

# Improved random forest model for smoke detection and early warning based on sparrow search algorithm

Jun Gao<sup>1,2</sup>, Duo Jia<sup>1</sup>, Nan Hu<sup>1</sup>

<sup>1</sup>Inner Mongolia Institute of Metrology Testing and Research, Hohhot, Inner Mongolia, 010050, China

<sup>2</sup>513687033@qq.com

**Abstract.** Smoke detection technology is of great significance in the field of fire safety, aiming at the timely detection of smoke released from fires or other combustion events to safeguard people's lives and properties. In this paper, a smart smoke alarm device is developed based on AI algorithm using UCL open source dataset, which is processed by AI recognition after collecting data through IoT to achieve accurate judgement of smoke situation. The dataset is divided by 6:4 ratio, and the training set confusion matrix shows that all 1159 smoke alarm tests are correctly predicted, of which 445 times predict no smoke and 714 times predict smoke, with an accuracy rate of 100%. The test set confusion matrix shows that only 1 out of 1158 tests was incorrectly predicted, of which 184 predicted no smoke, 312 predicted smoke, and 1 was misjudged as no smoke even though it should have been smoke, with an accuracy rate of 99.7%. The successful application of this technology provides a reliable guarantee for fire warning and demonstrates the great potential of AI in the field of security.

**Keywords:** Random Forest, Sparrow Search Algorithm, Smoke Detection.

## 1. Introduction

Smoke detection is an important safety technology aimed at the early detection of smoke from fires or other combustion events and the timely implementation of measures to protect people's lives and property. The background of the research stems mainly from the awareness of fire hazards and the urgent need for fire prevention. Fires are often accompanied by the release of large amounts of toxic gases and smoke, which contain large amounts of carbon monoxide and other harmful substances that are extremely hazardous to human health [1,2]. Therefore, it is of great significance to carry out the research and application of smoke detection technology to improve the fire warning capability and reduce casualties and property losses.

Machine learning algorithms play a key role in smoke detection and warning. The traditional smoke detection system is usually based on a fixed threshold to determine whether to trigger the alarm, and this method has a high false alarm rate and a high leakage rate [3]. Machine learning algorithms, on the other hand, can learn and analyse a large amount of data, find out the laws and build models, and achieve a more intelligent and more accurate smoke detection system.

First of all, machine learning algorithms can train the model through supervised learning, so that it can accurately determine whether there is a fire hazard based on the data obtained from the sensors [4,5].

For example, neural network algorithms can be used to identify different types of smoke features and compare them with pre-fire baseline data to achieve more accurate warnings.

Secondly, machine learning algorithms can also be used to mine potential correlations and regularities between data through unsupervised learning. This approach is suitable for those cases where labels or categories cannot be determined in advance, allowing the system to learn the intrinsic structure between the data on its own and make judgements accordingly [6].

In addition, in practical applications, machine learning algorithms can also be combined with multiple sensor data for comprehensive analysis to improve detection accuracy and reliability [7]. For example, combining temperature, humidity, light and other multiple parameter data for comprehensive analysis, to verify whether there is a fire risk in multiple dimensions.

Machine learning algorithms play an increasingly important role in smoke detection and early warning, providing new ideas and solutions to improve the efficiency of fire early warning and reduce the rate of false alarms and missed alarms. In this paper, the random forest model improved based on sparrow search algorithm perceives various environmental variables and predicts smoke alarms, which provides a certain research basis for subsequent research. In the future, with the continuous progress of technology and algorithm performance improvement, we believe that machine learning will play a greater potential in the field of smoke detection, and bring people a more intelligent, safe and reliable living environment.

## 2. Data set sources and data analysis

The dataset used in this paper is from the UCL open source dataset, based on which this paper develops a smoke alarm device based on artificial intelligence algorithms, which is able to collect data through IoT after which it is recognised and processed by the artificial intelligence to determine whether a smoke alarm is required or not. The training data is collected with the help of IoT devices as the goal is to develop an AI based smoke detector device.

The dataset consists of various data collected by the IoT device including environmental metrics such as temperature, humidity, TVOC, dioxide emissions, raw material, H2, raw ethanol, pressure, PM1.0, PM2.5, NC0.5, NC1.0, and NC2.5, as well as the final judgement (whether a smoke alarm is required or not). Descriptive statistics were performed on the data and the results are shown in Table 1.

