Research On Maximizing Crop Planting Returns Based on Genetic Algorithm

Chenbo Rao *

School of optoelectronic Engineering, Changchun University of Science and Technology, Changchun, China, 130022

* Corresponding Author Email: raoshuai926@gmail.com

Abstract. With the rapid development of agricultural modernization, the traditional planting method relying on experience makes it difficult to effectively cope with market fluctuations and uneven distribution of resources, resulting in the difficulty of maximizing agricultural returns. To solve this problem, some crop data were pre-processed and the extreme values in the data were processed by IQR (Interquartile Range). On this basis, the agricultural planting income maximization model was constructed, and the Genetic Algorithm (GA) was used to optimize the model's prediction. Through genetic operations such as selection, crossover, and mutation, the planting plan was gradually optimized, and the optimal planting planning plan for a rural plot in 2024 was finally obtained. For example, plot A1 with an area of 85 acres of arid land is suitable for growing sorghum, and plot A2 with 55 acres is ideal for growing red beans. The results of this study provide a reference value for the optimization of agricultural planting and are of great significance for improving the efficiency and income of rural resource utilization.

Keywords: Genetic Algorithm, Agricultural Planting, Interquartile Range, Prediction Model.

1. Introduction

With the advancement of agricultural modernization, agricultural planting no longer relies on traditional methods, and the scientific and reasonable planning of crop planting programs has become an important means to improve the efficiency of agricultural production. Traditional crop planting programs rely on experience and historical data, but it is difficult to effectively cope with problems such as market demand fluctuations and uneven resource allocation with this method. This results in frequent occurrences of unsalable or discounted sales, which affect the maximization of planting income.

In recent years, the application of intelligent optimization algorithms in the agricultural field has gradually attracted attention. Among them, as a heuristic algorithm based on the principle of biological evolution, GA is widely used in complex combinatorial optimization problems because of its strong global search ability and flexibility. The GA was proposed and designed by John Holland of the United States in the 1970s according to the evolutionary laws of organisms in nature. At present, GAs have been widely used in microorganisms, neural networks, image processing, adaptive control, and other fields. Wang et al published a paper on optimizing the BP neural network based on a GA, which reduces the cost of agricultural production and provides a value orientation for the optimal planting scheme in this study [1]. Ye et al published a paper proposing the selection of the optimal chromosome by GA, which provides a reference for the selection of the optimal planting scheme based on the GA in this study [2]. Emily Neil published a paper using GAs to calculate and calibrate multi-constraint ensemble ecosystem models [3]. Maionchi published a paper on geometric shape prediction models and optimization based on GAs, which provides suggestions for the prediction model in this study [4]. Bhatia published a paper on the application of GAs in agricultural problem species, which provided a reference for agricultural planting in this study [5]. Lu et al published a paper on microbial feature selection based on GAs, which compared microorganisms to obtain optimal feature solutions [6]. Li et al. provided suggestions for the establishment of a prediction model based on a GA to improve the prediction accuracy and model generalization ability of the BP neural network model [7]. Yang et al have also published a paper on the establishment of profit functions for vegetables and the use of GAs to solve the programming model, which provides a reference for the maximum benefit model in this study [8]. Ma et al. have analyzed the income and cost of agriculture in China, which provides a reference for this study [9]. At present, there is still a research gap in the large-scale optimization analysis of different crops, and this study discusses this problem in depth, focusing on the design of fitness function and optimization of planting schemes around the problem of maximizing the benefits of different crop planting, to provide more possibilities for future agricultural optimization.

