Hyperspectral prediction of pigment content in tomato leaves based on logistic-optimized sparrow search algorithm and back propagation neural network

Jiangui Zhao,[1] Tingyu Zhu,[1] Zhichao Qiu,[1] Tao Li,[1] Guoliang Wang,[2] Zhiwei Li,[1,3] Huiling Du[4]

[1]College of Agricultural Engineering, Shanxi Agricultural University, Taigu; [2]Institute of Millet Research, Shanxi Agricultural University, Changzhi; [3]College of Information Science and Engineering, Shanxi Agricultural University, Jinzhong; [4]Department of Basic Sciences, Shanxi Agricultural University, Taigu, China

Abstract

Leaf pigment content can reflect the nutrient content of the cultivation medium indirectly. To rapidly and accurately predict the pigment content of tomato leaves, chlorophyll a, chlorophyll b, chlorophyll and carotenoid were extracted from the leaves of tomato seedlings cultured at different nitrogen concentrations. The visible/near-infrared hyperspectral imaging non-destructive measurement technology, 430-900 nm and 950-1650 nm, with total variables of 794, was used to obtain the reflection spectra of leaves. An improved strategy of the sparrow search algorithm (SSA) based on logistic chaotic mapping was proposed, and it optimized the back propagation neural network to predict the pigment content of leaves. Different pretreatment methods were used to effectively improve the prediction accuracy of the model. The results showed that when the nitrogen concentration in the nutrient solution was 302.84 mg·L^{-1}, the pigment content of the leaves reached its maximum. Meanwhile, the inhibition effect of high concentrations was much stronger than that of low concentrations. To address the problem that the SSA is prone to premature convergence due to the reduction of population diversity at the end of the iteration, the initialization of the SSA population by logistic chaotic mapping improves the initial solution quality, convergence speed, and search capacity. The root mean squared error (RMSE), coefficient of determination (R^2) and relative percent deviation (RPD) of chlorophyll a were 0.77, 0.77, and 2.08, respectively. The RMSE, R^2 and RPD of chlorophyll b were 0.30, 0.66, and 1.71, respectively. The RMSE, R^2 and RPD of chlorophyll were 0.88, 0.81, and 2.28, respectively. The RMSE, R^2 and RPD of carotenoid were 0.14, 0.75, and 2.00, respectively. Hyperspectral imaging technology combined with machine learning algorithms can achieve rapid and accurate prediction of crop physiological information, providing data support for the precise management of fertilization in facility agriculture, which is conducive to improving the quality and output of tomatoes.

Correspondence: Zhiwei Li, College of Information Science and Engineering, Shanxi Agricultural University, Jinzhong 030801, China. Tel.: 86.0354.6289.384. E-mail: lizhiweitong@163.com

Key words: hyperspectral imaging, logistic chaotic mapping, pigments content, sparrow search algorithm, tomato leaf.

Contributions: all the authors made a substantial intellectual contribution, read and approved the final version of the manuscript and agreed to be accountable for all aspects of the work.

Conflict of interest: the authors declare no conflict of interests. The funders have no role in the experimental design, data collection and analysis, decision to publish, or preparation of the manuscript.

Funding: this work was supported by Major Special Projects of National Key R&D (2021YFD1600301-4), Major Special Projects of National Key R&D (2017YFD0701501), Major Special Projects of Shanxi Province Key R&D (201903D211005), Construction Project of Shanxi Modern Agricultural Industry Technology System.

Availability of data and materials: data and materials are available from the corresponding author upon request.

Received: 12 November 2022.
Accepted: 13 March 2023.
Early view: 20 June 2023.

Journal of Agricultural Engineering 2023; LIV:1528
doi:10.4081/jae.2023.1528

Introduction

Nitrogen is an essential element for tomato growth and development. Appropriate nitrogen concentration can promote the growth of stem and leaf, pigment synthesis, photosynthetic rate, and organic matter conversion efficiency, and effectively improve the quality and output of tomatoes (Fontes *et al.*, 1997). If tomato seedlings are deficient in nitrogen, plant growth is retarded, flower bud differentiation is low, stems and leaves are small, and growth and development are inhibited. If there is an excess of nitrogen, the plant grows wildly, with large leaves but stunted roots, leading to a decrease in resistance, an irreversible process (Baglieri *et al.*, 2014). Therefore, exploring the optimal nitrogen concentration of fertilization in the tomato seedling stage is the basis of the precision management of facility agriculture. According to the principle of plant physiology, plant leaves are the most sensitive to nitrogen, and the nitrogen content in the cultivation medium will cause the difference in pigment content in the leaves (Flores *et al.*, 2001). By detecting pigments in leaves, we can indirectly diagnose the nitrogen deficiency of plants. Visible/near-infrared (VIS/NIR) hyperspectral imaging (HIS) is an effective technique

for rapid and non-invasive analysis to obtain the required continuous spectral and image information. In recent years, using spectral analysis to obtain physiological information about crops has gradually become a research focus.

