Prediction of spring and summer maize yield in China based on feature analysis and hybrid DHKELM algorithms

Long Zhao[1], Fei Wang[2], Zizhe Xu[2], Dan Meng[3*], Shunhao Qing[2], Shuangchen Chen[1], Xuguang Xing[4]

(1. *College of Horticulture and Plant Protection, Henan University of Science and Technology, Luoyang 471000, Henan, China*;
2. *College of Agricultural Equipment Engineering, Henan University of Science and Technology, Luoyang 471000, Henan, China*;
3. *Chinese Society of Agricultural Engineering, Beijing 100125, China*;
4. *Key Laboratory for Agricultural Soil and Water Engineering in Arid Area of Ministry of Education, Northwest A&F University, Yangling 712100, Shaanxi, China*)

Abstract: Crop yield prediction helps to enhance the stability of agricultural product supply and promote sustainable agricultural development, both of which are crucial for food production and security. To develop simple yet highly accurate crop yield prediction models, this study proposed a spring- and summer-maize yield prediction model based on the deep hybrid kernel extreme learning machine (DHKELM) algorithm. In this study, four tree-based feature importance analysis algorithms, including classification and regression tree, gradient boosting decision tree, random forest, and extreme gradient boosting algorithms, were utilized to analyze the importance of the factors affecting the yield of spring and summer maize. Then, based on the analysis of the four algorithms, different combinations of factors were established to obtain the optimal combination of features. Moreover, to improve the prediction accuracy of the machine learning model, this study utilized three optimization algorithms, including the bald eagle search algorithm, chaos game optimization (CGO) algorithm, and carnivorous plant algorithm, to optimize the hyperparameters in the DHKELM algorithm. The results of the study showed that planting density and plant height were important factors affecting maize yield, and net solar radiation (R_n) received during the reproductive period exhibited the highest relative importance. Appropriate feature combinations can effectively improve model prediction accuracy. The optimal feature combination for spring maize included planting density, plant height, R_n, mean temperature (T_{mean}), minimum temperature (T_{min}), and cumulative temperature, and the optimal feature combination for summer maize included R_n, plant height, planting density, T_{min}, and T_{mean}. Among the three optimization algorithms, the CGO algorithm exhibited the best optimization effect and could significantly improve the prediction accuracy of the DHKELM algorithm. When the optimal combination of features was used as input, the CGO–DHKELM model used for maize yield prediction provided the following values: RMSE=1.488 t/hm², R^2=0.862, MAE=1.051 t/hm², and NSE=0.852 for spring maize; RMSE=1.498 t/hm², R^2=0.892, MAE=1.055 t/hm², and NSE=0.891 for summer maize. Thus, the findings of the study provide a reference for high-precision prediction of spring and summer maize yields in China.

Keywords: yield, feature analysis algorithm, deep hybrid kernel extreme learning machine, maize, Chaos game optimization
DOI: 10.25165/j.ijabe.20251806.9749

Citation: Zhao L, Wang F, Xu Z Z, Meng D, Qing S H, Chen S C, et al. Prediction of spring and summer maize yield in China based on feature analysis and hybrid DHKELM algorithms. Int J Agric & Biol Eng, 2025; 18(6): 191–201.

1 Introduction

Crop yield prediction, which provides information on crop growth to farmers and related enterprises, as well as the government, is beneficial for the rational formulation of agricultural production policies as it can guarantee food safety and reduce economic losses caused by production risks[1]. Furthermore, accurate and efficient crop yield prediction is the key to sustainable food production and guaranteeing national food security and can provide a reliable reference for agricultural policy adjustment and scientific planning of agricultural production decisions[2]. Therefore, highly operational and accurate crop yield prediction models are necessary for the development of sustainable agriculture in China.

Crop yield prediction is a multivariate, nonlinear process, as crop yield is affected by a variety of factors such as climatic conditions, soil conditions, and planting management measures[3]. Traditional crop yield prediction builds on farmers' labor and expertise, a method that is inefficient and destructive[4]. In recent years, with the improvement in the efficiency of data extraction and the increase in the amount of available data, previously developed linear analysis methods have also failed to meet the requirements of complex yield prediction[5]. To address this issue, researchers have developed models for simulating crop growth processes based on their physiological characteristics and immediate environment, and these models have been widely used for crop yield prediction[6-8]. These crop growth models accurately describe crop growth and development in relation to the environment and management

Received date: 2025-02-25 **Accepted date:** 2025-08-31

Biographies: Long Zhao, PhD, Associate Professor, research interest: agricultural informatization, Smart agriculture, Email: hkdzhaolong@163.com; **Fei Wang**, Undergraduate, research interest: agricultural informatization. Email: wf01082021@163.com; **Zizhe Xu**, Undergraduate, research interest: agricultural informatization. Email: 2410812873@qq.com; **Shunhao Qing**, MS, research interest: smart agriculture, Email: 220320261805@stu.haust.edu.cn; **Shuangchen Chen**, PhD, Professor, research interest: agricultural informatization, smart agriculture, Email: chen_shuangchen@126.com; **Xuguang Xing**, PhD, Associate Professor, research interest: agricultural water and soil environment, Email: xgxing@nwsuaf.edu.cn.

***Corresponding author: Dan Meng**, MS, Engineer, research interest: agricultural informatization. Chinese Society of Agricultural Engineering, Beijing 100125, China. Email: mengd202502@163.com.

practices, but they are data-driven, with uncertainty in the input parameters, and need to be simulated under specific conditions[9].

In this regard, machine learning, a powerful data analysis technique, has exhibited better performance in dealing with the nonlinear relationship between crop yield and various independent variables[10]. Kheir et al.[11] stated that machine learning is superior to crop growth models as it adapts to diverse data and features, can accurately capture complex relationships in the data, and provides more accurate predictions. Neural network—a powerful tool used to solve complex problems in the field of machine learning—can learn relationship mapping from input to output by adjusting weights in the network and has strong adaptability and modeling ability in dealing with complex nonlinear problems[12]. Extreme learning machine (ELM) is a novel single hidden-layer feedforward neural network algorithm proposed by Huang et al.[13]. ELM combines the advantages of machine learning and neural networks, with its primary feature being the fact that its input weights are randomly selected, which gives ELM algorithms improved training speed and generalization ability[13] compared to back propagation (BP) neural networks, support vector machines (SVM), and other state-of-the-art algorithms[14,15].

To further improve the performance of the ELM model, researchers have combined the ELM algorithm with deep learning ideas and kernel functions to obtain two models with better performance, i.e., deep extreme learning machine (DELM) and kernel extreme learning machine (KELM). The DELM algorithm uses ELM–Autoencoder (ELM-AE) as the basic learning unit, to train the data layer by layer, so that the model exhibits stronger fitting ability and adaptability, and also prevents model overfitting[16]. In contrast, KELM is an ELM algorithm based on the kernel function that uses kernel mapping instead of random mapping, which improves the learning speed, generalization ability, and prediction performance of the ELM model[17]. The choice of kernel function plays a crucial role in KELM, which is directly related to the performance of the model[18]. However, a single kernel function may face difficulty adapting to complex changes in the data. To overcome this limitation, researchers have proposed the hybrid kernel extreme learning machine (HKELM), which aims to improve the prediction accuracy and generalization ability of the model by combining multiple kernel functions. The hybrid kernel function of HKELM enables the model to better capture complex fluctuations in the data, thus demonstrating stronger learning ability and better generalization performance in prediction tasks[19]. In this study, the DELM and HKELM algorithms were combined to constitute the deep hybrid kernel extreme learning machine (DHKELM) algorithm for maize yield prediction, which is an HKELM algorithm with ELM–AE as the basic training unit that can further improve the performance of the model while retaining the advantages of the DELM and HKELM algorithms.