This study aims to use a GA to optimize and predict the agricultural planting strategy in a village, aiming to improve the efficiency of agricultural production and provide a scientific basis for agricultural decision-making. First, the IQR method was used to preprocess the agricultural data and deal with the extreme values in the data. IQR effectively identifies and eliminates extreme values by calculating the difference between the 25th percentile (Q1) and the 75th percentile (Q3) of the data, thus avoiding the impact of abnormal data on the analysis of results and ensuring the accuracy and robustness of the data. Based on data processing, the objective function of crop income was further established according to the yield of different crops. Specifically, the study considers the situation where a crop is sold at a lower price (e.g., 50% of the market price) when the crop yields exceed the expected sales volume or the yield, and this price adjustment will have an impact on the final yield. Therefore, the objective function not only needs to maximize the yield of the crop but also dynamically adjusts to different market situations in response to yield fluctuations. Finally, the GA algorithm was used to select the ones with greater adaptability and formulate the optimal planting plan to obtain the best planting plan for a rural plot in 2024.

2. The data comes from extreme-value processing

Data for this study are from: https://www.mcm.edu.cn/index_cn.html. IQR was produced in the era of manual calculation and plotting, which involves a small data set, and is an effective method for labeling outliersProposed in 1977 by John Tukey in his book Exploratory Data Analysis, it is an important method used in statistics to characterize the distribution of data. Ming et al. published a paper in 2016 using IQR to detect gross errors in residuals, which provided a reference for the treatment of extreme values in this study [10]. IQR is an important statistic in the distribution of a dataset, which represents the range of the middle 50% of the data in the dataset. Specifically, IQR is the difference between the upper quartile (Q3) and the lower quartile (Q1), reflecting how discrete the data is. IQR is more effective at measuring the central trend and variability of data than extreme values (such as maximum and minimum values) or averages, especially when there are outliers or skewed distributions in the data. By eliminating the effects of extreme values, IQR provides a more robust measure of variability, making it a more reliable tool than standard deviation when analyzing statistics with asymmetric distributions or outliers. Therefore, IQR not only helps to improve the accuracy of data analysis but also enables a better understanding of the middle part of the data in practical applications to avoid being misled by extreme values.

In simple terms, IQR is achieved by dividing the dataset into 4 parts, i.e., quartiles. The four data divided are equal data sets, namely Q1, Q2, Q3, and Q4, where Q1 is the 25th percentile of the data (the first quartile and Q2), Q3 is the 75th percentile of the data (the third quartile and Q4) and Q4 (the fourth quartile is shown in Figure 1).

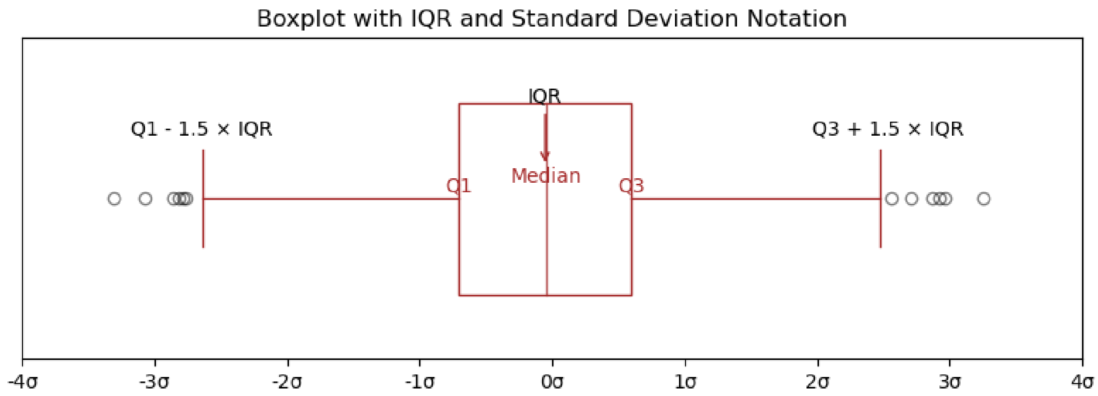

Figure 1. Schematic diagram of the IQR method

The magnitude of the value of IQR is:

$$\mathrm{IQR} = \mathrm{Q3} - \mathrm{Q1} \tag{1}$$

Q3 and Q1 are symmetrical concerning the intermediate values, and the magnitude of the IQR values is obtained. John Tukey believes that the IQR of the data is 1.5 times lower than the first quartile point, or the IQR of the data is 1.5 times higher than the third quartile point, the data point can be considered to be out of the data range, and the data can be considered to be outlier, if the deviation is too much, it is excessive outlier, that is, the upper limit of the dataset and the lower bound of the dataset are solved, and the upper and lower limits of the IQR are calculated as follows

$$\mathrm{upperlimit} = \mathrm{Q3} + 1.5 * \mathrm{IQR} \tag{2}$$

$$\mathrm{lowerlimit} = \mathrm{Q1} - 1.5 * \mathrm{IQR} \tag{3}$$

With the IQR calculation formula, the data can be processed with extreme values, but due to the large number of datasets, only a part of them is shown here, as shown in Figure 2 below.

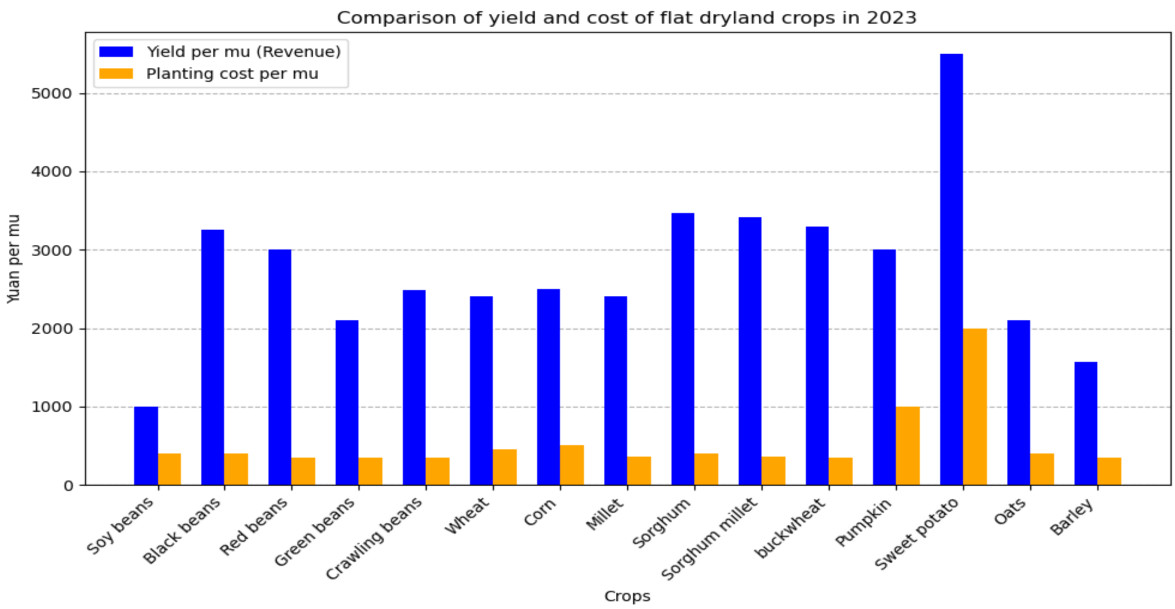

Figure 2. Comparison of benefits and costs per acre of flat dryland crops

The figure shows the comparison chart of income per acre and cost of flat and dryland crops in 2023, and the IQR analysis and calculation of the income per acre of flat and dryland crops are carried out, and the following results are shown in Table 1

Table 1. IQR of yield per acre of crop

IQR	Q1	Q3	lower limit	upper limit
1025	2250	3275	712.5	4812.5

As can be seen from Table 1, the yield dataset ranges from (712.5, 4812.5), and under the same planting conditions, the yield per acre beyond this range is the extreme value and should be excluded.

The IQR can be used to effectively process the data and judge the possible extreme values in the data, which can effectively avoid losses in agricultural planting and provide a reference value for formulating agricultural planting plans.