The reflectance of plant leaves is mainly determined by the surface properties and internal structure of the leaves, as well as by the concentration and distribution of chemical components. In the visible spectral range (400-750 nm), it is mainly the absorption of photosynthetic pigments in the leaves. In the short wavelength near infrared (750-1300 nm), there is no strong absorption feature, and the reflectance is determined by the discontinuous structure encountered in the leaves. The long wavelength near-infrared (1300-2500 nm) is mainly related to the absorption characteristics of water and other compounds (Yao *et al*., 2009). Various biochemical factors, such as the content of pigment and nutrients, have been found to affect the optical properties of tissues (Xu *et al*., 2007). Due to physiological stress, the spectra of plant leaves change in both the visible and near-infrared regions (Xu *et al*., 2007). The nitrogen concentration in the culture medium can directly affect the pigment content and its distribution in the leaves, which in turn affects the internal structure of the leaves. Owing to the large dimension of the hyperspectral detection data, the strategy of key band screening can reduce the amount of modeling data (Ouyang *et al*., 2021). However, it may lead to the loss of valid information, resulting in low prediction accuracy and insufficient generalization ability of the model. Therefore, HSI non-destructive measurement technology combined with machine learning algorithms is being more and more widely applied to the detection of agricultural information and agricultural products (Li *et al.*, 2018; Wang *et al.*, 2021).

The back propagation (BP) neural network algorithm has good self-learning, self-adaptation, generalization, and fault tolerance capabilities, but it has some drawbacks of slow convergence speed and local miniaturization (Zhang and Lou, 2021). To solve these problems, the optimization algorithm can optimize and reconstruct the BP network. At present, the classical population optimization algorithms include particle swarm optimization, ant colony optimization and bee colony optimization. The sparrow search algorithm (SSA) is a new population optimization algorithm designed with the idea of collaborative search of sparrow populations in foraging (Xue *et al.*, 2020). It has the advantages of fewer iterations, faster convergence, and higher search efficiency. Tuerxun *et al.* (2021) used SSA to optimize the penalty factor and kernel function parameters of the support vector machine (SVM) and established the SSA-SVM fault diagnosis model of the wind turbine. Zhang *et al.* (2022) put forward an aberration optimization method based on the SSA. The results show that the aberration combination distribution optimized by the SSA method is more remarkable than that under zero aberration. Fathy *et al.* (2022) proposed that SSA be applied to manage the operation of the power grid system in the best manner. The results show that SSA can effectively reduce the cost and emissions of power grid energy management and has good robustness. The above results show that the SSA has the advantages of fewer iterations, faster convergence speed, and higher search efficiency. At present, it has been widely applied in the fields of system diagnosis, image processing, mechanical analysis, and so on. However, compared with other optimization algorithms, it is easy to get into premature convergence due to the reduction of population diversity at the end of the iteration. Yuan *et al.* (2021) proposed to initialize the population using the gravity center reverse learning mechanism to make the population have a better spatial solution distribution. Gao *et al.* (2022) proposed to optimize the diversity of population initialization with 10 chaotic mapping, which accelerated the convergence rate and improved the convergence accuracy of the algorithm. Combined with the greedy strategy, the ability of the algorithm to deal with the locally optimal solution is improved, and each sparrow is fully utilized. Yan *et al.* (2021) proposed to improve SSA by adopting an iterative local search strategy. In the local search stage, the improved iterative local search strategy is adopted to improve the search accuracy and prevent missing the optimal solution.

In this study, chlorophyll a, chlorophyll b, chlorophyll and carotenoid in the leaves of tomato seedlings were researched. The nutrient solution was prepared with 10 kinds of nitrogen concentrations, and the leaves were picked for VIS/NIR hyperspectral collection. Aiming at the problems in the SSA, the logistic chaotic mapping strategy was proposed to improve population initialization and establish the logistic-SSA-BP prediction model. Different pretreatment methods were adopted for different pigments to improve the prediction accuracy of the model and to explore the optimum nitrogen concentration of the nutrient solution in the greenhouse tomato seedling stage and the prediction effect of the optimization algorithm.