The performance of a machine learning model heavily depends on hyperparameter settings, and manual parameter tuning is a time-consuming and laborious process. To address this issue, researchers have combined machine learning algorithms with optimization algorithms, which systematically determine the optimal parameter configuration to minimize the loss function and achieve more accurate predictions[20,21]. The bald eagle search (BES) algorithm simulates the hunting strategy of bald eagles when searching for prey, which is characterized by high search efficiency, finding the optimal solution in a short time, and is suitable for dealing with a variety of complex optimization problems[22]. Huang[23] used the BES algorithm to optimize the hyperparameters in the SVM model, and the results verified that the optimized model exhibited good prediction performance and generalization ability. The theme of another algorithm—chaos game optimization (CGO) algorithm—is based on a few principles of the chaos theory, providing the algorithm with high convergence speed and the ability to not get easily affected by the local optimal solution and quickly obtain a better solution, making it an optimization algorithm with strong search ability and adaptability[24]. He et al.[25] combined the CGO and multi-output least squares support vector regression machine (MLSSVR) algorithm and found that the CGO–MLSSVR prediction model was able to make effective predictions with high accuracy. The carnivorous plant algorithm (CPA) is a meta-heuristic algorithm inspired by the predation process of carnivorous plants on their prey, which avoids falling into local optimal solutions due to the diversity of their population and exhibits a greater advantage in solving global optimization problems[26]. Wang et al.[27] used the CPA algorithm to optimize the BP neural network and showed that CPA effectively reduced the prediction error of the BP neural network. Therefore, in this study, BES, CGO, and CPA were selected to optimize the hyperparameters in the DHKELM model and construct a high-precision maize yield prediction model.

In constructing a high-precision yield prediction model, the analysis of input factor importance can reduce data dimensions and improve model estimation efficiency, which can lead to the optimization of model accuracy[28,29]. Tree-based models can select the most influential features based on the intrinsic structure of the data, can accurately capture complex relationships, and are less prone to overfitting[30]. Common tree-based models include classification and regression tree (CART), gradient boosting decision tree (GBDT), random forest (RF), and extreme gradient boosting (XGBoost) models. Peng et al.[31] used Pearson's correlation coefficient, least absolute shrinkage and selection operator, and GBDT algorithms to determine the characteristic variables to estimate soil nutrient content, and the results of the study verified that the selection of the characteristic variables was the key to estimating the soil nutrient content with high accuracy, and the GBDT algorithm provided accurate information about the characteristic variables. Mohammadi and Mehdizadeh[32] utilized relief, RF, principal component analysis, and Pearson's correlation methods to preprocess data and construct a support vector regression-based daily reference evapotranspiration prediction model, and showed that input variables identified by the RF method produced more accurate results. Zheng et al.[33] analyzed the influence coefficients of soil parameters and other factors on soybean yield using a general linear model and CART algorithm and demonstrated that the prediction results of the CART model exhibited low error rates. Gill et al.[34] suggested that the XGBoost model exhibited superior predictive performance and provided a highly accurate ranking of factor importance. Thus, in this study, CART, GBDT, RF, and XGBoost algorithms were selected to analyze the importance of the characteristic factors affecting the yield of spring and summer maize. The complementary advantages of multiple models can be used to obtain more complete and accurate results of factor importance analysis and determine the optimal input feature combinations.

To obtain an operable maize-yield prediction model, this study first utilized the CART, GBDT, RF, and XGBoost algorithms to rank the importance of multiple factors affecting the yield of spring and summer maize, and then combined them into different combinations of input features based on the results of the importance ranking, for model construction. In the process of model construction, this study used the DHKELM model as the base model, and applied three optimization algorithms (BES, CGO, and CPA) to tune the hyperparameters of the model.

The purpose of this study was as follows: 1) To rank the importance of factors affecting the yield of spring and summer maize using four tree-based factor analysis methods (CART, GBDT, RF, and XGBoost) and analyze the results. 2) To determine the optimal combination of input factors affecting the yield of spring and summer maize according to the results of the importance analysis. 3) To construct a hybrid model using BES, CGO, and CPA optimization algorithms to adjust and optimize the parameters in the DHKELM model, and to analyze the comprehensive prediction performance of the independent and the hybrid models.

2 Materials and methods

2.1 Study area

Data for this study were obtained from the National Ecological Science Data Center, China, and included meteorological, phenological, and yield data of spring and summer maize. The different types of sample plots set up within each site exhibited different soil and crop management conditions, which led to differences in phenotypic characteristics of the crop. The sites exhibited different meteorological conditions owing to their geographic locations, and these differences resulted in a greater diversity in maize yield data. The data used in this study included planting density and plant height, as well as net solar radiation (R_n), mean temperature (T_{mean}), maximum temperature ($T_{\max}$), minimum temperature ($T_{\min}$), cumulative temperature (T_a), relative humidity (R_H), wind speed (U), and precipitation (Pre) during the reproductive period.

In the modeling process, the spring maize dataset used in this study contained 865 samples, and the summer maize dataset contained 921 samples. Among them, 80% of the input data was used as the training set, and 20% of the input data was used as the test set. The coordinates of the spring and summer maize study sites, along with the annual means of the meteorological data, are listed in Table 1.

Table 1 Geographic location and annual mean values of meteorological data of the spring and summer maize study sites

Station	Longitude (E)/(°)	Latitude (N)/(°)	Elevation/ m	R_n/ MJ·m^{-2}	T_{mean}/ °C	R_H	U/ m·s^{-1}	Pre/ mm
Ansai	109.19	36.53	1068.3	1320.85	19.18	64.53	0.87	419.56
Changwu	107.48	35.12	1206.5	1383.21	19.02	70.19	1.25	363.94
Fengqiu	114.25	35.02	69.6	1043.85	24.94	77.28	1.10	306.77
Hailun	126.58	47.27	239.2	1329.36	17.88	69.25	2.34	422.02
Huanjiang	108.32	24.73	208.5	919.46	22.57	79.94	0.66	482.64
Linze	100.10	39.09	1453.7	1229.65	19.97	40.79	1.87	102.37
Luancheng	114.38	37.53	52.9	1068.79	24.24	71.44	1.46	329.25
Naiman	120.39	42.51	362.9	1494.53	20.26	43.78	1.72	249.03
Shapotou	104.95	37.45	1225.7	1261.60	20.87	45.06	2.38	150.85
Yanting	105.45	31.27	421.3	935.93	25.21	78.61	0.50	502.77
Yucheng	116.34	36.56	23.6	1114.19	24.89	76.15	1.40	300.31

2.2 Tree-based algorithms for factor importance analysis

2.2.1 Classification and regression tree

CART is an algorithm based on decision trees, where variance minimization methods are used to obtain increasingly homogeneous subsets by recursively partitioning the dataset into subsets[33,35]. The CART algorithm needs to consider the degree of contribution of each feature in the construction of the decision tree when performing the importance analysis. The algorithm usually uses node purity to measure the relative importance of each feature. The information gain of each feature is an indicator of the degree of node purity improvement.

2.2.2 Random forest

RF is a nonparametric estimation algorithm based on decision tree integration, which provides more accurate and comprehensive importance assessment results. The Gini index is the primary basis used for measuring the importance of each feature in random forests. The relative importance of features is assessed by comparing the reduction in the average Gini index for different features[30].

2.2.3 Gradient boosting decision tree

GBDT is an iterative decision tree algorithm that consists of multiple decision tree models[36]. When the frequency of segmentation of a feature in constructing decision tree models is high, the information gain is greater and the feature exhibits higher importance. The GBDT model takes the average of the importance of the features in each tree as the final importance result.

2.2.4 Extreme gradient boosting

XGBoost is an efficient gradient-boosting algorithm that focuses on optimizing the objective function and evaluating the importance of each feature by calculating its gain in a decision tree split[37]. If a feature is used multiple times for key decision points, its importance score increases accordingly. The XGBoost algorithm takes the weighted average of the results of the feature across all decision trees as its result.

2.3 Machine learning algorithms

The DELM algorithm uses ELM–AE as the basic unit and combines the idea of deep learning to form a deep network structure, which can improve the learning ability and prediction performance of the model. ELM–AE can orthogonalize the randomly generated weights and biases and map the input data as it achieves feature extraction for different requirements, and can also effectively reduce the noise in the data to enhance the generalization ability of the model. The DELM uses ELM–AE to initialize the weights of the hidden layer, and the layer-by-layer training can get more comprehensive feature information. However, unlike traditional deep learning models, the DELM model does not need to fine-tune the parameters and exhibits an increased training speed.

