

## RESEARCH ARTICLE

# Research on Flow Decision-Making Model of Plant Protection UAV Based on Feature Selection

MENG WANG<sup>1,2,3</sup>, ZHIHAO BIAN<sup>1,2,3</sup>, YU YAN<sup>1,2,3</sup>, MUJAHID HUSSAIN<sup>1,2,3</sup>,  
GUOBIN WANG<sup>1,2</sup>, CANCAN SONG<sup>1,2,3</sup>, AND YUBIN LAN<sup>1,2,3</sup>

<sup>1</sup>College of Agricultural Engineering and Food Science, Shandong University of Technology, Zibo, Shandong 255022, China

<sup>2</sup>Shandong University of Technology Sub-center of National Center for International Collaboration Research on Precision Agricultural Aviation Pesticide Spraying Technology, Zibo, Shandong 255022, China

<sup>3</sup>Academy of Ecological Unmanned Farm, Shandong University of Technology, Zibo, Shandong 255022, China

Corresponding authors: Cancan Song (songcc@sdut.edu.cn) and Yubin Lan (ylan@scau.edu.cn)

This work was supported in part by the Special Funding Project for Shandong Province Natural Science Foundation under Grant ZR2021QC154, in part by the Research and Demonstration on Key Technologies of Agricultural Drone Monitoring and Prevention and Control of Dominant Crop Pests and Diseases in Ningxia under Project 2023BCF01051, and in part by the National Natural Science Foundation of China under Grant 32302410.

**ABSTRACT** The field environment is complex and variable, and multiple factors constrain the effectiveness of UAV applications, and a single flow applications may result in over- or under-use of pesticides in plots with different requirements. Therefore, it is crucial to study a decision-making model of flow rate for plant protection UAVs under multi-factor interaction. In this paper, based on a large amount of experimental data, combined with Pearson correlation analysis and random forest variable importance score ranking, screening the features obtained from the experiment increases the correlation between input and output, making the output results more reliable. The model evaluation results showed that the GA-BP neural network model has a correlation coefficient of 0.99 between the true value, predicted value, and a coefficient of determination of 0.98, which is better than the general regression model. A validation test was conducted to test the effectiveness of the model for new data. The final result yields an error value within  $\pm 20\%$  for the GA-BP model to predict the flow rate. At the same time, the BP neural network fluctuated more for some of the predicted values, which caused a 50% error in fitting results. It proves the feasibility of the BP neural network optimized based on feature screening and genetic algorithm in plant protection UAV flow rate decision-making, which can provide a reference basis and scientific guidance for precise variable spraying operation of plant protection UAVs.

**INDEX TERMS** Plant protection drone, BP neural network, genetic algorithm, variable spraying, decision model, spraying flow rate.

## I. INTRODUCTION

Pests, diseases, and weeds pose a significant challenge to cultivating high-quality and high-yielding crops worldwide and are often accompanied by outbreaks of severe and frequent nature [1]. Although chemical control is currently the primary means of weed control, disease prevention, and pest management, the improper use of pesticides can result in issues like environmental pollution and low pesticide efficiency. Existing technology allows for real-time adjustment of spraying parameters based on the canopy characteristics of crops and weeds, as crops and weeds, and the operating parameters of plant protection drones. This effectively reduces pesticide

residues, increases pesticide utilization, improves control effects, and is expected to become a significant method for achieving pesticide efficiency and reducing usage in the future [2], [3], [4]. Research manuscripts reporting large datasets that are deposited in a publicly available database should specify where the data have been deposited and provide the relevant accession numbers. At present, there has been a lot of research on plant protection UAV variable spraying technology, such as variable spraying technology based on prescription maps, according to crop growth, soil fertility, or the severity of pests and weeds and other information to create prescription maps to realize variable spraying [5], [6], [7] ultimately. Through the network RTK Positioning and PWM technology to adjust the duty cycle, the UAV variable spray can be realized, or the flow of UAV can be through the

The associate editor coordinating the review of this manuscript and approving it for publication was Junhua Li<sup>1</sup>.

flight altitude, speed, wind speed, and other factors to learn the function of UAV variable spraying and achieve the desired effect [8], [9]. However, most of the existing studies are based on the linear model between these factors and flow value to achieve variable spraying, which is challenging to achieve the desired effect. In practical applications, many factors affect droplet deposition and drift, and the interaction between these factors will lead to the difference between the expected value and the actual value of droplet deposition. Therefore, it is necessary to study a dynamic decision-making model of plant protection UAV flow based on multiple factors.

The BP neural network proposed by D.E. Rumelhart in the 1980s is a parallel nonlinear extensive dynamic system, which can solve the learning problem of multi-layer neural networks by establishing complex nonlinear mapping and building relational models based on pre-provided input and output data sets. It has been widely used in agriculture [10], [11], [12], [13]. Established the neural network decision model of plant protection UAV using neural network technology and image processing technology. Azizpanah et al. Used neural network technology and image processing technology to establish the prediction model between spray drift and input parameters in the wind tunnel, and the prediction accuracy was higher than 0.96. Lazarovitch used the BP neural network model to predict soil hydraulic properties, obtained the relationship model between input and output through training, and used the actual test parameters to obtain the output value to evaluate soil water capacity, which achieved the expected effect [14]. Through the above research, the BP neural network can perform nonlinear fitting on multi-source data, exert self-learning and self-adaptive ability, adjust the connection relationship between internal nodes of factors, and finally establish a high-precision model. However, due to the randomness of model weights and threshold initialization, BP neural network training results are prone to instability and overfitting, and the optimal network model cannot be obtained. Therefore, it is necessary to find an optimization method to improve the stability and prediction accuracy of neural network.

As one of the most widely used optimization algorithms, genetic algorithm (GA) is a mathematical model established by simulating the biological evolution principle of “natural selection and survival of the fittest” in nature. It can compare and screen the parameters of the original model and finally select the optimal parameters of the model through continuous iterative adjustment [10], [15]. Yajie Shi used BP neural network optimized based on a genetic algorithm (GA-BP) to input geographical environment and terrain elements into the model to obtain the soil moisture prediction model. The cross-validation  $r$  value of the model was 0.86, and the ubmse was 0.03 [16]. Jian Gu used GA-BP neural network to establish the irrigation water model of corn yield under different irrigation regimes under subsurface drip irrigation. The model’s average error was only 0.71%. The model accelerates the convergence speed of the network, improves prediction accuracy, and more accurately describes the rela-

tionship between yield and irrigation water under subsurface drip irrigation [17]. Hu Jin constructed a prediction model of photosynthetic rate based on the GA-BP neural network. The correlation coefficient between the predicted and measured values was 0.98, and the absolute error of photosynthetic rate was less than  $\pm 0.5 \mu\text{mol}/\text{m}^2/\text{s}$ . The model predicted the yield and irrigation water under different irrigation systems under subsurface drip irrigation conditions. The performance and accuracy of the model was significantly better than that of the single neural network prediction model [18].

Most of the current studies only focus on single-factor variable spraying, and model training mostly relies on the empirical method or trial and error method to select network features and parameters, which easily leads to the neglect of essential features, reduces the accuracy of the network, and leads to the occurrence of over-fitting or under-fitting. To solve these problems, this paper combines a variety of parameters obtained from deposition and field experiments and increases data relevance by analysing correlations between parameters as well as filtering features and increasing data relevance by ranking the importance of Random Forest features and conducting ablation experiments. At the same time, the GA-BP neural network is constructed to improve the instability of network training caused by random parameters in BP neural network. The final expected droplet deposition can improve the spraying quality and reduce the use of chemical pesticides, thus reducing environmental pollution and improving crop quality. Materials and Methods.