3. Multi-crop planting decision model based on yield optimization

In order to maximize the benefits of agricultural planting, this study establishes the income maximization function, first of all, it is necessary to determine the main decision-making variable and the objective function, here this study defines xij(t) as the planting area (acre) of the ith crop in the t-year of the j-block land, xij(t) is the main decision-making variable of the model, and the objective function is to establish the objective function, the objective function is to maximize the income of agricultural planting, which is divided into two situations, one is that the yield is greater than the expected sales volume, then the remaining products cannot be sold, resulting in unsalable wasteFor case one, establish a revenue model, known output is greater than or equal to the sales volume, in the model establishment, you can use min{A,B} to obtain the minimum value in A and B, and use it to achieve the minimum value of the output and sales volume is the actual sales volume, then the total income is the product of the actual sales volume and the sales price, and at the same time consider the income of different crops in different plots, then the maximum income model for case one is

$$\mathrm{Profit}_1 = \sum_{t=2024}^{t'} \sum_{i=1}^{n} \sum_{j=1}^{m} \min\left(S_i^{(t)}, D_i^{(t)}\right) \cdot P_i^{(t)} \tag{4}$$

The goal is to maximize earnings from 2024 to the next few years, in the model $S_i^{(t)}$is the total output of clock i crop in year t;$D_i^{(t)}$is the expected sales volume of a clock i crop in year t;$P_i^{(t)}$ is the selling price of the crop of clock i in year t;

In case 1, the minimum value of the total production volume and the expected sales volume is obtained, which is used as the actual sales volume, and the actual sales volume is sold, and if there is a surplus, a loss will be incurred

For case 2, relative to case 1, if there is a surplus of production, then sell at 50% of the price, you can try to avoid too much loss, the following to establish its maximum return function, as we all know, based on case 1, this study needs to add the price of the remaining production to sell, but also need to consider that if the remaining production cannot be sold all, you may wish to use max{A, B} to obtain the maximum value of A and B, here take the actual sales volume of the remaining production and 0 as A and B, you can get the actual sales volume of the remaining production. Add the sales gain in case 1 to get the maximum return function in case 2.

$$Profit_2 = \sum\left[min\left(S_i(t),\ D_i(t)\right) \cdot P_i(t) + max\left(S_i(t) - D_i(t), 0\right) \cdot 0.5 \cdot P_i(t)\right] \tag{5}$$

Once you have the return function, you also need to consider the constraints

$$\sum x_{ij}(t) \le A_j \tag{6}$$

That is, all planting areas cannot be larger than the total area, so far, the maximum return function model of this study has been established, which lays a foundation for the subsequent GA analysis and calculation of the optimal planting scheme.

4. Genetic algorithm model and its optimization solution in crop planting

After dealing with the extreme values and establishing the maximum planting benefit model, how to use the GA to optimize the model, for the organisms in nature, when under different environmental conditions have different adaptations to nature, it is also applicable to crops, in part 3, this study has established a decision model of the maximum return function, the maximum benefit indirectly reflects the degree of adaptation of the crop to the environment, and the corresponding fitness function should also be the same as the decision model of the maximum benefit function. That is, the fitness function is

$$\text{Fitness} = \sum_{t=2024}^{t'} \sum_{i=1}^{n} \sum_{j=1}^{m} \min\left(S_i^{(t)}, D_i^{(t)}\right) \cdot P_i^{(t)} \tag{7}$$

Roulette selection is a selection operation method commonly used in GAs, and its main role is to select the "parent" according to the individual's fitness value, that is, to determine which individuals can enter the next generation. The basic idea is to allocate the probability of being selected according to the fitness of the individual, and the higher the fitness, the greater the probability of the individual being selected. In this way, GAs can tend to select superior individuals and thus produce better solutions in subsequent evolution.