**Table 1.** Model evaluation parameter.

Variable name	Max	Min	Average	Standard deviation
Temperature	22.73	8.94	19.704	3.054
Humidity	66.44	44.69	51.733	4.106
eCO2	462	400	401.934	6.675
Raw Ethanol	20438	18520	19983.822	235.217
Pressure	939.832	938.811	939.603	0.227
PM1.0	27.42	0	1.691	0.952
PM2.5	28.49	0	1.774	0.979
NC0.5	188.7	0	11.602	6.613
NC1.0	29.426	0	1.835	1.012
NC2.5	2.78	0	0.057	0.146

## 3. Method

### 3.1. Sparrow Search Algorithm

The sparrow search algorithm is a heuristic optimization algorithm based on the foraging behaviour of birds, which simulates the strategy of sparrows in searching for food. The algorithm was proposed by a research team from Nanchang University of Aeronautics and Astronautics, China, in 2019, and implements search and optimisation in the solution space by simulating the sparrow's process of

searching for food [8]. The sparrow search algorithm abstracts the problem space into a two-dimensional plane and represents a candidate solution as a point on the plane, with each point corresponding to a solution. The sparrow will decide the direction and distance to move next during the search process based on the pheromone concentration and food concentration near the current position. Here the pheromone concentration can be understood as the attractiveness of the local optimal solution, while the food concentration represents the attractiveness of the overall optimal solution.

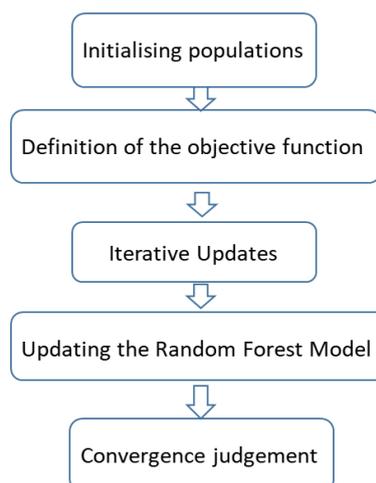
The sparrow search algorithm consists of two important phases: foraging behaviour and flocking behaviour. In the foraging behaviour phase, each sparrow explores locally based on the pheromone and food concentrations near its current location and chooses to move to a new location. In the cluster behaviour phase, on the other hand, the optimal solution is searched globally by exchanging information and cooperating and collaborating. During the iterative process, the sparrow search algorithm constantly updates the position of each sparrow and evaluates the merits of each candidate solution based on the objective function value [9]. By updating the optimal solution iteratively and combining the interaction of foraging behaviour and flocking behaviour, the global optimal solution or near-optimal solution is finally found.

Overall, the sparrow search algorithm achieves efficient search and optimisation in a complex problem space by simulating the intelligent strategy of the bird capability when searching for food. It has the characteristics of fast convergence speed, easy implementation, and applicability to a wide range of problem types, and has extensive potential and value in many application areas.

### 3.2. random forest algorithm

The Random Forest algorithm is an integrated learning method that performs classification or regression by constructing multiple decision trees and combining their predictions. The Random Forest algorithm, proposed by Leo Breiman and Adele Cutler in 2001, is a powerful and widely used machine learning algorithm. Random forest algorithm is a powerful and flexible integrated learning method, which is widely used in practical applications for tasks such as classification, regression, feature selection and has achieved good results in many fields [10].

In this paper, we use the sparrow search algorithm to optimise the random forest model for smoke alarm prediction, and the flowchart of the algorithm is shown in Figure 1.



**Figure 1.** Target detection outputs.  
(Photo credit : Original)

Initialising the population: First, the SSA algorithm is used to initialise a population of "sparrows", each representing a possible solution or parameter configuration. Here, the "sparrows" can be regarded as a set of decision tree parameters, including the number of trees, the maximum depth, the number of features to be considered when splitting nodes, and so on.

**Objective function definition:** For the random forest model, a fitness function needs to be defined as an objective function in SSA to evaluate the performance of each individual (i.e., parameter configuration) on the current problem. Usually the accuracy of the model on the validation set or other performance metrics can be chosen as the fitness function.

**Iterative update:** During the iterative process, each "sparrow" explores locally based on the pheromone and food concentrations near its current location and chooses to move to a new location. The movement here can be understood as updating the parameter configuration of the individual. By updating the optimal solution iteratively and combining the interaction of foraging and flocking behaviours, the optimal parameter configurations are finally found.