KELM is a neural network algorithm obtained by combining the kernel function and ELM. The kernel function exhibits excellent nonlinear mapping ability, which enhances the divisibility of data in high-dimensional space and effectively improves the learning ability and generalization ability of ELM. The KELM algorithm utilizes kernel mapping instead of random mapping in ELM to map low-dimensional space data to high-dimensional space through kernel function, which further improves the generalization ability and prediction performance while retaining the advantages of ELM[38,39].

The predictive performance of the KELM model is greatly affected by the type of kernel function, and a single kernel function is relatively weak for large-scale and multi-featured datasets. Thus, to further improve the performance of the KELM model, scholars proposed the HKELM algorithm, which uses a hybrid kernel function obtained from the combination of multiple kernel functions instead of a single kernel function of the KELM, effectively improving the overall performance of the model. Commonly used kernel functions include radial basis function kernel function, polynomial kernel function, and wavelet kernel function. The polynomial kernel function shows good global feature capture ability and strong generalization ability[40], whereas the wavelet kernel function is good at local feature capture. The combination of these two kernel functions is used to obtain more comprehensive data information and improve the learning ability of the model.

The DHKELM algorithm used in this study is the HKELM algorithm with ELM–AE as the basic unit, which has the feature extraction capability of deep learning as well as the powerful mapping capability of kernel functions, enabling the model to deal with complex nonlinear problems more efficiently[41-43]. The structure of the DHKELM model is shown in Figure 1.

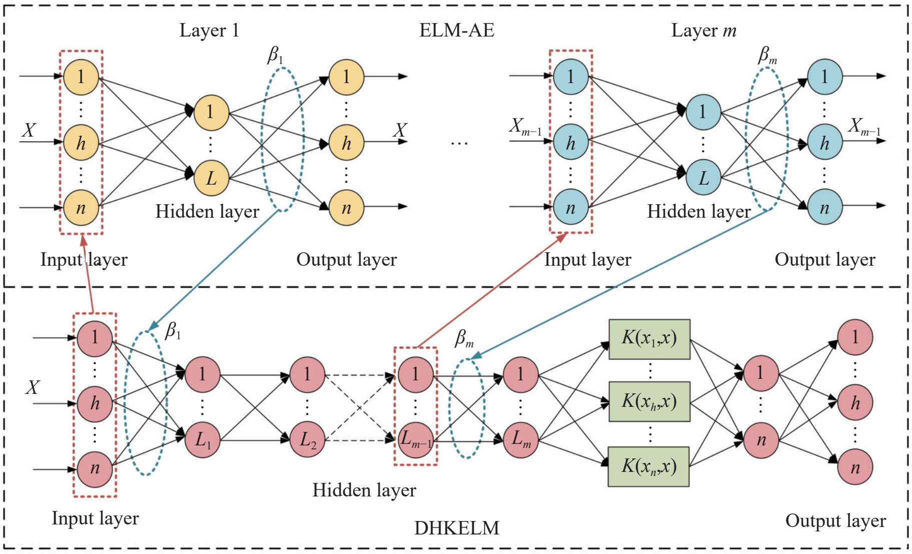

Figure 1 Structure of the DHKELM algorithm

2.4 Meta-heuristic optimization algorithms

2.4.1 Bald eagle search algorithm

BES algorithm, a meta-heuristic optimization algorithm with strong search capability, simulates the hunting strategy of the bald eagle in searching for prey and searches for the optimal solution through iterative search in the hunting process[22]. The optimization search of the BES algorithm is divided into three phases, namely selection, searching, and swooping. In the selection phase, the bald eagle selects a search space based on individual fitness value to facilitate the search for prey. The bald eagle then searches for prey within the selected search space and flies in a spiral form to speed up the search process and determine the optimal swooping location. In the swooping phase, the bald eagle flies from the optimal location in the search space to the target in a fast swoop, and the other individuals also move toward the optimal location and attack the prey.

2.4.2 Chaos game optimization algorithm

CGO algorithm is a meta-heuristic optimization algorithm based on the principles of the chaos theory[24]. This optimization method approaches the problem by considering candidate solutions that are also embodiments of the eligible seeds in the Sierpinski triangle. In the mathematical model of CGO, the Sierpinski triangle is the search space for candidate solutions. During the iterative process, to create a new eligible seed in the search space, the eligible seed in the search space constructs temporary triangles based on three parameters, namely, the position of the ith candidate solution as the selected seed, the position of the mean group, and the position of the so-far found global best, after which a new eligible seed is created.

2.4.3 Carnivorous plant algorithm

CPA simulates the process whereby carnivorous plants adapt to survive in harsh conditions[26]. First, the carnivorous plant population is initialized, and the individuals are sorted according to their fitness values in ascending order. Then, the individuals in the population are classified into carnivorous plants and prey, and grouped. During reproduction, only the optimal group in the population is allowed to reproduce, and the newly generated populations of carnivorous plants and prey will combine with the previous populations to form new populations, thereby repeating the process of categorical grouping, growth, and reproduction until the termination conditions are met. The principles of the three optimization algorithms are shown in Figure 2.

The implementation of the algorithm in this study was performed using MATLAB, while the graphics were drawn using Origin.

2.5 Model evaluation indices

In this study, root mean square error (RMSE), coefficient of determination (R^2), mean absolute error (MAE), and Nash–Sutcliffe efficiency (NSE) were used to assess the accuracy of model prediction results, and global evaluation index (GPI) was used for a comprehensive assessment of model performance. The formulas for the indicators are shown below:

$$R^2 = 1 - \frac{\left[\sum_{i=1}^{n}\left(P_i - \bar{P}\right)\left(T_i - \bar{T}\right)\right]^2}{\sum_{i=1}^{n}\left(P_i - \bar{P}\right)^2 \sum_{i=1}^{n}\left(T_i - \bar{T}\right)^2} \tag{1}$$

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(T_i - P_i)^2}{n}} \tag{2}$$

$$\mathrm{MAE} = \frac{\sum_{i-1}^{n}|T_i - P_i|}{n} \tag{3}$$

$$\mathrm{NSE} = 1 - \frac{\sum_{i=1}^{n}(T_i - P_i)^2}{\sum_{i=1}^{n}\left(T_i - \bar{T}\right)^2} \tag{4}$$