## II. MATERIALS AND METHODS

### A. DATA COLLECTION

The field test of this model was conducted in Linzi District, Zibo City, Shandong Province (E118°12'50"; N36°57'47") at the ecological unmanned farm base of the Shandong University of Technology, and the model training was carried out through the data obtained from the experiment.

According to the spraying width of the aircraft, the field test size is set as 9m\*24m, and the sampling point interval is 1m, as shown in figure 1. DJ T20 and XAG XP2020 plant pro-

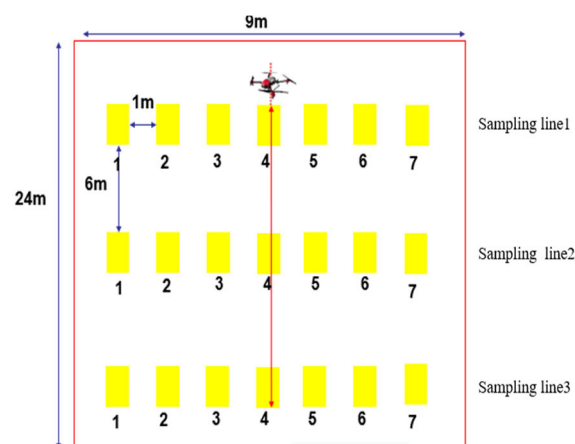


FIGURE 1. Layout of experimental site.

tection UAVs were selected as the spraying platform to spray 5g/L Allura red solution, record the flow rate, UAV parameters, and other data, and obtain the temperature, humidity, wind speed, wind direction, and other meteorological data by setting a Kestrel portable weather station 2m away from the test point during the test. Finally, at the end of the experiment, the deposition amount was obtained by using ImageJ software to scan the copper plate card [19]. Finally, we get Deposition quantity (X1), Application amount (X2), angle between wind direction and flight direction (Afterwards, it was called direction)(X3) [20], [21], rote (X4), nozzle type (X5) [21], [22], flight height (X6), wind speed (X7), flight speed (X8), number of nozzles (X9), temp (X10), wheelbase (X11), number of rotor (X12) [23], humidity (X13), flow rate (X14) for a total of 14 types of parameters.

**B. FEATURE SELECTION**

The data obtained from the experiment contains a large number of different data, which will lead to the model being too complicated, so the data should be filtered for features before being input into the model.

**1) PEARSON CORRELATION ANALYSIS**

In the natural sciences, Pearson’s correlation coefficient is widely used to measure the degree of correlation between two variables, with a value between -1 and 1.

The formula for calculating Pearson’s correlation coefficient is as follows [24], [25]:

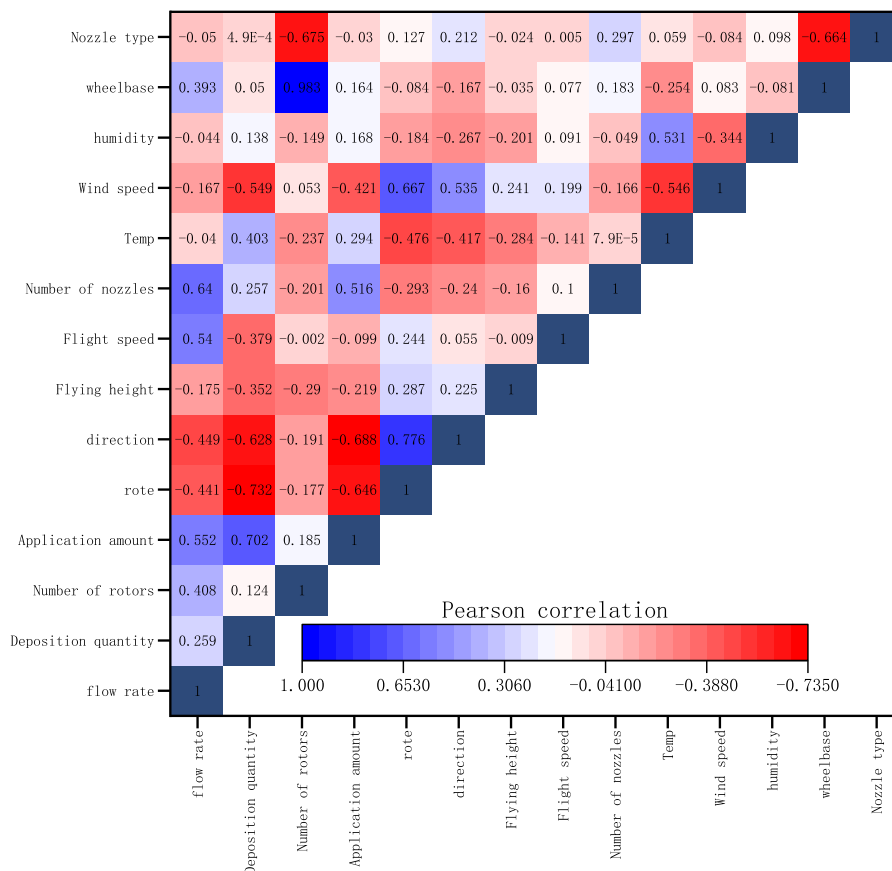
$$Cov(X, Y) = \frac{\sum_{i=1}^n (x_i - E(X))(y_i - E(Y))}{n} \tag{1}$$

$$\rho_{X,Y} = \frac{Cov(X, Y)}{\rho_X * \rho_Y}$$

$$= \frac{\sum_{i=1}^n (x_i - E(X))(y_i - E(Y))}{\sqrt{\sum_{i=1}^n (x_i - E(X))^2} * \sqrt{\sum_{i=1}^n (y_i - E(Y))^2}} \tag{2}$$

$Cov(X, Y)$  is the covariance,  $\rho_{X,Y}$  is the Pearson’s coefficient, x, y are the bivariate data, E(X), E(Y) is the expectation of the bivariate data.  $\rho_X, \rho_Y$  are the standard deviation of the bivariate data.

After integrating the obtained experimental data, Pearson correlation analysis and statistical analysis were performed on the filtered data using SPSS software to get the correlation heat map of the factors affecting the deposition amount (Fig. 2) and the statistical data of spraying parameters (Table 1), to eliminate redundant features, further improve the prediction accuracy of the model and avoid overfitting of the model. Combining the correlation heat map with the existing literature analysis ten categories of features X1, X3, X4, X5, X6, X7, X8, X9, X10, X12 are selected. Thus, after the



**FIGURE 2. Correlation thermogram of influencing factors.**

TABLE 1. Field spraying parameters of plant protection UAV.