Materials and Methods

Experimental design and sample collection

The experiment was carried out in the scientific greenhouse of the College of Agricultural Engineering, Shanxi Agricultural University, on November 12, 2021. Tomato seedlings (*Provence),* purchased from the seedling company, were grown in the transplanting pot (diameter of 23.8 cm and height of 19.5 cm) on coconut chaff. The nutrient solution was a water-soluble fertilizer formulated by the Netherlands Institute of Greenhouse Horticulture. A total of 10 nitrogen concentrations (nitrogen concentration range of 59.64~605.68 mg·L^{-1}, step length of 60 mg·L^{-1}, denote: N20, N40, N60, N80, N100, N120, N140, N160, N180, N200) were adjusted with urea. Ca^{2+} was supplemented with pure calcium fertilizer (Ca^{2+}≥94%), keeping the Ca^{2+} concentration consistent.

A total of 10 plants were cultivated at all concentrations, except for 5 plants at the N200 concentration, for a total of 95 plants. When tomato plants bloom by more than 50%, they are identified as reaching the flowering stage. Tomato leaves were sampled at the 'transplant-flowering' stage (December 13, 2021). At this time, the main stem stretched out 10-11 branches, and the tomato leaves were arranged according to Figure 1A. Among them, 9-7 branches were in the upper position, 6-4 branches were in the middle position, and 3-1 branches were in the lower position. The branches and leaves of tomatoes are "single-branch and multi-leaf". To obtain a sample of the corresponding leaf position as comprehensively as possible, refer to the marked parts for sampling in Figure 1B. According to the position of the leaves, a total of 1710 leaves (285 samples) of uniform size were collected (6 leaves for each sample). The samples were placed in sealed bags, numbered, and stored in an incubator filled with dry ice.

Hyperspectral imaging system

Leaf images were collected by the VIS/NIR hyperspectral scanning platform (Headwall Photonics, Boston, MA, USA). This system (Figure 1C) mainly includes 2 hyperspectral imagers, a lifting platform, a light source, a controller, a scanning platform controlled by a stepper motor, and a military computer. The resolution of the hyperspectral imager in the spectral range of 380-1000 nm

and 900-1700 nm is 0.727 nm and 4.715 nm, respectively. Due to the noise near the measuring range of the imager, a total of 794 variables in the spectral range of 430-900 nm and 950-1650 nm are intercepted as effective variables for modeling. System parameter settings: the movement speed was 2.721 mm·s^{-1}, the advance stroke was 100 mm, and the distance from the lens to the leaf was 28 cm, to obtain a clear and distortion-free image.

To reduce the interference of the system light source and dark current on the image, the black-and-white correction of the HSI is carried out according to Eq. 1. Firstly, the dust and impurities on the leaf surface were washed clean with deionized water. Secondly, use filter paper to absorb the moisture on the surface. Finally, put 3 leaves on the stage. Two hyperspectral images were collected from each sample.

$$R = \frac{R_0 - R_b}{R_w - R_b} \quad (1)$$

where R represents the image obtained after correction, R_0 represents the original image, R_w represents the whiteboard calibration image (>99.9% reflectance), and R_b represents the dark background calibration image (<0% reflectance).

Determination of leaf pigment content

After collecting spectral images of the samples, the contents of chlorophyll a, chlorophyll b, chlorophyll and carotenoid were measured by a spectrophotometer (Wellburn, 1994). After removing the leaf veins, each sample was cut into small pieces of about 2×2 mm, mixed, weighed 0.2 g, and poured into a test tube. First, 10 mL of anhydrous ethanol was added, shaken well, and extracted in the dark room for 10 hours. Then, 10 mL of anhydrous ethanol was extracted for 14 hours. Finally, 5 mL of anhydrous ethanol in a constant volume. The absorbance values of the prepared pigment extract were measured at wavelengths of 665 nm, 649 nm, and 470 nm, respectively. Each sample was repeated 3 times. The pigment content is calculated according to Eqs. 2-5:

$$Chla = 13.95 \times A_{665} - 6.88 \times A_{649} \quad (2)$$

$$Chlb = 24.96 \times A_{649} - 7.32 \times A_{665} \quad (3)$$

$$Chll = Chla + Chlb \quad (4)$$

$$Caro = \frac{1000 \times A_{470} - 2.05 \times Chla - 114.8 \times Chlb}{245} \quad (5)$$

where *Chla*, *Chlb*, *Chll* and *Caro* represent chlorophyll a, chlorophyll b, chlorophyll and carotenoid content, respectively. A_{665}, A_{649} and A_{470} represent the absorbance of the solution to be measured wavelengths at 665 nm, 649 nm and 470 nm, respectively.