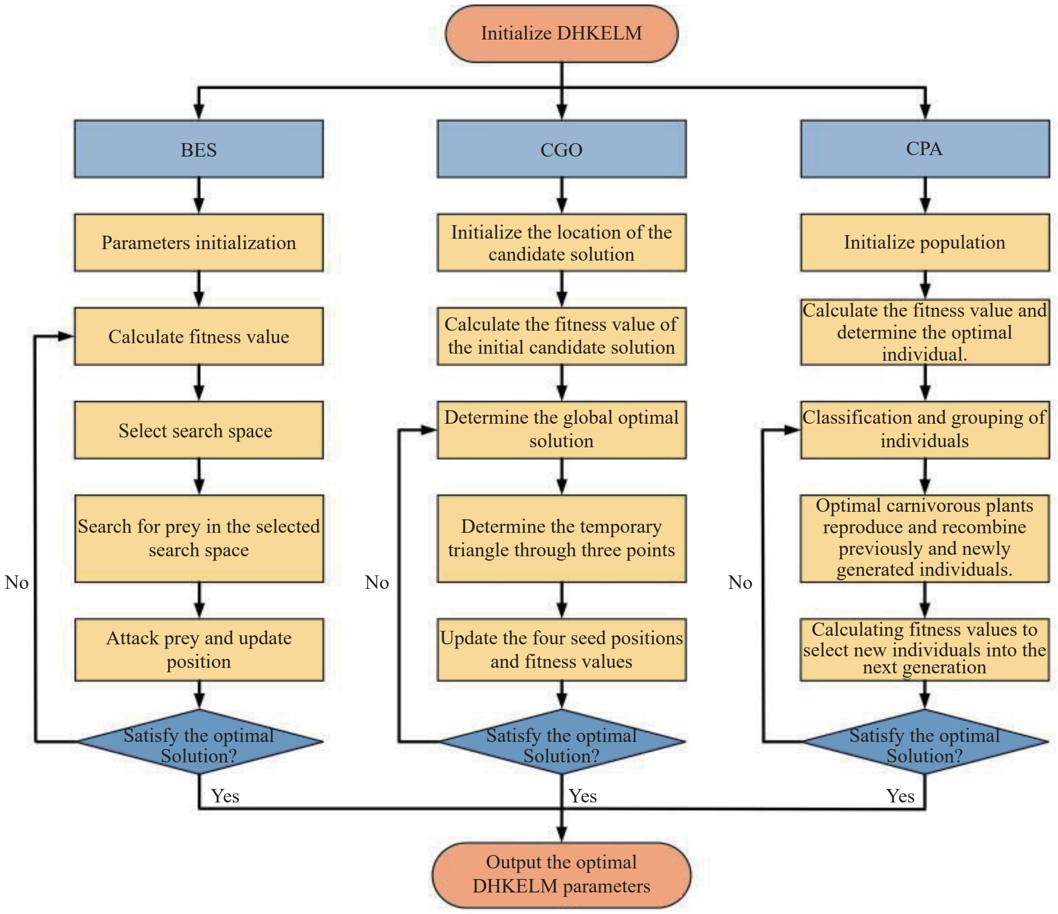

Figure 2 The principles of the BES, CGO, and CPA algorithms

$$\mathrm{GPI} = \sum_{j=1}^{4} \alpha_j \left(g_j - y_j\right) \tag{5}$$

where, T_i denotes the true value of yield; P_i denotes the predicted value of yield; $\bar{T}_i$ denotes the average of the true value of yield; and $\bar{P}_i$ denotes the average of the predicted value of yield. Additionally, g_j is the normalized value of RMSE; R^2, MAE, and NSE, and y_j is the median of each parameter. a_j was −1 for RMSE and MAE, and 1 for all other cases.

The larger the R^2 and NSE values, the smaller the RMSE and MAE values, and the better the predictive performance of the model. Furthermore, models with higher GPI values or higher rankings exhibited better overall performance.

3 Results and discussion

3.1 Importance analysis of factors affecting the yield of spring and summer maize

In this study, four tree-based feature analysis models, i.e., CART, GBDT, RF, and XGBoost algorithms, were used to estimate the relative importance of the factors affecting the yield of spring and summer maize. This study used the median percentage of the analysis results for each factor as the final relative importance coefficient. The median better reflects the central tendency of the data and avoids the influence of extreme values on the results. The specific results are shown in Figure 3. Regarding the importance analysis of factors affecting the yield of spring maize, planting density was the most important influencing factor, with a relative importance coefficient of 36.276%, followed by plant height and R_n, with relative importance coefficients of 19.469% and 16.486%, respectively. T_{mean}, T_{min}, and T_a were ranked fourth to sixth with relative importance coefficients of 7.259%, 5.913%, and 4.882%, respectively. These were followed by R_H and U, both with relative importance coefficients of 3.107%, and Pre and T_{max} in that order (with relative importance coefficients of 1.874% and 1.627%, respectively). However, the importance analysis of factors affecting the yield of summer maize indicated that R_n was the most important factor, with a relative importance coefficient of 23.612%, and plant height (23.167%) exhibited a comparable relative importance to R_n. Planting density ranked third with a relative importance coefficient of 16.040%, followed by T_{min} and T_{mean} with relative importance coefficients of 11.004% and 10.248%, respectively. T_a, U, and Pre were sixth to eighth with relative importance coefficients of 6.208%, 4.743%, and 2.050%, respectively. Additionally, R_H and T_{max} exhibited the lowest relative importance coefficients of 1.523% and 1.405%, respectively.

These results suggested differences in the results of factor importance analysis between spring and summer maize, which may be related to the immediate environment of the crop. Spring maize may be more dependent on early planting density and growth conditions, whereas summer maize may be more influenced by solar radiation. Planting density and plant height are important factors affecting the yield of both spring and summer maize and are key parameters for yield prediction[44,45]. Proper plant height and planting density help improve photosynthetic efficiency in maize so that maize plants can grow and develop better, thereby increasing maize yield. Solar radiation is a key energy source for photosynthesis in crops[46,47]. The results of factor analysis for both spring and summer maize demonstrated that the most important meteorological factor was R_n, which is the difference between the amount of solar radiation reaching the ground and the amount of radiation emitted from the ground. This suggests that R_n was a crucial determinant of crop growth and yield and was one of the important parameters for crop yield prediction.

Nonetheless, R_n exhibited different importance coefficients for spring and summer maize, probably because of the differences in the fertility periods. Spring maize is mostly productive in the spring when solar radiation is relatively weak, but as the season changes and the temperature gradually rises, the growth of the crop gradually accelerates. Spring maize may require additional management measures to promote photosynthesis in the early stages to ensure adequate growth even under reduced solar radiation conditions. In contrast, the effect of R_n on the yield of summer maize was particularly significant during the fertility period owing to the higher intensity of solar radiation. Adequate solar radiation can promote photosynthesis and increase biomass accumulation in the crop, thus improving yield. Therefore, monitoring and utilizing solar radiation for the cultivation and management of summer maize is of great significance in improving yield[48]. Moreover, T_{mean}[49], T_{min}[40], and T_a[50] during the fertility period were important meteorological factors affecting maize yield. In addition, R_H, Pre, U, and T_{max}, although less important compared to other factors, contributed adequately to maize yield.

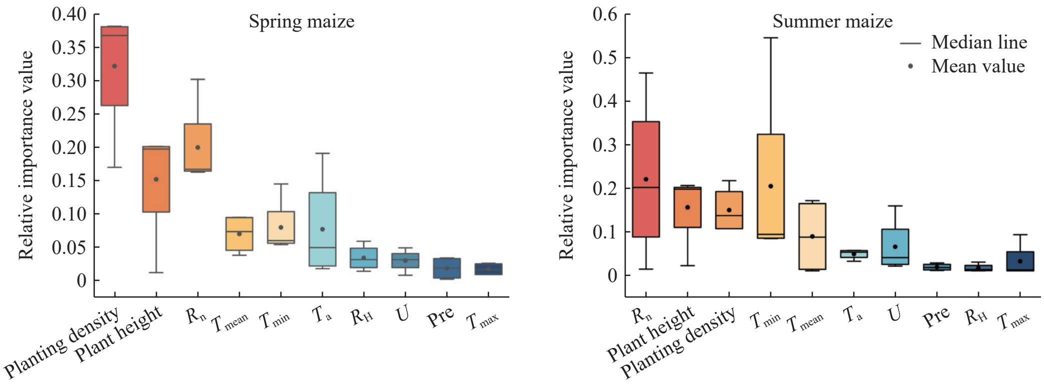

Figure 3 Box line plots of the results of the relative importance analysis of factors affecting the yield of spring and summer maize

To select appropriate combinations of input factors and construct yield prediction models with high operability for spring and summer maize, this study combined input factors according to the relative importance coefficients of each factor. The combinations were bound by the sum of relative importance coefficients of 70%, 80%, 90%, and 100%. The input combinations for spring and summer maize are listed in Table 2. Spm1 reflects the input combination with the sum of the importance coefficients of the factors influencing spring maize yield exceeding 70%, and Sum1 is the input combination with the sum of the importance coefficients of the factors influencing summer maize yield exceeding 70%.

Table 2 Combination of factors affecting the yield of spring and summer maize

Input factor combination	Input variables
Spm1	Planting density, Plant height, R_n
Spm2	Planting density, Plant height, R_n, T_{mean}, T_{min}
Spm3	Planting density, Plant height, R_n, T_{mean}, T_{min}, T_a
Spm4	Planting density, Plant height, R_n, T_{mean}, T_{min}, T_a, R_H, U, Pre, T_{max}
Sum1	R_n, Plant height, Planting density, T_{min}
Sum2	R_n, Plant height, Planting density, T_{min}, T_{mean}
Sum3	R_n, Plant height, Planting density, T_{min}, T_{mean}, T_a
Sum4	R_n, Plant height, Planting density, T_{min}, T_{mean}, T_a, U, Pre, R_H, T_{max}

3.2 Analysis of optimal factor combinations for the models predicting spring and summer maize yield

To obtain the optimal combination of input factors, this study used the DHKELM and hybrid models to predict the yield of spring and summer maize. Different combinations of inputs can result in the same machine learning model generating different prediction results. Figure 4 demonstrates the changes in the accuracy of the prediction results of the models predicting the yield of spring and summer maize under different combinations of inputs. The specific accuracy indices of the prediction results are listed in Tables 3 and 4. From the results, it is noted that the prediction accuracy of the model for spring and summer maize yield increased gradually with the increase in the number of input factors. When a certain number of input factors was reached, the prediction accuracy peaked. Thereafter, the predictive accuracy of the model did not improve significantly even if the input factors continued to be added. Thus, the input combination that resulted in the highest accuracy of the model was the optimal combination of inputs for predicting spring and summer maize yield.