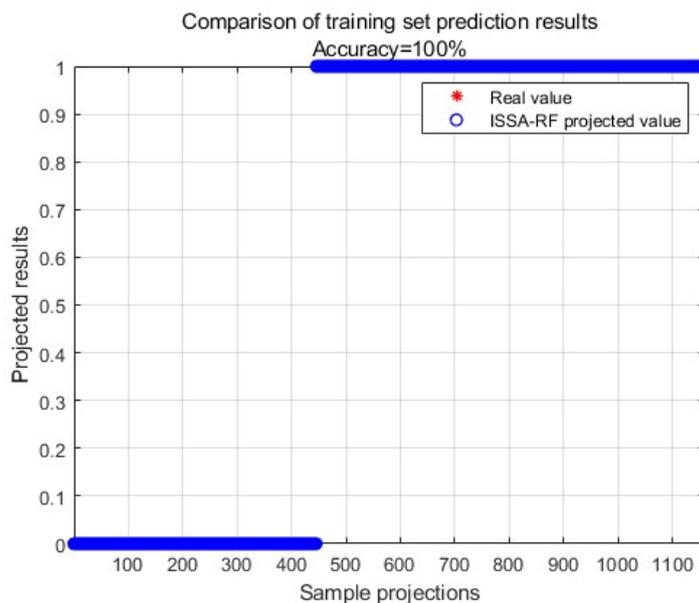
**Updating the random forest model:** After each iteration, the optimal parameter configuration obtained is used to train a new random forest model. This makes the random forest model more adaptable to the current problem and improves its prediction performance.

**Convergence judgement:** The algorithm is judged to have converged by setting a stopping condition or reaching the maximum number of iterations. Generally, the algorithm is considered to have converged when there is no significant improvement in the fitness function value for several consecutive generations.

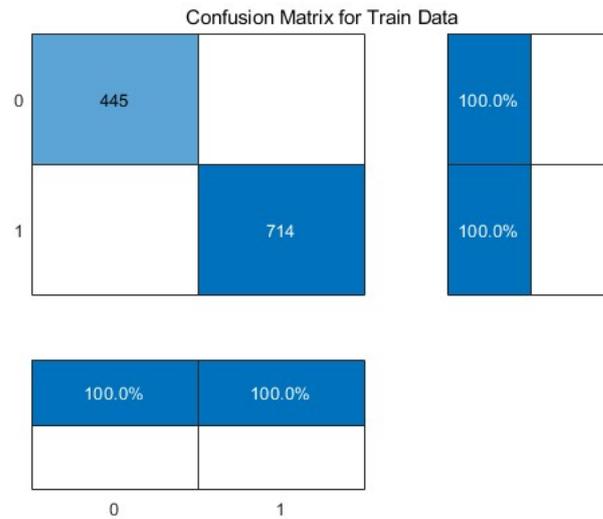
By combining the sparrow search algorithm with the random forest model, the performance of the random forest model can be effectively optimised for a particular problem. This combination exploits the characteristics of SSA's ability to perform both global and local searches, making the resulting random forest model more generalisable and accurate.

#### 4. Result

The dataset was preprocessed to remove data with outliers and missing values, then the data was divided in a ratio of 6:4, 40% of the data was used for model testing and 60% of the data was used for model training, using the test set to output the confusion matrix and accuracy of the smoke alarms, the predictions for the model training set are shown in Fig. 2, the confusion matrix for the training set is shown in Fig. 3, and the confusion matrix for the test set is shown in Fig. 4 shown.

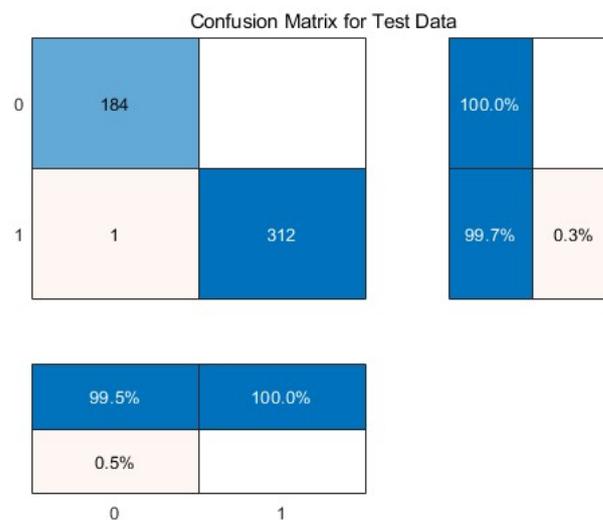


**Figure 2.** The predictions for the model training set.  
(Photo credit : Original)



**Figure 3.** The confusion matrix for the training set.  
 (Photo credit : Original)

From the confusion matrix of the training set, it can be seen that all 1159 smoke alarm tests were predicted correctly, of which 445 were predicted as no smoke present and 714 were predicted as smoke present. The accuracy is 100 per cent.