Logistic improved sparrow search algorithm

SSA is based on the cooperative layout of the sparrow population in the form of predators, followers, and warnings in the process of finding food. The population simulates sparrow individuals in matrix form, expressed as follows (Eq. 6):

$$X = \begin{Bmatrix} x_1^1 & x_1^2 & \text{L} & x_1^d \\ x_2^1 & x_2^2 & \text{L} & x_2^d \\ \text{M} & \text{M} & \text{O} & \text{M} \\ x_n^1 & x_n^2 & \text{L} & x_n^d \end{Bmatrix} \quad (6)$$

where X^d_n represents the position of sparrows, n represents the number of sparrows, and d represents the dimension of variable

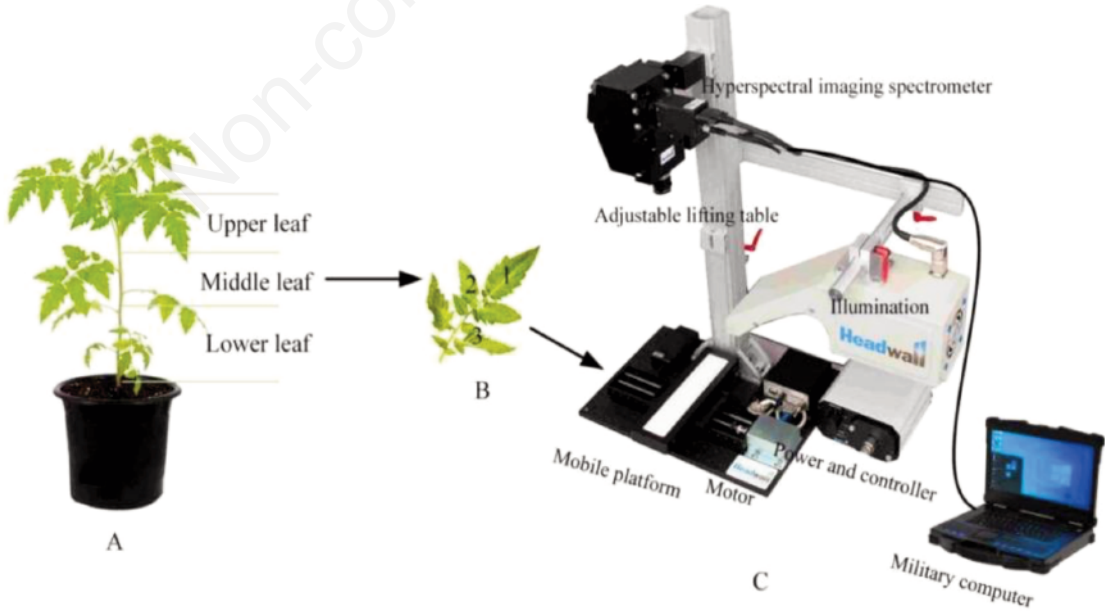

Figure 1. Sampling rules and hyperspectral imaging system. **A**) Main stem and tomato leaves; **B**) marked parts for sampling;) system including 2 hyperspectral imagers, an adjustable lifting platform, illumination, a controller, a mobile platform controlled by a stepper motor, and a military computer.

space. The population fitness value is expressed as (Eq. 7):

$$F_x=\begin{Bmatrix} f\left(\left[x_1^1 \quad x_1^2 \quad \mathrm{L} \quad x_1^d\right]\right) \\ f\left(\left[x_2^1 \quad x_2^2 \quad \mathrm{L} \quad x_2^d\right]\right) \\ \mathrm{L} \\ f\left(\left[x_n^1 \quad x_n^2 \quad \mathrm{L} \quad x_n^d\right]\right) \end{Bmatrix} \tag{7}$$

where F_x represents the individual fitness value of the sparrow.

Aiming at the existing problems of SSA, the logistic chaotic mapping strategy was proposed to initialize the population. It increases the population diversity, makes the population distribution more uniform, and improves the initial solution quality of the population, thus upgrading the optimization efficiency and ergodicity of the algorithm. Its expression is as follows (Eq. 8):

$$X(t+1) = m \times X(t) \times (1 - X(t)) \tag{8}$$

where $X(t+1) \in [0, 1]$. $m \in [0, 4]$ represent logistic parameters. According to relevant literature (Kanso and Smaoui, 2009), the closer to 4, the uniform distribution of X in the [0, 1] region, that is, the complete chaotic state, so $\mu = 4$; t is the iteration time step.