As shown in Figure 4 and Table 3, the prediction accuracy of the model displayed a significant upward trend during the input combination from Spm1 to Spm3. When Spm1 was used as the input combination, the accuracy of the model prediction results was low, and the values of the evaluation indices were as follows: RMSE=2.067±0.400 t/hm^2, R^2=0.715±0.108, MAE=1.579±0.360 t/hm^2, and NSE=0.708±0.101. When the input combination of the model was extended from Spm1 to Spm2, the prediction accuracy of the model improved, and the values of the evaluation indices were as follows: RMSE=1.975±0.403 t/hm^2, R^2=0.775±0.078, MAE=1.459±0.308 t/hm^2, and NSE=0.771±0.076. The input combination Spm3 was the optimal input combination, as it performed best in terms of reduced prediction error and enhanced model prediction performance. The model's prediction accuracy was at its highest for Spm3, and the values of the evaluation indices were as follows: RMSE=1.640±0.152 t/hm^2, R^2=0.837±0.025, MAE=1.206±0.155 t/hm^2, and NSE=0.832±0.023. When the input combination was changed to Spm4, the prediction accuracy of the model did not improve, and the values of the evaluation indices were as follows: RMSE=1.844±0.254 t/hm^2, R^2=0.804±0.038, MAE=1.329±0.186 t/hm^2, and NSE=0.802±0.039. Similarly, Figure 4 and Table 4 show that the input combinations for summer maize exhibit similar trends to those for spring maize, with Sum2 being the optimal input combination.

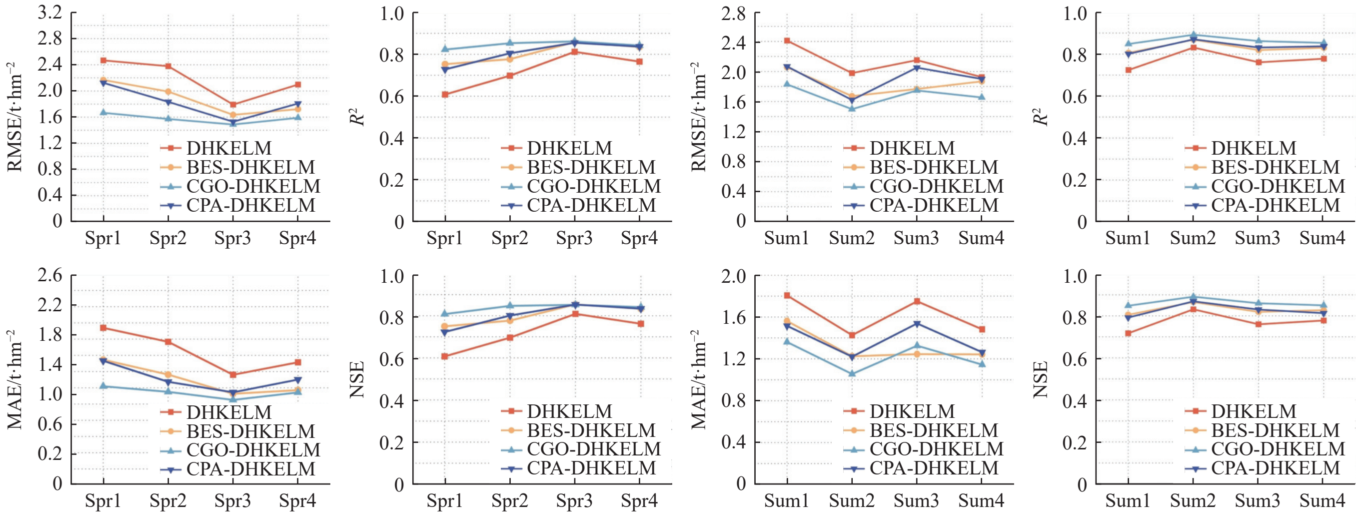

Figure 4 Changes in the precision of prediction results of spring and summer maize yield

Table 3 Values of evaluation indices for model prediction results under different combinations of inputs for spring maize

Model	Input combination	Training				Testing			
		RMSE/t·hm^{-2}	R^2	MAE/t·hm^{-2}	NSE	RMSE/t·hm^{-2}	R^2	MAE/t·hm^{-2}	NSE
Spm1	DHKELM	2.524	0.623	1.905	0.622	2.467	0.608	1.939	0.607
	BES–DHKELM	1.541	0.852	1.066	0.851	2.168	0.753	1.546	0.750
	CGO–DHKELM	1.434	0.880	0.999	0.879	1.666	0.823	1.219	0.808
	CPA–DHKELM	1.933	0.776	1.424	0.776	2.124	0.728	1.533	0.723
Spm2	DHKELM	2.003	0.751	1.463	0.750	2.378	0.698	1.767	0.696
	BES–DHKELM	1.328	0.892	0.974	0.892	1.990	0.777	1.363	0.777
	CGO–DHKELM	1.354	0.890	0.978	0.890	1.572	0.853	1.151	0.847
	CPA–DHKELM	1.557	0.854	1.147	0.853	1.834	0.805	1.274	0.801
Spm3	DHKELM	1.655	0.835	1.236	0.834	1.792	0.812	1.361	0.809
	BES–DHKELM	1.203	0.911	0.858	0.910	1.635	0.858	1.127	0.854
	CGO–DHKELM	1.119	0.926	0.771	0.926	1.488	0.862	1.051	0.852
	CPA–DHKELM	1.387	0.885	1.021	0.885	1.530	0.855	1.145	0.853
Spm4	DHKELM	1.781	0.805	1.308	0.803	2.098	0.765	1.515	0.762
	BES–DHKELM	1.213	0.910	0.891	0.909	1.722	0.835	1.172	0.833
	CGO–DHKELM	1.206	0.913	0.862	0.913	1.590	0.842	1.143	0.841
	CPA–DHKELM	1.488	0.859	1.107	0.859	1.808	0.837	1.302	0.834

Table 4 Values of evaluation indices for model prediction results under different combinations of inputs for summer maize

Model	Input combination	Training				Testing			
		RMSE/t·hm^{-2}	R^2	MAE/t·hm^{-2}	NSE	RMSE/t·hm^{-2}	R^2	MAE/t·hm^{-2}	NSE
Sum1	DHKELM	2.149	0.750	1.653	0.748	2.411	0.724	1.806	0.717
	BES–DHKELM	1.545	0.869	1.188	0.868	2.056	0.806	1.561	0.805
	CGO–DHKELM	1.573	0.862	1.159	0.862	1.829	0.848	1.361	0.848
	CPA–DHKELM	1.933	0.806	1.463	0.797	2.067	0.801	1.515	0.792
Sum2	DHKELM	1.499	0.874	1.101	0.874	1.978	0.831	1.426	0.831
	BES–DHKELM	1.182	0.924	0.795	0.924	1.672	0.868	1.225	0.866
	CGO–DHKELM	1.267	0.913	0.880	0.913	1.498	0.892	1.055	0.891
	CPA–DHKELM	1.532	0.874	1.147	0.874	1.621	0.870	1.221	0.869
Sum3	DHKELM	2.075	0.771	1.605	0.770	2.151	0.761	1.750	0.760
	BES–DHKELM	1.299	0.913	0.924	0.912	1.767	0.819	1.245	0.819
	CGO–DHKELM	1.435	0.886	1.068	0.886	1.747	0.862	1.326	0.860
	CPA–DHKELM	1.657	0.842	1.276	0.842	2.051	0.831	1.538	0.830
Sum4	DHKELM	1.866	0.822	1.398	0.820	1.925	0.778	1.482	0.778
	BES–DHKELM	1.159	0.928	0.759	0.927	1.868	0.830	1.243	0.827
	CGO–DHKELM	1.364	0.902	0.955	0.902	1.654	0.853	1.145	0.850
	CPA–DHKELM	1.402	0.895	1.029	0.895	1.900	0.837	1.263	0.813

The prediction models with Spm1 and Sum1 as input combinations for both spring and summer maize presented high errors and uncertainty, which may be attributed to the limitations of the input features. Spm1 included three factors, namely, planting density, plant height, and R_n, whereas Sum1 included four factors, namely, R_n, plant height, planting density, and $T_{\min}$. Although these features were significantly correlated with the yield of both spring and summer maize, such feature sets were not sufficient to provide the model with enough information to make accurate predictions. Spm2 exhibited increased T_{mean} and $T_{\min}$ compared to Spm1, with relative importance coefficients of 7.259% and 5.913%, respectively; Sum2 exhibited increased T_{mean} compared to Sum1, with a relative importance coefficient of 10.248%. With the introduction of more relevant features, the prediction accuracy of the models significantly improved, probably because Spm2 and Sum2 included more temperature-related environmental factors, which provided ample information to help the prediction models capture yield variations more accurately.