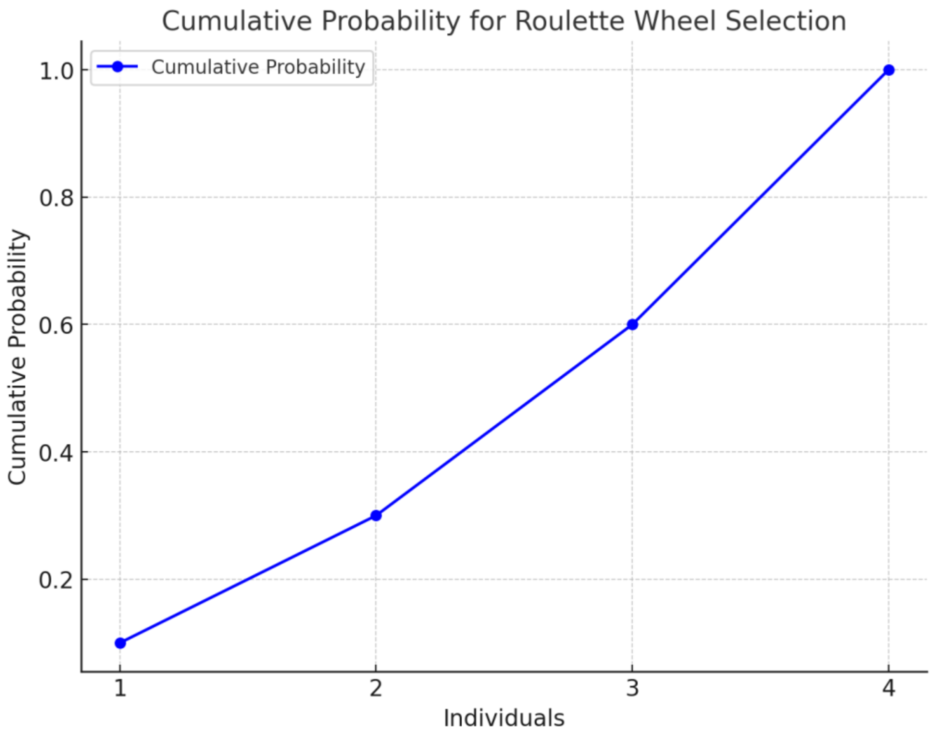

Figure 3. Cumulative probability curves of individuals

Figure 3 illustrates the cumulative probability curve for individuals in the roulette selection method. As an individual's fitness increases, so does the probability of choosing on the wheel, and the curve shows an upward trend. Each dot represents the cumulative probability of the corresponding individual, indicating the cumulative probability of its selection.

In Figure 2 of Part 2, it is shown that the planting income per acre of crops sold at the lowest price and the planting cost per acre are shown to be processed by the extreme value, and it is found that the net income per acre of sweet potato is greater than the upper limit of IQR treatment, that is, the income per acre of sweet potato belongs to the extreme value, which should be excluded and analyzed on the remaining crops. Pick out the crops sold at the lowest price, and the difference between the planting income per acre and the planting cost per acre is the largest. Table 2 below shows the following.

Table 2. Profit per acre of crops

cropper	Income per acre (RMB)	Cost per acre (RMB)	Net income (RMB)
soybean	1000	400	600
Black beans	3250	400	2850
Red beans	3000	350	2650
mung bean	2100	350	1750
Climb the beans	2490	350	2140
wheat	2400	450	1950
corn	2500	500	2000
millet	2400	360	2040
sorghum	3465	400	3065
Millet	3412.5	360	3052.5
buckwheat	3300	350	2950
pumpkin	3000	1000	2000
Wheat	2100	400	1700
barley	1575	350	1225

The table shows the difference between the planting income per acre and the planting cost per acre of the crops sold at the lowest price, and the net income per acre, to maximize the planting income, only the top 5 crops with the largest difference and the largest net income per acre are taken, as shown in Table 3