Type	Number of rotors	Acre amount (L·hm <sup>-2</sup> )	Altitude (m)	Speed (m·s <sup>-1</sup> )	Number of nozzles	Nozzle type	Temperature (°C)	Humidity (%)	Wind speed (m·s <sup>-1</sup> )	Wheel base (mm)	Flow rate (L·min <sup>-1</sup> )
T20	6	1~3	2~5	3~5	4、8	SX110015vs、 SX11001vs、 SX11002vs	11.2~35.80	23.80~85.70	0~7.10	1883	1.23~4.00
XP2020	4	1~3	3~5	3~5	4	XR11002vs Centrifugal atomizing nozzles, droplet size selection 150μm	27.80~31.10	56.00~72.80	1.80~3.80	1680	0.77~2.06

feature selection process, new input features will be obtained as follows:

$$x_t = (X1, X3, X4, X5, X6, X7, X8, X9, X10, X12) \quad (3)$$

## 2) RANDOM FOREST VARIABLE IMPORTANCE SCORE

In order to observe the importance of features on decision flow this paper uses random forests for variable importance scoring. Random Forest is an efficient integrated learning algorithm that cleverly incorporates decision trees and bootstrap resampling techniques. In the process of constructing a random forest, each feature is assigned an importance score that reflects the extent of the feature’s contribution in constructing all the decision trees. The importance score of a feature in a random forest is calculated based on the number of times it participates in leaf node splitting in all decision trees. The more times a feature participates in splitting, the more useful it is in constructing the decision tree, and therefore the higher its score. In order to assess the importance of each feature, the random forest algorithm calculates the average of each feature’s contribution over all the decision trees and selects features by comparing the size of the contribution between different features. In this paper, the Gini index (GI) is used as a criterion for calculating the contribution of features, and its calculation formula is as follows [26], [27], [28]:

$$GIC = 1 - \sum_{k=1}^{|k|} p_{Ck}^2 \quad (4)$$

k denotes the number of features and p<sub>Ck</sub> denotes the proportion of category k in node C.

The importance of feature XJ at node q of the ith tree, i.e., the amount of change in the Gini index before and after branching of node q is:

$$VIM_{jq}^{(Gini)(i)} = GI_q^{(i)} - GI_l^{(i)} - GI_r^{(i)} \quad (5)$$

where  $GI_q^{(i)}$  and  $GI_l^{(i)}$  denote the Gini indices of the two new nodes after branching, respectively.

If, feature XJ appears in decision tree i in the set of nodes Q, then the importance of XJ in the ith tree is::

$$VIM_j^{(Gini)(i)} = \sum_{q \in Q} VIM_{jq}^{(Gini)(i)} \quad (6)$$

Suppose there are I trees in the random forest

$$VIM_j^{(Gini)} = \sum_{i=1}^I VIM_j^{(Gini)(i)} \quad (7)$$

Finally, all the obtained importance scores were normalised.

$$VIM_j^{(Gini)} = \frac{VIM_j^{(Gini)}}{\sum_{j=1}^J VIM_j^{(Gini)(i)}} \quad (8)$$

The features screened by Pearson’s correlation were subjected to Random Forest Variable Importance Scoring to finally get the importance ranking of the features as: X10(0.242), X7(0.164), X3(0.15), X1(0.146), X8(0.092), X5(0.088), X2(0.041), X6(0.039), X9(0.024), X4(0.013).

## C. NEURAL NETWORK CONSTRUCTION

In this paper, a BP neural network structure model was built based on Python, as shown in Fig 3. The training data are 160 groups, divided into the training set and test set according to the ratio of 8:2. The BP neural network consists of two processes: forward propagation of input signal and backpropagation of error, the core of which is to compare the desired output with the operational output, and iteratively update the network by backpropagation of error according to the gradient descent algorithm so that the final error meets the requirements [29].

The genetic algorithm optimizes the weights and thresholds of the BP neural network by selecting individuals with higher fitness to obtain the optimal parameters as the initial thresholds and weights of the BP neural network. The specific operation process of the GA-BP neural network is shown in Fig 4.

- (i) Data pre-processing: This study on the data set within the text-type data to take the unique thermal encoding and data standardization way to process the data standardization processing formula as shown in (9).

$$x = \frac{X - \text{mean}.X}{\text{std}.X} \quad (9)$$

where X is the original data, mean. X is the mean, and std.X is the standard deviation.

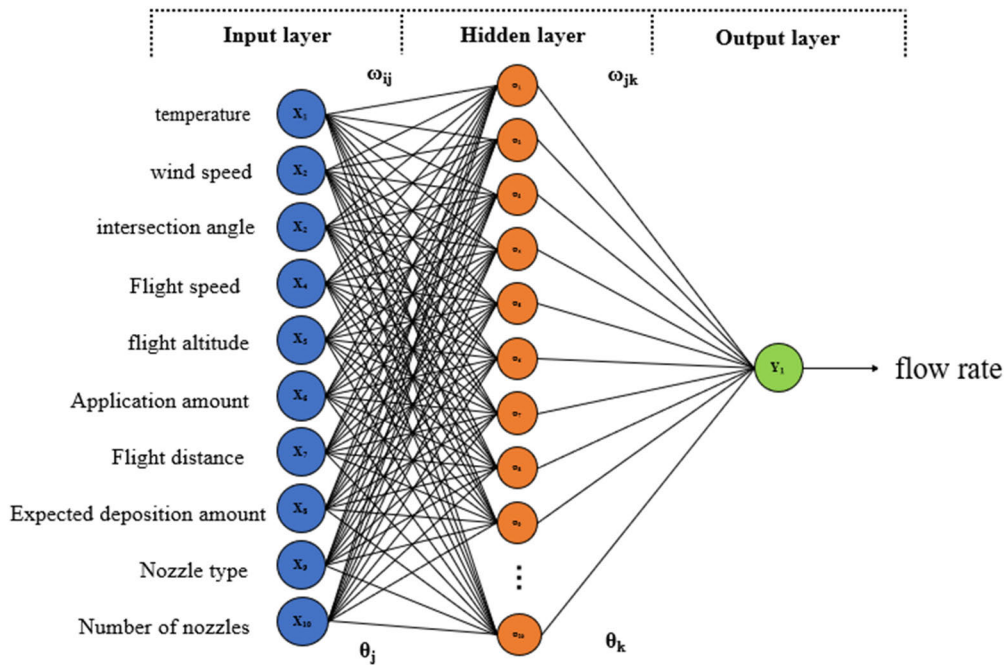


FIGURE 3. BP neural network structure diagram.

- (ii) Confirmation of BP neural network topology: The model adopts a single hidden layer structure, in which the number of neurons in the input layer is 15, and to prevent the occurrence of underfitting, which may be caused by the amount of data as well as the number of features, the number of neurons in the hidden layer is first set to 14 by the empirical formula (10), and is set to 20 by the neural network parameter test, and the number of neurons in the output neural network is 1.

$$h = \sqrt{i + k} + a \tag{10}$$

where h is the number of neurons, i is the number of neurons in the input layer, k is the number of neurons in the output layer, and a is any number between 2 and 10.

- (iii) Initialising the population: setting the population size and the number of iterations and randomly generating a population where the individuals are an array of weights and thresholds of the neural network.
- (iv) Setting the Fitness Function: The sum of the absolute value of the error between the true value and the predicted value is set as the fitness function, as shown in equation(11).

$$fit = \frac{1}{n} \sum_{i=0}^k (y_i - y_{\hat{a}i})^2 \tag{11}$$

where fit is the fitness, k is the size of the data set, y and y\_hat\_i are the actual and predicted values, respectively, and n is the size of the data set.

- (v) Selection: This study uses a tournament algorithm to randomly select individuals in the population while

using an elite retention strategy to ensure that better individuals can be selected to form a new population.