The prediction models were more likely to achieve the highest accuracy when the input combination (Spm3 and Sum2) contained the key factors affecting maize yield. Thus, the introduction of R_H (3.107%), U (3.107%), Pre (1.874%), and $T_{\max}$ (1.627%) to the yield prediction model of spring maize, and T_a (6.208%), U (4.743%), Pre (2.050%), R_H (1.523%), and $T_{\max}$ (1.405%) to the yield prediction model of summer maize did not lead to further improvement in the prediction accuracy of the models. This suggested that the prediction model may have captured the primary information in the data after a certain level of complexity, that the introduction of additional factors did not provide significant accuracy gains, and that the introduction of several features may have led to model overfitting. These findings demonstrated that, when constructing a model, the more critical input factors need to be selected to avoid model overfitting and unnecessary complexity. Feature selection techniques, which can effectively determine the most helpful factors for prediction, can simplify the model while maintaining high prediction accuracy.

3.3 Analysis of optimal yield prediction models for spring and summer maize

To improve the accuracy of the maize yield prediction model, three optimization algorithms (BES, CGO, and CPA) were used in this study to adjust and optimize the hyperparameters in the DHKELM model, which included, but were not limited to, penalty coefficients, regularization coefficients, and the number of nodes in the hidden layer, and which exhibited an important impact on the learning process of the model and the final prediction performance. By optimizing these parameters, the model could use the most useful features of the data for data training and obtain the best generalization ability. Thus, to display the prediction performance of the model intuitively, this study used scatter plots for the prediction results under different combinations of inputs (shown in Figure 5), and used GPI values to evaluate the comprehensive performance of the model (as listed in Tables 5 and 6). The GPI value accounted for multiple aspects of the predicted results and provided a more comprehensive view of the overall performance of the model. This study revealed that the application of the optimization algorithm resulted in a significant improvement in the prediction accuracy of the DHKELM model, which could provide the expected results. Among the three hybrid models, the CGO–DHKELM model exhibited the optimal prediction performance, followed by the CPA–DHKELM and BES–DHKELM models.

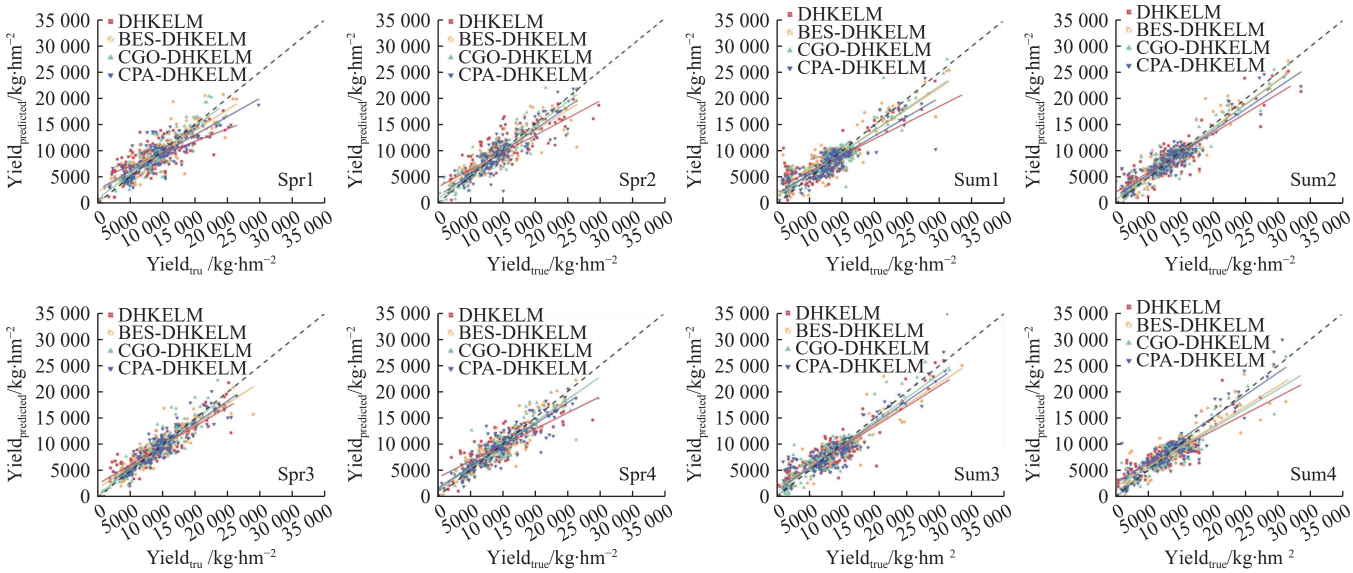

Figure 5 Scatter plot of predicted and true values of spring and summer maize yield

Table 5 GPI and ranking of yield prediction results for spring maize under different input combinations

Input combination	DHKELM		BES–DHKELM		CGO–DHKELM		CPA–DHKELM	
	GPI	Rank	GPI	Rank	GPI	Rank	GPI	Rank
Spm1	−2.000	16	−0.100	13	1.290	7	−0.252	14
Spm2	−1.004	15	0.485	11	1.737	4	0.954	10
Spm3	0.960	9	1.749	3	1.991	1	1.818	2
Spm4	0.103	12	1.435	6	1.658	5	1.214	8

Take the spring maize yield prediction model as an example. The accuracy of the independent DHKELM model in predicting the yield of spring maize provided the following values: RMSE=2.130±0.337 t/hm^2, R^2=0.710±0.102, MAE=1.650±0.289 t/hm^2, and NSE=0.708±0.101. Among the three optimization algorithms, the CGO algorithm best optimized the DHKELM model, providing the following values: RMSE=1.577±0.089 t/hm^2, R^2=0.843±0.019, MAE=1.135±0.084 t/hm^2, and NSE=0.830±0.022. The CPA also improved the accuracy of the DHKELM model substantially, with the following values: RMSE=1.827±0.297 t/hm^2, R^2=0.791±0.063, MAE=1.339±0.194 t/hm^2, and NSE=0.788±0.065. Furthermore, the prediction performance of the BES–DHKELM model was better than that of the independent DHKELM model, with the following values: RMSE=1.902±0.266 t/hm^2, R^2=0.806±0.053, MAE=1.337±0.209 t/hm^2, and NSE=0.802±0.052.