Table 3. The five crops with the highest profits per acre

cropper	Income per acre (RMB)	Cost per acre (RMB)	Net income (RMB)
sorghum	3465	400	3065
Millet	3412.5	360	3052.5
buckwheat	3300	350	2950
Black beans	3250	400	2850
Red beans	3000	350	2650

The GA was performed on the top five crops with the largest income per acre, the population size was set to 50 and the crossover probability was 08. The mutation probability is 0.1, and the maximum number of iterations is 100. The specific process is to set the population parameters and calculate the fitness (i.e., profit) of each individual. The top 50% of individuals (individuals with high fitness) are selected based on fitness. The selected individuals are crossover (genetically recombinant). By selecting the intersection point, the genes of the two parent individuals are exchanged to generate two offspring. Randomly change an individual's genes with a certain probability (0.1) (i.e., randomly select a plot of crops). Cross and mutate individuals to form new populations. The GA in this study optimizes the crop planting scheme to maximize the total profit for a given plot area and crop cost and benefit. Through genetic operations such as selection, crossover, and mutation, the algorithm gradually optimizes the planting plan and finally returns an optimal solution.

It is important to note that the optimal results of this study are not the only ones due to factors such as the randomness of the initial population, the randomness in the crossover operation, the randomness in the mutation operation, the uncertainty in the selection operation, the accumulation effect of randomness, and the randomness of the sales price. After considering the above factors, the optimal planting plan for crops can be predicted.

After GA processing of the data, the fitness (profit) curve with algebra was obtained, as shown in Figure 4

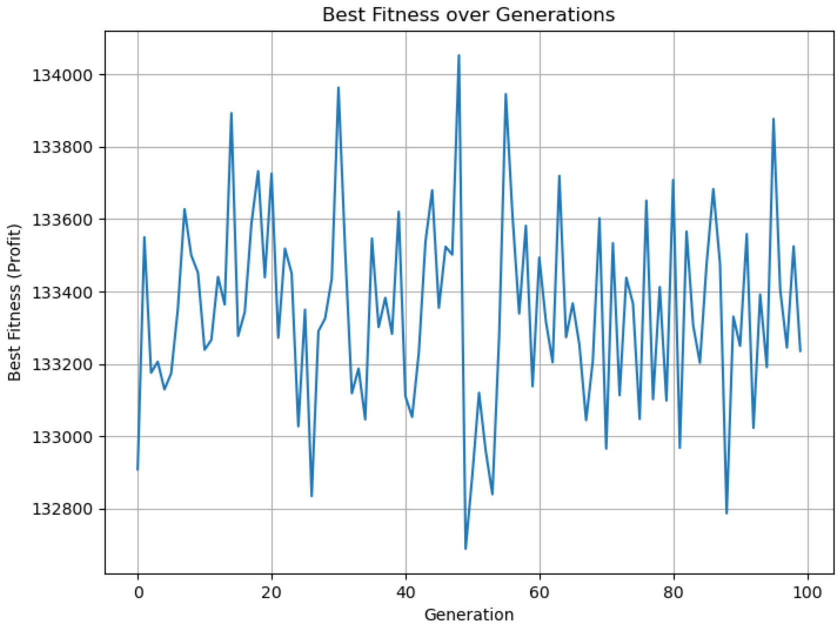

Figure 4. Fitness changes with algebra

Figure 4 illustrates the convergence process of the GA, and the optimal planting scheme for the 2024 dryland was obtained by combining all the above calculation processes, as shown in Table 4 below

Table 4. Dryland crop projections for 2024

Drylands	Areas	Optimal planting predictions
A1	85	sorghum
A2	55	Red beans
A3	35	sorghum
A4	72	sorghum
A5	68	Red beans
A6	55	Black beans

Table 4 shows the forecast of crop planting in rural arid land in 2024, which can maximize the benefits to a certain extent, and this study provides some suggestions for agricultural planting.