- (vi) Crossover and variation: Two individuals in the population are randomly selected to form new individuals based on a uniform crossover operator for crossover operation and Gaussian variation. In this paper, we adopt the method of adaptive crossover probability and mutation probability so that the neural network can jump out of the local minimum and find the optimal solution.

$$P = \begin{cases} P_{max} & f_{ave} > f_{best\_ave} \\ P_{max} - \frac{i}{n} (P_{max} - P_{min}) & f_{ave} \leq f_{best\_ave} \end{cases} \tag{12}$$

f\_ave is the average fitness to carry out the current generation, f\_best\_ave is the optimal average fitness in the iterative process, n is the total number of iterations, i is the f\_ave is the average fitness to carry out the current generation, P\_max, P\_min, is the maximum and minimum value of crossover probability and variation probability.

- (vii) Calculate the fitness: Calculate the fitness of the individuals in the population after selection, crossover and mutation, get the individual corresponding to the optimal fitness, and iterate until the optimal solution is gradually approached.
- (viii) Input the optimal weights and thresholds into the GA-BP neural network for training.

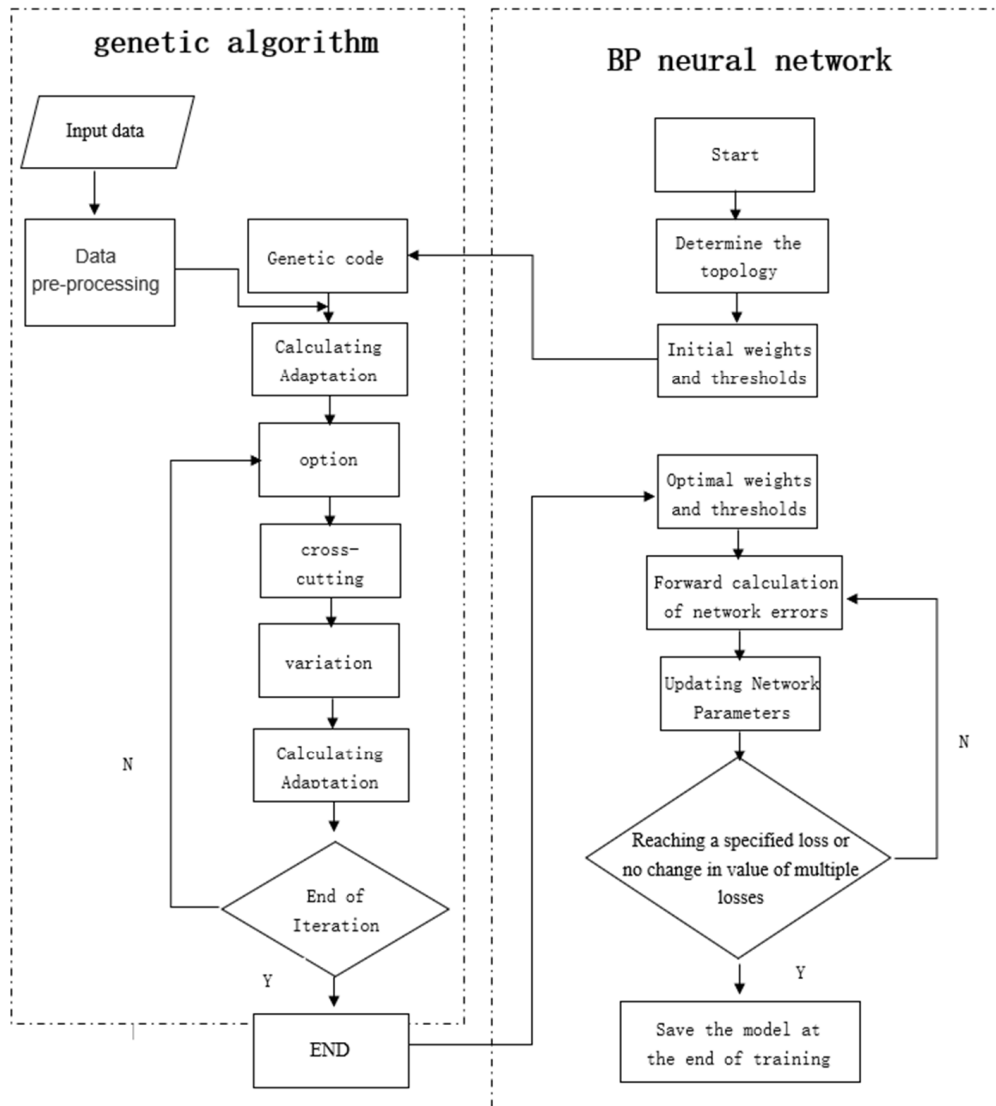


FIGURE 4. Flow chart of BP neural network optimized by genetic algorithm.

**D. MODEL EVALUATION METHOD**

The error analysis of the model prediction results uses MSE (Mean Square Error), RMSE (Root Mean Square Error), MAE (Mean Absolute Error), MAPE (Mean Absolute Percentage Error) with R2 (Coefficient of Determination) as the evaluation criterion to observe the difference between the two models. The specific calculation formula is as follows:

$$MSE = \frac{1}{n} \sum_{i=1}^k (y_i - y_{\hat{i}})^2 \tag{13}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^k (y_i - y_{\hat{i}})^2} \tag{14}$$

$$MAE = \frac{1}{n} \sum_{i=1}^k |y_i - y_{\hat{i}}| \tag{15}$$

$$MAPE = \frac{100\%}{n} \sum_{i=1}^k \left| \frac{y_i - y_{\hat{i}}}{y_i} \right| \tag{16}$$

$$R^2 = 1 - \frac{\sum_{i=1}^k (y_i - y_{\hat{i}})^2}{\sum_{i=1}^k (y_i - \bar{y})^2} \tag{17}$$

where k is the dataset size, y and y\_hati are the actual and predicted values, respectively, and n is the dataset size.

**III. RESULTS AND DISCUSSION**

**A. SELECTION OF MODEL HYPERPARAMETERS**

To avoid overfitting or underfitting due to the inappropriate selection of model hyperparameters, this paper adopts the tenfold cross-validation method to select hyperparameters for the model. The method of tenfold cross-validation is to

TABLE 2. Model training effect based on feature selection.

Feature	MSE	RMSE	MAE	MAPE/%	R2
X10, X7, X1, X3, X5	0.0569	0.2386	0.1914	0.1150	0.8867
X10, X7, X1, X3, X5, X8	0.0156	0.1250	0.0952	0.0514	0.9688
X10, X7, X1, X3, X5, X8, X6	0.0074	0.0860	0.0642	0.0378	0.9852
X10, X7, X1, X3, X5, X8, X6, X9	0.0081	0.0900	0.0667	0.0388	0.9839
X10, X7, X1, X3, X5, X8, X6, X9, X4	0.0070	0.0834	0.0547	0.0323	0.9862
X10, X7, X1, X3, X5, X8, X6, X9, X4, X12	0.0105	0.1025	0.0677	0.0398	0.9791
All	0.0096	0.0979	0.0705	0.0385	0.9808

divide the training set into 10 equal parts, select one part as the validation set in turn, set the remaining 9 training sets for network training, and obtain the average value and loss curve of the loss of the training set and the validation set after the completion of 10 times of training, which is used as a judgment criterion to determine whether the model is overfitting or underfitting [30], [31]. Finally, after various parameter combinations, the parameters of the BP neural network are as follows: weight decay of 3e-2, the learning rate of 0.0001, the number of training times is 1000. A small-batch stochastic gradient descent algorithm is used, and training is terminated when the number of training times is satisfied.