Table 6 GPI and ranking of yield prediction results for summer maize under different input combinations

Input combination	DHKELM		BES–DHKELM		CGO–DHKELM		CPA–DHKELM	
	GPI	Rank	GPI	Rank	GPI	Rank	GPI	Rank
Sum1	−2.000	16	−0.295	12	0.716	6	−0.354	13
Sum2	0.268	10	1.294	3	2.000	1	1.381	2
Sum3	−1.180	15	0.604	8	1.004	5	0.029	11
Sum4	−0.366	14	0.606	7	1.240	4	0.504	9

As listed in Tables 5 and 6, the yield prediction models for spring and summer maize exhibited the highest GPI values for each model with the optimal combination of factors as inputs; and the hybrid model exhibited higher GPI values than the independent DHKELM model for the same combination of input factors. This indicated that the optimization algorithm effectively adjusted the parameters of the independent DHKELM model and significantly improved its overall performance. All of these CGO–DHKELM models presented the most significant optimization results with the highest GPI ranking, which may be attributed to the high efficiency of the CGO algorithm regarding parameter optimization. This algorithm is a parameter-free meta-heuristic algorithm that explores the parameter space based on some principles of the chaos theory and determines the optimal model parameters without requiring set internal parameters[51]. Nonetheless, the BES–DHKELM and CPA–DHKELM models outperformed the unoptimized DHKELM model in terms of prediction performance but exhibited slightly lower performance than the CGO–DHKELM model. This may be because, although CPA could perform parallel computation and was easy to implement, the optimization search of the algorithm, which was based on the predation behavior of carnivorous plants, lacked mathematical support and may have been affected by the initial population. In addition, CPA may be affected by the attraction rate parameter and may converge prematurely, making it difficult to find a globally optimal solution[52]. The BES algorithm presents a strong global search capability and can cope with large-scale optimization problems. However, the BES algorithm tends to reduce local optimization, which may cause the algorithm to fall into local optimal solutions. In addition, the BES algorithm may use unnecessary repetitive searches, which can reduce its convergence speed[53].

From these findings, it is revealed that the introduction of the optimization and feature analysis algorithms improved the performance of the DHKELM model in predicting the yield of spring and summer maize. The CGO algorithm exhibited the most outstanding performance at addressing this optimization problem and effectively improved the prediction accuracy and reliability of the DHKELM algorithm. Thus, the CGO–DHKELM model with Spm3 and Sum2 as input combinations was the optimal yield prediction model for spring and summer maize.

4 Conclusions

The aim of this study was to develop an accurate and highly operational maize-yield prediction model based on the DHKELM algorithm. In this regard, to understand the correlation of each factor affecting the yield of spring and summer maize, four feature importance analysis algorithms were used in this study, including CART, RF, GBDT, and XGBoost, which could provide reliable and accurate results. Based on the results of the analysis, this study used different combinations of input factors to determine the optimal combination. To further improve the prediction accuracy and generalization ability of the model, three optimization algorithms (BES, CGO, and CPA) were combined with the DHKELM model and determined the optimal parameter configurations, to construct a maize yield prediction model with higher accuracy and good applicability. The results of the study were as follows:

1) Four tree-based feature analysis models (CART, RF, GBDT, and XGBoost) provided important information, thus improving the accuracy of the yield prediction model. It is noted that planting density and plant height were the most important factors and R_n was the main meteorological factor affecting maize yield during the fertility period. These were followed by T_{mean}, T_{min}, and T_a. However, the relative importance of R_H, Pre, U, and T_{max} was less.

2) The optimal combination of factors for predicting the yield of spring maize included planting density, plant height, R_a, T_{mean}, T_{min}, and T_a, whereas that for predicting the yield of summer maize included R_a, plant height, planting density, T_{min}, and T_{mean}.

3) The BES, CGO, and CPA algorithms effectively adjusted and optimized the hyperparameters of the DHKELM model, thereby significantly improving the prediction accuracy of the model. Among the three optimization algorithms, the CGO algorithm optimization was the best. When the optimal factor combination was used, the CGO–DHKELM model showed optimal performance at predicting the yield of both spring and summer maize, with the following values: RMSE=1.488 t/hm^2, R^2=0.862, MAE=1.051 t/hm^2, and NSE=0.852 (spring maize); RMSE=1.498 t/hm^2, R^2=0.892, MAE=1.055 t/hm^2, and NSE=0.891 (summer maize).

This study also compared the yield prediction performance of the independent DHKELM and hybrid models under different input combinations and constructed a high-precision maize-yield prediction system. The findings of this study can thus provide farmers and agricultural specialists with accurate crop yield-prediction results to develop effective planting strategies and resource management plans. Nonetheless, the applicability of these models to multi-region and multi-year datasets and the performance of models under different agricultural management strategies can be further explored. Additionally, exploring the performance of similar predictive models for crop yield estimation can simplify the model application process for farmers.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grant No. 52309050 and 32372680), Youth Backbone Teacher Project of Henan University of Science and Technology (Grant No. 13450013 and 3450010), Key Scientific Research Projects of Colleges and Universities in Henan Province (Grant No. 24B416001), and Innovative Research Team (Science and Technology) in the University of Henan Province (Grant No. 23IRTSTHN024).

[References]

[1] Han J C, Zhang Z, Cao J, Luo Y C, Zhang L L, Li Z Y, et al. Prediction of winter wheat yield based on multi-source data and machine learning in China. Remote Sensing, 2020; 12(2): 236.

[2] Cheng M H, Penuelas J, McCabe M F, Atzberger C, Jiao X Y, Wu W B, et al. Combining multi-indicators with machine-learning algorithms for maize yield early prediction at the county-level in China. Agricultural and Forest Meteorology, 2022; 323: 109057.

[3] Li M Y, Zhao J, Yang X G. Building a new machine learning-based model to estimate county-level climatic yield variation for maize in Northeast China. Computers and Electronics in Agriculture, 2021; 191: 106557.

[4] Geipel J, Link J, Claupein W. Combined spectral and spatial modeling of corn yield based on aerial images and crop surface models acquired with an unmanned aircraft system. Remote Sensing, 2014; 6(11): 10335–10355.

[5] Zhao Y X, Xiao D P, Bai H Z, Tang J Z, Liu D L, Qi Y Q, et al. The prediction of wheat yield in the North China Plain by coupling crop model with machine learning algorithms. Agriculture, 2022; 13(1): 99.

[6] Tao F, Palosuo T, Rötter R P, Díaz-Ambrona C G H, Mínguez M I, Semenov M A, et al. Why do crop models diverge substantially in climate impact projections? A comprehensive analysis based on eight barley crop models. Agricultural and Forest Meteorology, 2020; 281: 107851.

[7] Berghuijs H N C, Silva J V, Reidsma P, de Wit A J W. Expanding the WOFOST crop model to explore options for sustainable nitrogen management: A study for winter wheat in the Netherlands. European Journal of Agronomy, 2024; 154: 127099.

[8] Gautam V, Gani A, Pathak S, Shukla A K. Evaluating crop yield prediction models in illinois using aquacrop, semi-physical model and artificial neural networks. Scientific Reports, 2025; 15(1): 27494.

[9] Xu X B, He W, Zhang H Y. Random hierarchical model for estimation of wheat yield in the North China Plain at different spatial scales. Field Crops Research, 2024; 306: 109226.

[10] Ren Y T, Li Q Z, Du X, Zhang Y, Wang H Y, Shi G W, et al. Analysis of corn yield prediction potential at various growth phases using a process-based model and deep learning. Plants, 2023; 12(3): 446.

[11] Kheir A M S, Mkuhlani S, Mugo J W, Elnashar A, Nangia V, Deware M, et al. Integrating APSIM model with machine learning to predict wheat yield spatial distribution. Agronomy Journal, 2023; 115(6): 3188–3196.

[12] Meng X, Hou Q Z, Quan L M, Qiao J F. Self-organizing fuzzy neural network with adaptive evolution strategy for nonlinear and nonstationary processes. Artificial Intelligence Review, 2025; 58(9): 280.

[13] Huang G B, Zhu Q Y, Siew C K. Extreme learning machine: a new learning scheme of feedforward neural networks. In: 2004 IEEE International Joint Conference on Neural Networks (IEEE Cat. No. 04CH37541), Budapest, Hungary: IEEE, 2004; pp.985–990.

[14] Huang G, Huang G B, Song S J, You K Y. Trends in extreme learning machines: A review. Neural Networks, 2015; 61: 32–48.

[15] Huang J C, Ko K M, Shu M H, Hsu B M. Application and comparison of several machine learning algorithms and their integration models in regression problems. Neural Computing and Applications, 2020; 32(10): 5461–5469.

[16] Jiang X F, Duan H C, Liao J, Guo P L, Huang C H, Xue X. Estimation of soil salinization by machine learning algorithms in different arid regions of Northwest China. Remote Sensing, 2022; 14(2): 347.

[17] Ji J T, Li N N, Cui H W, Li Y C, Zhao X B, Zhang H L, et al. Study on monitoring spad values for multispatial spatial vertical scales of summer maize based on UAV multispectral remote sensing. Agriculture, 2023; 13(5): 1004.

