1763	8.1965	13.2972
PSO-RF	0.9441	0.9075	11.84	18.55	14.1946	19.5681	7.2296	12.9467
QPSO-RF	0.9702	0.9418	8.29	9.61	9.8425	12.367	7.0325	10.8637
HGS-RF	0.9781	0.9556	8.65	11.13	10.0396	13.9179	6.8507	8.339

The results showed that the prediction model based on HGS-RF model was the best in terms of the MAE, MAPE, RMSE and R2, compared with the RF, SSA-RF and PSO-RF models. The MAPE was 1.52% higher than the QPSO-RF model, and the RMESE was 1.5509 higher than the QPSO-RF model. However, the R2 was also higher by 0.0138, the MAE was lower by 2.5247 and the running speed was the fastest among the five models, 2.371 s faster than the QPSO-RF model. Therefore, HGS-RF showed good accuracy, generalization performance and fast operating speed. The results show that the HGS-RF algorithm has strong universality, stability and excellent practicability.

5. Conclusions

(1)   By combining the Hunger Games search algorithm and the Random Forest algorithm, we constructed an early warning model for the hazard level of spontaneous combustion in boreholes based on the HGS-RF and compared the predictions with the RF, SSA-RF, PSO-RF and QPSO-RF models. The results show that the predictions of the HGS-RF model were closer to the actual situation, while the RF model had strong generalization performance but poor prediction accuracy. The SSA-RF and PSO-RF models were prone to overfitting, Although the prediction result of QPSO-RF model

is similar to that of HGS-RF model, the running time of this model is relatively long and more preparation is required, so the HGS-RF model is the most practical.

(2)   Compared to the RF, SSA-RF, PSO-RF and QPSO-RF models, the HGS-RF model showed a decrease in the MAE of 7.0189, 5.0831, 4.8982 and 0.1637, respectively, in the test samples. MAPE decreased by 3.6564%, 3.823%, 4.5266%, but it increased by 0.5805% compared to the QPSO-RF model, respectively; RMSE decreased by 9.3955, 3.0326 and 4.84, but it increased by 0.1812 compared to the QPSO-RF model, respectively; R^2 increased by 0.0674, 0.0402 0.0451 and 0.0049, respectively. The HGS-RF based drilling spontaneous combustion degree warning model can achieve more accurate prediction results without complex parameter settings and optimization, and it is robust and generalizable.

(3)   The reliability of the HGS-RF model was verified by taking the data of the 1024 coal face of the Zhengjia Coal Industry as an example. The results show that the HGS-RF model had certain reliability, and the results were the closest to the actual situation. In order to further verify the universality and stability of the HGS-RF model, it was applied to the Jinniu Coal Mine in Shanxi Province and compared with other regression models. The results show that the regression results of the HGS-RF model were accurate and reliable, and better results have been obtained in the prediction of borehole spontaneous combustion in different mines, indicating that the HGS-RF model can accurately predict the borehole spontaneous combustion temperature.

Author Contributions: Conceptualization, Y.Q. and K.X.; Validation, K.X., Y.Q. and W.W. Formal analysis, K.X. Research, K.X.; Resources provided by K.X. and X.C.; Data collation, K.X. Original draft written, K.X. Writing—Review and Editing, K.X. Visualization, R.L.; Monitoring work, Y.Q.; Project Management, W.W.; Funding access, Y.Q. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: No need to declare.

Informed Consent Statement: Not applicable.

Data Availability Statement: All data generated or analyzed during this study are included in this published article.

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

References

1.   Qiao, M.; Ren, T.; Roberts, J.; Yang, X.; Li, Z.; Wu, J. New insight into proactive goaf inertisation for spontaneous combustion management and control. *Process Saf. Environ. Prot.* **2022**, *161*, 739–757. [CrossRef]
2.   Wang, W.; Qi, Y.; Liu, J. Study on multi field coupling numerical simulation of nitrogen injection in goaf and fire-fighting technology. *Sci. Rep.* **2022**, *12*, 1–18. [CrossRef]
3.   Zhang, Y.; Niu, K.; Du, W.; Zhang, J.; Wang, H.; Zhang, J. A method to identify coal spontaneous combustion-proneregions based on goaf flow field under dynamic porosity. *Fuel* **2021**, *288*, 119690. [CrossRef]
4.   Jia, X. Study on Influencing Mechanism and Prevention Technology of Coal Spontaneous Combustion Caused by Borehole Gas Extraction. Ph.D. Thesis, Liaoning Technical University, Fuxin, China, 2022.
5.   Lu, H.; Zhao, Y.; Fang, Z. Study on the spontaneous combustion and ignition law of lignite in the coal mine of Shuilianta based on variable oxygen conditions. *Coal Sci. Technol.* **2021**, *49*, 152–158.
6.   Wen, H.; Wang, W.; Tao, W.; Chen, X.; Jiang, X.; Cheng, B. Research on spontaneous coal combustion prediction and prevention and control technology during the withdrawal of ultra-long synthesized mining face. *Coal Sci. Technol.* **2020**, *48*, 167–173.
7.   Lei, C.; Deng, J.; Cao, K.; Ma, L.; Xiao, Y.; Ren, L. A random forest approach for predicting coal spontaneous combustion. *Fuel* **2018**, *223*, 63–73. [CrossRef]
8.   Wang, F.; Wang, J.; Gu, L.; Liu, P. Determination method for coal spontaneous combustion degree based on grey target decision model. *China Sci. Pap.* **2019**, *14*, 980–984.