The individual coding of the genetic algorithm part takes the floating-point coding, and the optimization parameters are 381, including 360 weights and 21 thresholds. Other parameters in the genetic algorithm part of this paper are set: the maximum number of evolutionary generations is 100, the initial population size is 40, the maximum value of crossover probability is 0.7, the minimum value is 0.1, the maximum value of variation probability is 0.1, the minimum value is 0.01, and the number of comparison individuals in the tournament algorithm is 2.

When determining the model parameters, the loss curves of the two models' training set and validation set (Fig. 5) are used to determine whether the model is overfitting or underfitting. Finally, after the parameter adjustment, the model loss gradually decreases, and the overfitting is alleviated.

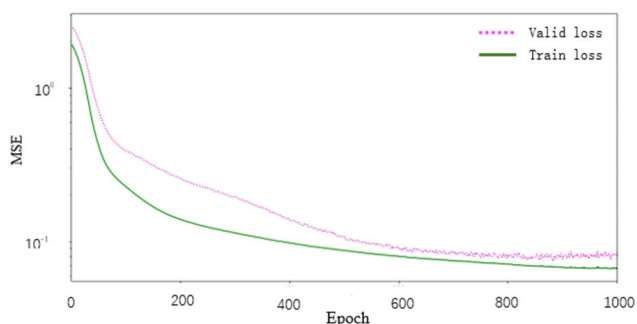


FIGURE 5. Loss of cross-validation of the GA-BP neural networks.

### B. FEATURE VARIABLE SELECTION FOR MODELS BASED ON FEATURE SELECTION

To verify the effect of feature selection on model effectiveness, this paper ranks the feature variables after the importance score of random forest variables, selects the top five features to input into the network, and increases the feature variables to compare the model performance under different numbers of features, as shown in Table 2. With the increase of the number of feature variables, the model's indexes have different degrees of improvement, reaching the optimum at 9 feature variables. To exclude the enhancement of the model by the number of features, the performance of the network with all 13 feature inputs is compared with that of the network with 9 features, and it can be obtained that the indexes under all features are lower than that of the network with 9 features after feature selection. It can be proved that the effectiveness of feature selection on model performance enhancement.

### C. ANALYSIS OF THE NEURAL NETWORK MODEL

The iterative evolution curve of the population-optimal average fitness of the GA-BP neural network model is shown in Fig 6. From the figure, It can be seen from the figure that the individual fitness value of the GA-BP neural network is large at the initial stage. The individual fitness in the population is continuously reduced by the selection of genetic algorithm and adaptive crossover mutation, and the model training error is gradually reduced. With the increase of evolution algebra, the new individuals generated by the genetic algorithm are always close to the global optimal solution, and the optimal average moderate value of the population gradually decreases and tends to be stable. After several times of moderate constancy, stop training and finally get the best individual moderation. This shows that the parameter settings of the genetic algorithm are reasonable and have good convergence. The weights and thresholds of the optimal fitting individuals of the GA-BP neural network are entered into the model for training to obtain two model training losses, as shown in Fig. 7. The final training losses of the BP neural network and the GA-BP neural network are 0.066 and 0.053 respectively. After the optimization of the genetic algorithm, the loss of the model is reduced by 5.08%, which shows that the random determination of weights and

TABLE 3. Error analysis of BP neural network and GA-BP neural network.

Model	MSE	RMSE	MAE	MAPE/%	R2
GA-BP	0.007	0.083	0.055	0.032	0.986
BP	0.067	0.259	0.188	0.098	0.867
xgboost	0.026	0.161	0.090	0.049	0.945
Lightgbm	0.048	0.219	0.171	0.219	0.904
RandomForest	0.024	0.155	0.080	0.039	0.952
SVM	0.054	0.234	0.141	0.069	0.891

thresholds will lead to the BP neural network easily falling into the local optimal solution or a wide range of flat areas, resulting in a large loss of BP neural network training. Ultimately, the model can not achieve high-precision prediction of new data, affecting the model’s prediction accuracy and generalization ability. Comparing the loss curves of the two models in the figure, it can be seen that the training effect of the GA-BP neural network and the convergence speed of the model have been significantly improved, which is substantially better than the neural network without genetic algorithm optimization therefore, the phytosanitary UAV flow rate decision model constructed with the neural network optimized by genetic algorithm effectively improves the model training performance and is suitable for this type of problem.

D. MODEL EVALUATION

In order to observe the performance of the model in this paper more intuitively, the test set prediction results of the model in this paper and several existing regression models were compared with the true values to derive the MSE, RMSE, MAE and MAPE of the prediction results of several models as well as the model comparison graphs as shown in Table 3, the RMSE and MAPE of the GA-BP neural network model are reduced to a certain extent compared to the several regression models. The generalization ability of the model is improved. From the test set, the determination coefficient R2 is increased by 0.119, and the model’s generalization ability is improved. It shows that the weights and thresholds obtained by the global search of the genetic algorithm can prevent the neural network from falling into a wide range of flat areas and local minimum areas, which makes the GA-BP neural network optimized by genetic algorithm better predict the flow of plant protection UAV than several regression models, and improves the prediction accuracy of the model.

To further analyze the model in this paper, the performance of the model is compared with that of a BP neural network, which is also a type of neural network. this paper compares the prediction results of the test set with the actual values. It obtains the prediction regression curves of the two models, as shown in Figure 8. From the data in Figure 8(a), it can be seen that the correlation coefficient  $r$  of the BP neural network model prediction data is 0.931, and the coefficient of determination R2 is 0.862, while from Figure 8(b), it can be seen that the correlation coefficient  $r$  of the GA-BP neural network prediction data is increased to 0.993. The coefficient of determination R2 is increased to 0.986. By comparing the regression curves of the prediction results of the two models, it can be seen that the BP neural network has the same performance as the test set due to the randomness of the weights and thresholds. While the optimized GA-BP neural network model predicts stable results, the fitting curve of the test set has a high degree of overlap with  $y=x$  with a slope of 1. It has a high degree of linearity and goodness of fit and can make accurate decisions about the flow of crop protection UAVs.

E. MODEL VALIDATION

In order to test the effectiveness of the model for new data, the model performance was verified using field experiments

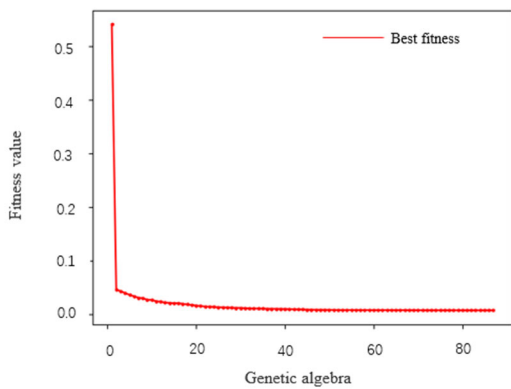


FIGURE 6. Iterative diagram of average fitness of GA-BP neural network.

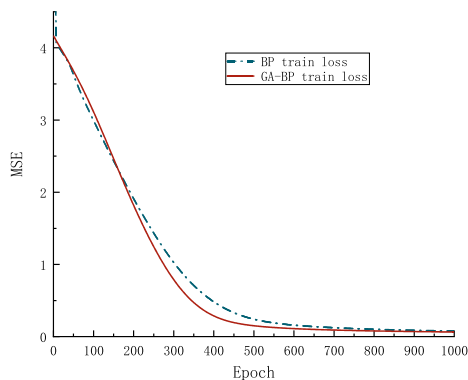


FIGURE 7. Train curve based on neural network model.