5. Conclusion

In this study, a GA model was used to predict and optimize crop planting in arid land in 2024 to optimize the planting and marketing of rural crops. In this study, the planting cost, yield per acre, sales price, and historical sales volume of crops were extracted from the actual data, and a model was constructed based on these data. Firstly, IQR was used to process the extreme value of the data, the maximum return model was established, and finally, the GA was used to optimize the planting plan. The goal of the GA is to maximize future economic gains by adjusting the planting area of each plot. GA parameter setting: population size is 50, crossover probability is 0.8, mutation probability is 0.1, and the maximum number of iterations is 100 In the evolution process of each generation, the GA gradually optimizes the planting scheme through selection, crossover, and mutation. Each individual in the population represents a possible crop cultivation combination, which after 100 generations of evolution, finally converges to the optimal solution. The convergence curve of GA clearly shows the improvement of fitness in the optimization process, and the income gradually stabilizes with the increase of algebra, indicating that the planting plan is gradually approaching the optimum.

This study proposes an innovative research approach and framework, aiming to provide new research tools for agricultural economics, agricultural planting, and related disciplines. By introducing the GA, the maximum return model of crops is established, and the intelligent optimization method is used to solve the complex decision-making problems in agricultural production. The model can not only verify the feasibility of income optimization at the theoretical level but also provide a scientific basis and practical guidance for actual agricultural production. The results of this study provide practical suggestions for planting planning, resource allocation, and income improvement in agriculture and rural areas, and provide theoretical support and decision-making reference for promoting agricultural modernization and high-quality development of rural economy.

References

[1] Wang Jianan, Wang Yuying, He Shulin, et al. Soil Moisture Prediction Model Based on Improved Genetic Algorithm for Optimizing BP Neural Network [J]. Journal of Computer Systems Application, 2022, 31 (02): 273-278.

[2] Ye Yu, Li Min, Wen Yan. Research on Classification and Prediction Algorithm for Agricultural Big Data [J]. Computer and Digital Engineering, 2022, 50 (03): 468-470+475

[3] Neil E, Carrella E, Bailey R. Calibrating multi-constraint ensemble ecosystem models using genetic algorithms and Approximate Bayesian Computation: A case study of rewilding at the Knepp Estate, UK [J]. Ecological Modelling, 2025, 500 110948-110948.

[4] Maionchi O D D, Coimbra S D N, Silva D G J, et al. Predictive model and optimization of micromixers geometry using Gaussian process with uncertainty quantification and genetic algorithm [J]. Fluid Dynamics Research, 2024, 56 (6): 065504-065504.

[5] Bhatia, K. A . Applications of genetic algorithms in agricultural problems - an overview. (Special Issue: Artificial intelligence in agriculture.) [J]. Journal of the Indian Society of Agricultural Statistics, 2013, 67 (1): 13-22.

[6] Lu Fumei, Wen Liuying. Microbial Feature Selection Method Integrating Single-class F-score and Genetic Algorithm [J]. Information Technology, 2024, (11): 125-131.

[7] Li Yunfeng, Yan Sixing, Ran Binbin, et al. Prediction of Lateral Bearing Capacity of Pile Foundation Based on Genetic Algorithm for Improving BP Neural Network [J]. Science and Technology Innovation and Application, 2024, 14 (33): 30-33.

[8] Yang Jia, Ren Changhang. Decision Analysis of Automatic Pricing and Replenishment of Vegetable Commodities from the Perspective of Genetic Algorithm [J]. Modern Business, 2024, (20): 23-26.

[9] Ma Xiaohe. Structural analysis of changes in agricultural income and production cost in China [J]. China Rural Economy, 2011, (05): 4-11+56

[10] Ming Feng, Zeng Anmin, Jing Yifan. GNSS Coordinate Sequence Gross Error Detection Algorithm Based on L1 Norm and IQR Statistics [J]. Journal of Surveying and Mapping Science and Technology, 2016, 33 (02): 127-132