9. Qi, Y.; Qi, Q.; Wang, W.; Zhou, X. A forecast model for the spontaneous combustion risk in the goaf based on set valued statistics-entropy and its application. *J. Saf. Environ.* **2019**, *19*, 1526–1531.

10. Wang, W.; Jia, B.; Qi, Y. Prediction model of spontaneous combustion risk of extraction drilling based on improved CRITIC modified G2-TOPSIS method and its application. *China Saf. Sci. J.* **2019**, *29*, 26–31.

11. Zhou, X.; Zhu, Y.; Zhang, J.; Qin, S.; Wang, Y. Study on prediction model of coal spontaneous combustion based on PSO-XGBoost. *Min. Saf. Environ. Prot.* **2022**, *49*, 79–84.

12. Zhai, X.; Zhang, W.; Wang, K.; Wang, W. Research on comprehensive evaluation of spontaneous combustion hazard of mine coal based on SEM. *Coal Eng.* **2020**, *52*, 101–105.

13. Zhao, L.; Wen, G.; Shao, L. PCA-AdaBoost Prediction Model for Coal Spontaneous Combustion in Blanketed Areas with Unbalanced Data. *China Saf. Sci. J.* **2018**, *28*, 74–78.

14. Deng, J.; Lei, C.; Cao, K.; Ma, L. Random forest method for predicting coal spontaneous combustion in gob. *J. China Coal Soc.* **2018**, *43*, 2800–2808.

15. Zheng, X.; Li, M.; Zhang, Y.; Jiang, P.; Wang, B. Research on the prediction model of coal spontaneous combustion temperature based on random forest algorithm. *Ind. Mine Autom.* **2021**, *47*, 58–64.

16. Jia, P.; Lin, K.; Guo, F. A temperature prediction model for coal spontaneous combustion based on PSO-SRU deep artificial neural networks. *J. Mine Autom.* **2022**, *48*, 105–113.

17. Wen, T.; Yu, F. Research on the prediction of coal spontaneous combustion based on KPCA-Fisher discriminant analysis. *Min. Saf. Environ. Prot.* **2018**, *45*, 49–53+58.

18. Shukla, U.S.; Mishra, D.P.; Mishra, A. Prediction of spontaneous combustion susceptibility of coal seams based on coal intrinsic properties using various machine learning tools. *Environ. Sci. Pollut. Res.* **2023**, *30*, 69564–69579. [CrossRef] [PubMed]

19. Lawal, A.I.; Onifade, M.; Abdulsalam, J.; Aladejare, A.E.; Gbadamosi, A.R. On the Performance Assessment of ANN and Spotted Hyena Optimized ANN to Predict the Spontaneous Combustion Liability of Coal. *Combust. Sci. Technol.* **2022**, *194*, 1408–1432. [CrossRef]

20. Zhou, X.; Yang, J.; Zu, Y.; Zhang, J. Study on gas outflow prediction based on NMF-HGS-RF. *Min. Saf. Environ. Prot.* **2023**, *50*, 117–123.

21. Yang, Y.; Chen, H.; Heidari, A.A.; Gandomi, A.H. Hunger Games Search: Visions, Conception, Implementation, Deep Analysis, Perspectives, and Towards Performance Shifts. *Expert Syst. Appl.* **2021**, *177*, 114864. [CrossRef]

22. Yang, Y. Intelligent Optimization Algorithm for Node Localization Based on Multiple Communication Radii. Master's Thesis, Lanzhou University of Science and Technology, Lanzhou, China, 2022.

23. Li, C.; Zhao, L. CACC-RF-based risk prediction of rutters indicating notch jamming failures. *J. China Railw. Soc.* **2022**, *44*, 46–55.

24. Tian, R.; Meng, H.; Cheng, S.; Chen, S.; Wang, C.; Shi, L. Rockburst Intensity Classification Prediction Model under RF-AHP-Cloud Modeling. *China Saf. Sci. J.* **2020**, *30*, 166–172.

25. Cheng, J.; Shang, D.; Zhao, Z.; Chen, C. A machine learning-based method for predicting fracture behavior of rock samples containing filled fractures. *J. Rock Mech. Eng.* **2023**, *42*, 3458–3472.

26. Jiang, P. A Machine Learning-Based Model for Predicting the Temperature of Spontaneous Coal Combustion. Master's Thesis, Xi'an University of Science and Technology, Xi'an, China, 2020.

27. Sun, H.; Yang, J.; Jing h Tu, H.; Zhou, X. NOx prediction model for coal-fired boilers based on Bayesian optimization-random forest regression. *J. Chin. Soc. Power Eng.* **2023**, *43*, 910–916.