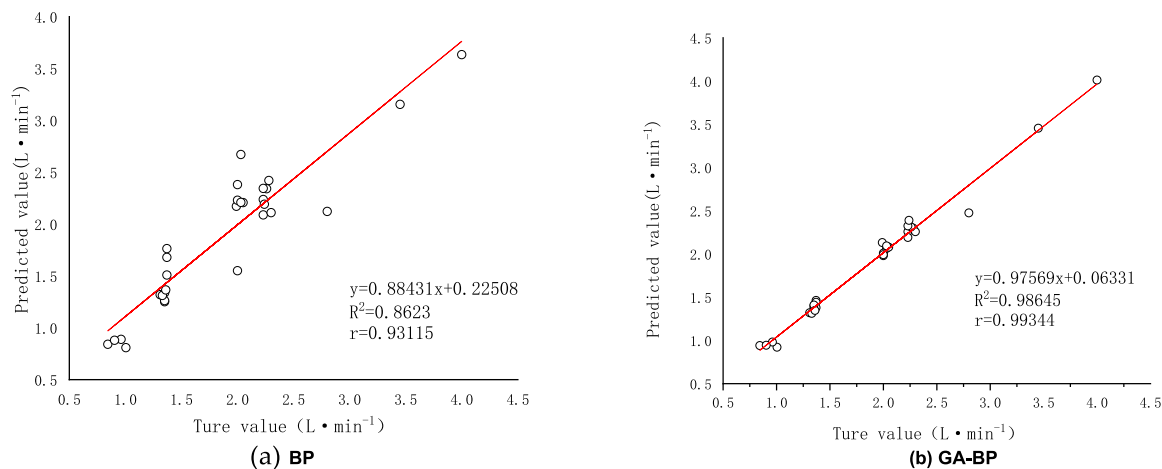


FIGURE 8. Regression curve of GA-BP neural network prediction value.

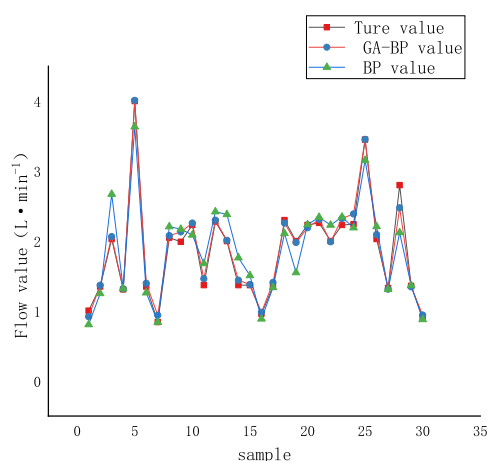


FIGURE 9. Model comparison diagram.

with a DJI T20 plant protection UAV at Shandong University of Technology, Zhangdian District, Zibo City, Shandong Province. The model validation results are shown in Table 4. The data in the table indicates that the bias of the BP neural network model for the flow rate is relatively large, The results show that the error between the real flow rate and the flow rate predicted by the GA-BP model in the new data obtained is within  $\pm 20\%$ . However, the BP neural network has a relatively large fluctuation in predicting some values, and the error in some fitting results exceeds 50 %, indicating that the model has good performance for new data and can be used to determine the required flow rate based on changes in factors such as wind speed and acreage usage. By comparing the two models, it is proved that the improved flow decision model meets the expected requirements when the expected deposition amount and other parameters are specified.

IV. DISCUSSION

Precision agriculture aviation technology allows for precise application of pesticides according to demand, while the

complexity of the farmland environment, the different needs of the plots and various factors can easily lead to a reduction in the efficiency of pesticide use. Therefore, this paper aims to study the factors affecting the deposition of droplets and establish a decision-making model with the flow rate. Many researchers have conducted research on this aspect in the past [32]. In the spray experiment conducted by Lan Yubin’s team at the South China Agricultural University with a multi-rotor drone and different orifice-size nozzles, it was found that with the increase in droplet size, droplet deposition rate and droplet penetration rate also increased [33]. Zhan et al. found that changes in UAV flight parameters can alter downwash airflow distribution in various directions, particularly when it impacts spray distribution under the UAV from the vertical direction [34], [35]. In addition, environmental factors are also important factors affecting the deposition effect of droplets [36], [37], [38]. Wang Juan conducted spray drift and deposition tests under various meteorological conditions. The experimental results showed that with the change in UAV operating altitude and wind speed, the maximum offset of the starting position of the droplet deposition zone in the extreme case was 4 m, and the percentage of total spray drift increased from 15.42% to 55.76% [39]. But the interaction between multiple factors leads to deviation from the expected deposition at a fixed flow rate, and the various types of features acquired and input into the model produce different effects. In this paper, the deposition volume is added to the dataset as a feature, and the feature selection method using the correlation and feature importance scores of random forests is used to select features for the collected data, and the feature-selected model is evaluated. In most of the existing variable spraying studies the flow decision component is based on prescription maps produced from remotely sensed imagery or visual recognition based flow decision making or spraying for ground application. In most of the existing variable spraying studies the flow decision component is based on prescription maps produced from remotely sensed imagery or

**TABLE 4. Validation of droplet deposition based on the flow results of two decision models.**

batch	Acre amount (L•hm-2)	Altitude (m)	Speed (m•s- 1)	Number of nozzles	Nozzle type	Temperature (°C)	Wind speed (m•s- 1)	Direction (°)	Flow rate (L•min- 1)	Actual deposition amount (μL/cm2)	BP decision flow (L•min- 1)	BP deviation (%)	GA-BP decision flow (L•min- 1)	GA-BP deviation (%)
1	1	3	3	8	SX110015vs	29.3	1.7	71	2.68	0.48	2.84	6.08	2.80	4.48
2	1.5	4	2	8	SX110015vs	28.8	2.7	22	2.46	0.50	2.38	-3.25	2.08	-15.51
3	2	5	1	4	SX110015vs	29.5	0.8	21	1.92	1.36	2.92	52.00	2.24	16.67
4	1	3	3	8	SX11001vs	24.3	3.8	83	2.74	0.17	3.10	13.19	2.72	-0.56
5	1.5	4	2	8	SX11001vs	24.6	2.4	75	2.50	0.21	2.32	-7.38	2.23	-10.83
6	2	5	1	4	SX11001vs	24.3	0.4	75	1.84	0.94	2.15	17.06	1.97	7.11
7	1	3	3	4	XR11002vs	28.7	3.7	71	2.67	0.27	1.49	-44.34	2.16	-18.93
8	1.5	4	2	4	XR11002vs	28.7	2.6	50	2.31	0.42	1.52	-34.27	1.86	-19.35
9	2	5	1	4	XR11002vs	25.1	0.7	72	1.91	0.96	1.97	3.18	1.87	-1.89

visual recognition based flow decision making or spraying for ground application. Hao, Z developed an Adaptive Spraying Decision System (ASDS) that recommended the minimum drone spraying volume and reasonable drone spraying speed, drone spraying height, and initial droplet size based on crop and environmental information, which minimised pesticide use by reducing the amount of spraying by 14% compared to using traditional parameters [40]. The study in this paper is similar to such studies, but takes into account the influence of environmental parameters during application, which are used as features combined with an improved neural network model to make decisions on the flow rate of plant protection UAVs, in anticipation of further reducing the error in the spraying process.