**Figure 4.** The confusion matrix for the test set.  
 (Photo credit : Original)

From the confusion matrix of the test set, it can be seen that all 1158 smoke alarm tests were predicted correctly, with only one incorrect prediction, of which 184 were predicted to be smoke absent, 312 were predicted to be smoke present, and one should have been predicted to be smoke present but was actually predicted to be smoke absent, giving an accuracy of 99.7%.

## 5. Conclusion

Smoke detection technology plays a vital role in the security sector of today's society. It aims to safeguard people's lives and property by timely detection of smoke released during fire or other combustion events. The dataset used in this paper is derived from the UCL open source dataset, based on which we have successfully developed a smoke alarm device based on artificial intelligence

algorithms. The device collects data through IoT and determines whether a smoke alarm is required after recognition and processing by AI .

For the training set, we divided the data in the ratio of 6:4. Confusion matrix analysis shows that out of 1159 smoke alarm tests, all predictions were accurate, with 445 being correctly predicted as no smoke and 714 being accurately predicted as smoke. The accuracy rate is an amazing 100 per cent. As for the test set, the confusion matrix shows that out of 1,158 smoke alarm tests, 1 prediction was incorrect. Specifically, 184 were correctly predicted as no smoke and 312 were accurately predicted as smoke. Despite the presence of 1 instance that should have been correctly predicted as having smoke but was misjudged as having no smoke, the overall accuracy was still 99.7%.

In summary, this smoke alarm device developed based on AI algorithms performed well, achieving satisfactory results on both the training and test sets. This not only validates the advantages of the device in terms of effectiveness and reliability, but also provides strong support for the future application of similar technologies in the security field. It is hoped that this technology can be further promoted and improved to better serve the public and contribute to building a safer and more reliable living environment.

## References

- [1] Dang, Bo, et al. "Enhancing Kitchen Independence: Deep Learning-Based Object Detection for Visually Impaired Assistance." *Academic Journal of Science and Technology* 9.2 (2024): 180-184.
- [2] Dong, Xinqi, et al. "The prediction trend of enterprise financial risk based on machine learning arima model." *Journal of Theory and Practice of Engineering Science* 4.01 (2024): 65-71.
- [3] Zang, Hengyi, et al. "Evaluating the social impact of ai in manufacturing: A methodological framework for ethical production." *Academic Journal of Sociology and Management* 2.1 (2024): 21-25.
- [4] Ma, Danqing et al. "Implementation of computer vision technology based on artificial intelligence for medical image analysis." *International Journal of Computer Science and Information Technology* (2023): n. pag.
- [5] Ma, Danqing, Bo, Dang, Shaojie, Li, Hengyi, Zang, Xinqi, Dong. "Implementation of computer vision technology based on artificial intelligence for medical image analysis". *International Journal of Computer Science and Information Technology* 1. 1(2023): 69–76.
- [6] Ao Xiang, , Zongqing Qi, Han Wang, Qin Yang, Danqing Ma. "A Multimodal Fusion Network For Student Emotion Recognition Based on Transformer and Tensor Product." (2024).
- [7] Emel S ,Tuncay S .A performance comparison of YOLOv8 models for traffic sign detection in the Robotaxi-full scale autonomous vehicle competition[J].*Multimedia Tools and Applications*,2023,83(8):25005-25035.
- [8] Mudawi A N,Qureshi M A ,Abdelhaq M , et al.Vehicle Detection and Classification via YOLOv8 and Deep Belief Network over Aerial Image Sequences[J].*Sustainability*,2023,15(19):
- [9] Karakuş S ,Kaya M,Tuncer A S.Real-Time Detection and Identification of Suspects in Forensic Imagery Using Advanced YOLOv8 Object Recognition Models[J].*Traitement du Signal*,2023,40(5):
- [10] HyoYeon J ,CheolHong H .Obscuration Threshold Database Construction of Smoke Detectors for Various Combustibles.[J].*Sensors (Basel, Switzerland)*,2020,20(21):6272-6272.