[18] Hai M W, Wang M, Zhou B, Zhang Q. Assessment of frost heave in coarse-grained soil: A novel application of multi-strategy enhanced dung beetle-optimized KELM model. Earth Science Informatics, 2025; 18(1): 63.

[19] Cheng Y H, Hu B B. Forecasting regional carbon prices in China based on secondary decomposition and a hybrid kernel-based extreme learning machine. Energies, 2022; 15(10): 3562.

[20] Ahmad F, Javed K, Tahir A, Khan M U G, Abbas M, Rabbani M, et al. Identifying key soil characteristics for Francisella tularensis classification with optimized machine learning models. Scientific Reports, 2024; 14(1): 1743.

[21] Zhao L, Qing S H, Li H, Qiu Z M, Niu X L, Shi Y, et al. Estimating maize evapotranspiration based on hybrid back-propagation neural network models and meteorological, soil, and crop data. International Journal of Biometeorology, 2024; 68(3): 511–525.

[22] Alsattar H A, Zaidan A A, Zaidan B B. Novel meta-heuristic bald eagle search optimisation algorithm. Artificial Intelligence Review, 2020; 53: 2237–2264.

[23] Huang Y. Improved SVM-based soil-moisture-content prediction model for tea plantation. Plants, 2023; 12(12): 2309.

[24] Talatahari S, Azizi M. Chaos game optimization: a novel metaheuristic algorithm. Artificial Intelligence Review, 2021; 54(2): 917–1004.

[25] He Y C, Yang K, Wang X Q, Huang H S, Chen J D. Quality prediction and parameter optimisation of resistance spot welding using machine learning. Applied Sciences, 2022; 12(19): 9625.

[26] Ong K M, Ong P, Sia C K. A carnivorous plant algorithm for solving global optimization problems. Applied Soft Computing, 2021; 98: 106833.

[27] Wang Y, Wang W, Chen Y. Carnivorous plant algorithm and BP to predict optimum bonding strength of heat-treated woods. Forests, 2022; 14(1): 51.

[28] Zhao L, Qing S H, Bai J Y, Hao H H, Li H, Shi Y, et al. A hybrid optimized model for predicting evapotranspiration in early and late rice based on a categorical regression tree combination of key influencing factors. Computers and Electronics in Agriculture, 2023; 211: 108031. doi: 10.1016/j.compag.2023.108031.

[29] Peng S R, Guo L J, Li Y S, Huang H Y, Peng J Y, Liu X X. Biogas production prediction based on feature selection and ensemble learning. Applied Sciences, 2024; 14(2): 901.

[30] Fei H, Fan Z H, Wang C K, Zhang N N, Wang T, Chen R G, et al. Cotton classification method at the county scale based on multi-features and random forest feature selection algorithm and classifier. Remote Sensing, 2022; 14(4): 829.

[31] Peng Y P, Wang L, Zhao L, Liu Z H, Lin C J, Hu Y M, et al. Estimation of soil nutrient content using hyperspectral data. Agriculture, 2021; 11(11): 1129.

[32] Mohammadi B, Mehdizadeh S. Modeling daily reference evapotranspiration via a novel approach based on support vector regression coupled with whale optimization algorithm. Agricultural Water Management, 2020; 237: 106145.

[33] Zheng H F, Chen L D, Han X Z, Zhao X F, Ma Y. Classification and regression tree (CART) for analysis of soybean yield variability among fields in Northeast China: The importance of phosphorus application rates under drought conditions. Agriculture, Ecosystems & Environment, 2009; 132(1–2): 98–105.

[34] Gill M, Anderson R, Hu H F, Bennamoun M, Petereit J, Valliyodan B, et al. Machine learning models outperform deep learning models, provide interpretation and facilitate feature selection for soybean trait prediction. BMC plant biology, 2022; 22(1): 180.

[35] Dastres E, Rabiei-Dastjerdi H, Esmaeili H, Amiri M, Sonboli A, Mirjalili M H. Modeling habitat suitability for endangered herb (Salvia leriifolia Benth) using innovative hybrid machine learning algorithms. Environmental and Sustainability Indicators, 2025; 26: 100694.

[36] Cheng E H, Zhang B, Peng D L, Zhong L H, Yu L, Liu Y, et al. Wheat yield estimation using remote sensing data based on machine learning approaches. Frontiers in Plant Science, 2022; 13: 1090970.

[37] Baloch N, Liu W M, Hou P, Ming B, Xie R Z, Wang K R, et al. Effect of latitude on maize kernel weight and grain yield across China. Agronomy Journal, 2021; 113(2): 1172–1182.

[38] Ustuner M, Simsek F F. An assessment of training data for agricultural land cover classification: a case study of Bafra, Türkiye. Earth Science Informatics, 2025; 18(1): 7.

[39] Miao H L, Zhang R, Song Z H, Chang Q R. Estimating winter wheat canopy chlorophyll content through the integration of unmanned aerial vehicle spectral and textural insights. Remote Sensing, 2025; 17(3): 406.

[40] Lv L, Wang W H, Zhang Z Y, Liu X G. A novel intrusion detection system based on an optimal hybrid kernel extreme learning machine. Knowledge-based systems, 2020; 195: 105648.

[41] Fu J C, Song Z X, Meng J H, Wu C L. Prediction of lithium-ion battery state of health using a deep hybrid kernel extreme learning machine optimized by the improved black-winged kite algorithm. Batteries, 2024; 10(11): 398.

[42] Wang S L, Guo X L, Sun T L, Xu L H, Zhu J F, Li Z C, et al. Short-term photovoltaic power forecasting based on the VMD-IDBO-DHKELM model. Energies, 2025; 18(2): 403.

[43] Wang Y G, Yu Y D, Ma Y C, Shi J. Lithium-ion battery health state estimation based on improved snow ablation optimization algorithm-deep hybrid kernel extreme learning machine. Energy, 2025; 323: 135772.

[44] Hou H P, Ma W, Noor M A, Tang L Y, Li C F, Ding Z S, et al. Quantitative design of yield components to simulate yield formation for maize in China. Journal of Integrative Agriculture, 2020; 19(3): 668–679.

[45] Luo S Z, Liu W W, Zhang Y Q, Wang C, Xi X H, Nie S, et al. Maize and soybean heights estimation from unmanned aerial vehicle (UAV) LiDAR data. Computers and Electronics in Agriculture, 2021; 182: 106005.

[46] Liu Q, Yang Z P, Zhou W, Wang T, Fu Y, Yue X P, et al. Solar radiation utilization of five upland–paddy cropping systems in low-light regions promoted by diffuse radiation of paddy season. Agricultural and Forest Meteorology, 2023; 338: 109527.

[47] Zhou N B, Zhang J, Fang S L, Wei H Y, Zhang H C. Effects of temperature and solar radiation on yield of good eating-quality rice in the lower reaches of the Huai River Basin, China. Journal of Integrative Agriculture, 2021; 20(7): 1762–1774.

[48] Gao Z, Feng H Y, Liang X G, Zhang L, Lin S, Zhao X, et al. Limits to

maize productivity in the North China Plain: A comparison analysis for spring and summer maize. Field Crops Research, 2018; 228: 39–47.

[49] Chen C, Pang Y M. Response of maize yield to climate change in Sichuan province, China. Global Ecology and Conservation, 2020; 22: e00893.

[50] Hou P, Liu Y E, Xie R Z, Ming B, Ma D L, Li S K, et al. Temporal and spatial variation in accumulated temperature requirements of maize. Field Crops Research, 2014; 158: 55–64.

[51] Talatahari S, Azizi M. Optimization of constrained mathematical and engineering design problems using chaos game optimization. Computers & Industrial Engineering, 2020; 145: 106560.

[52] Zhang P L, Sun X B, Wang J Q, Song H H, Bei J L, Zhang H Y. The discrete carnivorous plant algorithm with similarity elimination applied to the traveling salesman problem. Mathematics, 2022; 10(18): 3249.

[53] Wu J H, Hu Y, Wu D Q, Yang Z Y. An aquatic product price forecast model using VMD-IBES-LSTM hybrid approach. Agriculture, 2022; 12(8): 1185.