## V. CONCLUSION

This paper proposes a method to establish a multi-factor fusion flow decision-making model based on the feature selection of the GA-BP neural network. By conducting a large number of crop protection UAV spraying droplet deposition experiments, obtaining the droplet deposition data under different parameters, and after statistical analysis of the data and Pearson correlation analysis of the initial screening of the features of the Random Forest Importance Score finally identified nine types of feature variables. The correlation coefficient between the actual and predicted values of the test set after training of the established and improved GA-BP neural network model reaches 0.993, and the coefficient of determination reaches 0.986, which is significantly better than the existing prediction model in all aspects, with lower final loss and faster loss reduction, and the validation experiments prove the spraying accuracy of the GA-BP model. The results show that the error value between the actual flow rate and the predicted flow rate of the GA-BP model

is within 20%. In contrast, some of the expected values of the BP neural network fluctuated greatly. The error of some fitting results was more than 50%, and the validation results proved the feasibility and effectiveness of the BP neural network optimised by genetic algorithm in plant protection UAV flow decision-making. It can provide a reference for the research of precise variable spraying of plant protection UAV.

## REFERENCES

- [1] H. Chen, Y. Lan, B. K. Fritz, W. C. Hoffmann, and S. Liu, "Review of agricultural spraying technologies for plant protection using unmanned aerial vehicle (UAV)," *Int. J. Agricult. Biol. Eng.*, vol. 14, no. 1, pp. 38–49, 2021.
- [2] V. K. Tewari, C. M. Pareek, G. Lal, L. K. Dhruw, and N. Singh, "Image processing based real-time variable-rate chemical spraying system for disease control in paddy crop," *Artif. Intell. Agricult.*, vol. 4, pp. 21–30, 2020. [Online]. Available: <https://doi.org/10.1016/j.aiia.2020.01.002>
- [3] P. Chen, J. P. Douzals, Y. Lan, E. Cotteux, X. Delpuech, G. Pouxviel, and Y. Zhan, "Characteristics of unmanned aerial spraying systems and related spray drift: A review," *Frontiers Plant Sci.*, vol. 13, p. 2726, Aug. 2022.
- [4] L. Longlong, H. Xiongkui, S. Jianli, W. Xiaonan, J. Xiaoming, and L. Chaohui, "Design and experiment of automatic profiling orchard sprayer based on variable air volume and flow rate," *Trans. Chin. Soc. Agricult. Eng.*, vol. 33, no. 1, pp. 70–76, 2017.
- [5] R. A. Pierce and P. D. Ayers, "Evaluation of deposition and application accuracy of a pulse width modulation variable rate field sprayer," in *Proc. ASAE Annu. Meeting*, 2001, p. 1.
- [6] Z. Zhou, W. Xiu, and W. Jun, "Decision support system for variable pesticide spraying based on GIS," *Trans. CSAE*, vol. 24, no. 2, pp. 123–126, 2008.
- [7] J. Campos, M. Gallart, J. Llop, P. Ortega, R. Salcedo, and E. Gil, "On-farm evaluation of prescription map-based variable rate application of pesticides in vineyards," *Agronomy*, vol. 10, no. 1, p. 102, Jan. 2020.
- [8] P. Chen, Y. Lan, J.-P. Douzals, F. Ouyang, J. Wang, and W. Xu, "Droplet distribution of unmanned aerial vehicle under several spray volumes and canopy heights in the cotton canopy," *Int. J. Precis. Agricult. Aviation*, vol. 1, no. 1, pp. 74–79, 2018.
- [9] S. Ding, X. Xue, W. Qin, W. Gu, C. Cai, and L. Cui, "Influencing factors research and performance experiment on droplets deposition at low wind speed," *Int. J. Precis. Agricult. Aviation*, vol. 2, no. 1, pp. 46–51, 2019.

- [10] Y. Zhenzhen et al., "Prediction model of sugarcane yield in hot areas based on multi-source data and GA-BP," *J. Agricult. Machinery*, vol. 53, no. 10, pp. 277–283, 2022.
- [11] A. Azizpanah, A. Rajabipour, R. Alimardani, K. Kheiralipour, and V. Mohammadi, "Precision spray modeling using image processing and artificial neural network," *Agricult. Eng. Int., CIGR J.*, vol. 17, no. 2, 2015.
- [12] C. B. O. Zhu, "Prediction of winter wheat evapotranspiration based on BP neural networks," *Trans. Chin. Soc. Agricult. Eng.*, vol. 26, no. 4, 2010.
- [13] Z. Wang, F. Wang, and S. Su, "Solar irradiance short-term prediction model based on BP neural network," *Energy Procedia*, vol. 12, pp. 488–494, 2011, doi: [10.1016/j.egypro.2011.10.065](https://doi.org/10.1016/j.egypro.2011.10.065).
- [14] N. Lazarovitch, M. Poulton, A. Furman, and A. W. Warrick, "Water distribution under trickle irrigation predicted using artificial neural networks," *J. Eng. Math.*, vol. 64, no. 2, pp. 207–218, Jun. 2009.
- [15] D. Liu, C. Liu, Y. Tang, and C. Gong, "A GA-BP neural network regression model for predicting soil moisture in slope ecological protection," *Sustainability*, vol. 14, no. 3, p. 1386, Jan. 2022.
- [16] Y. Shi, C. Ren, Z. Yan, and J. Lai, "High spatial-temporal resolution estimation of ground-based global navigation satellite system interferometric reflectometry (GNSS-IR) soil moisture using the genetic algorithm back propagation (GA-BP) neural network," *ISPRS Int. J. Geo-Inf.*, vol. 10, no. 9, p. 623, Sep. 2021.
- [17] J. Gu, G. Yin, P. Huang, J. Guo, and L. Chen, "An improved back propagation neural network prediction model for subsurface drip irrigation system," *Comput. Electr. Eng.*, vol. 60, pp. 58–65, May 2017.
- [18] J. Hu, P. Xin, S. Zhang, H. Zhang, and D. He, "Model for tomato photosynthetic rate based on neural network with genetic algorithm," *Int. J. Agricult. Biol. Eng.*, vol. 12, no. 1, pp. 179–185, 2019.
- [19] W. Guobin, L. Xuan, and J. Andaloro, "Current status and prospects of precise sampling of pesticide droplets," *Trans. Chin. Soc. Agricult. Eng.*, vol. 37, no. 11, pp. 1–12, 2021.
- [20] S. Guo, W. Yao, T. Xu, H. Ma, M. Sun, C. Chen, and Y. Lan, "Assessing the application of spot spray in Nanguo pear orchards: Effect of nozzle type, spray volume rate and adjuvant," *Pest Manage. Sci.*, vol. 78, no. 8, pp. 3564–3575, Aug. 2022.
- [21] Q. Liu, Y. Lan, G. Wang, H. Zhu, and S. Chen, "The effects of crosswind speed on the droplet deposition distribution of thrush 510G aircraft," *Appl. Eng. Agricult.*, vol. 38, no. 4, pp. 659–667, 2022.
- [22] C. Shan, G. Wang, H. Wang, Y. Xie, H. Wang, S. Wang, S. Chen, and Y. Lan, "Effects of droplet size and spray, volume parameters on droplet deposition of wheat herbicide application by using UAV," *Int. J. Agricult. Biol. Eng.*, vol. 14, no. 1, pp. 74–81, 2021.
- [23] G. Wang, Y. Lan, H. Qi, P. Chen, A. Hewitt, and Y. Han, "Field evaluation of an unmanned aerial vehicle (UAV) sprayer: Effect of spray volume on deposition and the control of pests and disease in wheat," *Pest Manage. Sci.*, vol. 75, no. 6, pp. 1546–1555, Jun. 2019.
- [24] A. P. M. Ramos, L. P. Osco, D. E. G. Furuya, W. N. Gonçalves, D. C. Santana, L. P. R. Teodoro, C. A. da Silva Junior, G. F. Capristo-Silva, J. Li, F. H. R. Baio, J. Marcato Jr., P. E. Teodoro, and H. Pistori, "A random forest ranking approach to predict yield in maize with UAV-based vegetation spectral indices," *Comput. Electron. Agricult.*, vol. 178, Nov. 2020, Art. no. 105791.
- [25] P. Waldmann, "On the use of the Pearson correlation coefficient for model evaluation in genome-wide prediction," *Frontiers Genet.*, vol. 10, p. 899, Sep. 2019.
- [26] H. Fei, Z. Fan, C. Wang, N. Zhang, T. Wang, R. Chen, and T. Bai, "Cotton classification method at the county scale based on multi-features and random forest feature selection algorithm and classifier," *Remote Sens.*, vol. 14, no. 4, p. 829, Feb. 2022.
- [27] W. Huo, W. Li, Z. Zhang, C. Sun, F. Zhou, and G. Gong, "Performance prediction of proton-exchange membrane fuel cell based on convolutional neural network and random forest feature selection," *Energy Convers. Manage.*, vol. 243, Sep. 2021, Art. no. 114367.
- [28] S. S. Kumar and T. Shaikh, "Empirical evaluation of the performance of feature selection approaches on random forest," in *Proc. Int. Conf. Comput. Appl. (ICCA)*, Sep. 2017, pp. 227–231.
- [29] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, Oct. 1986.
- [30] L.-J. Qiu and C.-Z. He, "Cross validation model for neural network stability," *Comput. Eng. Appl.*, vol. 46, pp. 43–45, 2010.
- [31] D. M. Allen, "The relationship between variable selection and data augmentation and a method for prediction," *Technometrics*, vol. 16, no. 1, pp. 125–127, Feb. 1974.
- [32] Z. Lou, F. Xin, X. Han, Y. Lan, T. Duan, and W. Fu, "Effect of unmanned aerial vehicle flight height on droplet distribution, drift and control of cotton aphids and spider mites," *Agronomy*, vol. 8, no. 9, p. 187, Sep. 2018.
- [33] S. Chen, Y. Lan, Z. Zhou, F. Ouyang, G. Wang, X. Huang, X. Deng, and S. Cheng, "Effect of droplet size parameters on droplet deposition and drift of aerial spraying by using plant protection UAV," *Agronomy*, vol. 10, no. 2, p. 195, Feb. 2020.
- [34] Y. Zhan, P. Chen, W. Xu, S. Chen, Y. Han, Y. Lan, and G. Wang, "Influence of the downwash airflow distribution characteristics of a plant protection UAV on spray deposit distribution," *Biosystems Eng.*, vol. 216, pp. 32–45, Apr. 2022.
- [35] Y. Zhang, Y. Li, Y. He, F. Liu, H. Cen, and H. Fang, "Near ground platform development to simulate UAV aerial spraying and its spraying test under different conditions," *Comput. Electron. Agricult.*, vol. 148, pp. 8–18, 2018, doi: [10.1016/j.compag.2017.08.004](https://doi.org/10.1016/j.compag.2017.08.004).
- [36] H. Qi, Z. Lin, J. Zhou, J. Li, P. Chen, and F. Ouyang, "Effect of temperature and humidity on droplet deposition of unmanned agricultural aircraft system," *Int. J. Precis. Agricult. Aviation*, vol. 1, no. 1, pp. 41–49, 2018.
- [37] L. Yubin, S. Changfeng, and W. Qingyu, "Effects of different spray additives on droplet deposition characteristics during plant protection UAV spraying operations," *Trans. Chin. Soc. Agricult. Eng.*, vol. 37, no. 16, pp. 31–38, 2021.
- [38] W. Guobin, W. Shizhou, C. Pengchao, H. Xiaoqiang, S. Changfeng, C. Shengde, and L. Yubin, "Effect of spraying droplet size with drones on deposition, penetration, and cotton harvest-aid efficacy," *J. Plant Protection*, vol. 48, pp. 493–500, 2021, doi: [10.13802/j.cnki.zwbhxb.2021.2021837](https://doi.org/10.13802/j.cnki.zwbhxb.2021.2021837).
- [39] J. Wang, Y. Lan, H. Zhang, Y. Zhang, S. Wen, W. Yao, and J. Deng, "Drift and deposition of pesticide applied by UAV on pineapple plants under different meteorological conditions," *Int. J. Agricult. Biol. Eng.*, vol. 11, no. 6, pp. 5–12, 2018.
- [40] Z. Hao, X. Li, C. Meng, W. Yang, and M. Li, "Adaptive spraying decision system for plant protection unmanned aerial vehicle based on reinforcement learning," *Int. J. Agricult. Biol. Eng.*, vol. 15, no. 4, pp. 16–26, 2022.



**MENG WANG** was born in 1999. He is currently pursuing the M.S. degree in agricultural engineering with the Shandong University of Technology. His research interest includes precision agricultural aviation technology and equipment.



**ZHHAO BIAN** was born in 1998. He is currently pursuing the M.S. degree in agricultural engineering with the Shandong University of Technology. His research interests include plant protection UAV, mechatronics, and automation control.

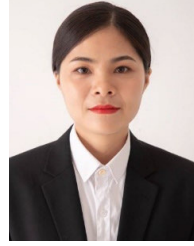


**YU YAN** was born in 1997. She received the B.S. degree from the Shandong University of Technology, Zibo, China, in 2021, where she is currently pursuing the M.S. degree in agricultural engineering. Her research interest includes intelligent spraying technology.



advancing UAV and precision agriculture aviation technology.

**MUJAHID HUSSAIN** was born in 1992. He is currently pursuing the Ph.D. degree with China Agricultural University. He is also a Postdoctoral Fellow with the College of Agricultural Engineering and Food Science, Shandong University of Technology, Zibo, China. His research interest includes precision agriculture aviation technology. He is mainly engaged in the research line for synthesizing plant growth regulators and nanomaterials and their application to various crops by



**CANCAN SONG** was born in Henan, China, in 1991. She received the Ph.D. degree, with a focus on precision agriculture aerial application technology and intelligent agricultural machinery.



**GUOBIN WANG** was born in Henan, China, in 1991. He received the Ph.D. degree, with a focus on precision agriculture aerial application technology and intelligent agricultural machinery.



**YUBIN LAN** was born in 1967. He received the B.S. degree in agricultural machinery design and manufacturing from Jilin University, Jilin, China, in 1982, and the Ph.D. degree in agricultural engineering from Texas A&M University, in 1994. He was a Professor with the Shandong University of Technology; a Foreign Academician with the European Academy of Sciences, Arts and Humanities, the Russian Academy of Natural Sciences, and the Georgian National Academy of Sciences; and the Director and the Lead Scientist of the National Center for International Collaboration Research on Precision Agricultural Aviation Pesticides Spraying Technology (NPAAC).

...