The role of predators is to provide searching direction for the foraging. In the process of searching for food (optimization), predators with higher adaptability get food first, so the predator has a larger search space. In the iterative optimization process, the position of the predator is expressed as (Eq. 9):

$$X_{i,j}^{t+1}=\begin{cases} X_{i,j} \times exp\left(-\dfrac{i}{\alpha \times inter_{max}}\right) & R_2 < ST \\ X_{i,j}^t + Q \times L & R_2 \geq ST \end{cases} \tag{9}$$

where $X_{i,j}$ represents the spatial position information of the i-th sparrow in the j-th dimension. t represents the current iteration number. $\alpha \in (0, 1]$ random number. $inter_{max}$ represents the maximum number of iterations (set to 20). Q is a random number that obeys the normal distribution. L represents a 1×d matrix with all elements of 1. $R_2 \in [0, 1]$ is the warning value. $ST \in [0.5, 1]$ is a safe value (set to 0.8). When $R_2 < ST$, the search space is safe, and predators can widely search for food. When $R_2 \geq ST$, the predator finds an alarm in the search space, and alert the community to fly away quickly. During foraging, the follower watches the predator in real time. If the predator finds food, the follower quickly occupies the best position for foraging, and its position is expressed as (Eq. 10):

$$X_{i,j}^{t+1}=\begin{cases} Q \times exp\left(\dfrac{X_{worst}^t - X_{i,j}^t}{i^2}\right) & i > 0.5n \\ X_p^{t+1} + \left|X_{i,j}^t - X_p^{t+1}\right| \times \left(A^T\left(AA^T\right)^{-1}\right) \times L & i \leq 0.5n \end{cases} \tag{10}$$

where X_{worst} represents the worst position in the current population space. X_p indicates the best position of the current predator. A is a 1×d matrix expressed as 1 and -1. When $i > 0.5n$, followers with low fitness values cannot catch food, so they need to search other areas for food. When $i \leq 0.5n$, followers will forage in the best position.

A certain number of sparrows in the population are randomly selected as warnings agents to assume the warning function. Its position is expressed as (Eq. 11):

$$X_{i,j}^{t+1}=\begin{cases} X_{best}^t + \beta \times \left|X_{i,j}^t - X_{best}^t\right| & f_i > f_g \\ X_{i,j}^t + K \times \left(\dfrac{\left|X_{ij}^t - X_{worst}^t\right|}{\left(f_i - f_w\right) + \varepsilon}\right) & f_i = f_g \end{cases} \tag{11}$$

where X_{best} represents the best position of the current population space. β represents the step size control parameter, which is a random number with standard normal distribution. The random number of $k \in [-1, 1]$. f_i is the individual fitness value of the current population. f_g and f_w are the best and worst fitness values of the current population space, respectively. ε represents a constant (set to 10^{-8}) to avoid the model being meaningless.

Logistic-sparrow search algorithm-back propagation model and evaluation index

In the logistic-SSA optimization algorithm, the ratio of predator, follower and warning was set to 0.7:0.3:0.2. When the maximum number of iterations was reached, the optimal fitness value outputs the optimal solution, so as to optimize the weights and thresholds of the BP neural network. The error BP was used for iterative training, and the preset effect was achieved.

The logistic-SSA-BP model flow is shown in Figure 2. The node number of the input layer, hidden layer and output layer of the BP neural network was 794, 30 and 1, respectively. This network has 2 layers, where the number of thresholds and weights in the first layer were set to 23,820 and 30, and the thresholds and weights in the second layer were set to 30 and 1. The settings of each parameter are shown in Table 1. In this paper, the fitness function is expressed as (Eq. 12):

$$f = argmin(mse(TE)+MSE(PE)) \tag{12}$$

where TE and PE are training set and prediction set errors, respectively, and mse is the mean square error function. The smaller the mean square error, that is, the smaller the fitness function value, indicating the higher the prediction accuracy of the model.

The root mean squared error (RMSE), coefficient of determination (R^2) and relative percent deviation (RPD) evaluated the prediction potential and performance of the models. The smaller the RMSE, the higher the prediction accuracy. The closer R is to 1, the higher the prediction accuracy is. When RPD>2, it indicates that the model achieves a better prediction effect on the index. When 1.4<RPD<2, it indicates that the model can predict the index to a certain extent. When RPD<1.4, it indicates that the model cannot predict the index (Wang *et al.*, 2019) (Eqs. 13-15).

Table 1. Parameters of the back propagation neural network model.

Parameters	Network function and parameter size
Hidden layer function	Logsig
Output layer function	Purelin
Training function	Traingdx
Maximum steps of training	300
Training accuracy	0.01
Learning rate	0.1

$$RMSE=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2} \quad (13)$$

$$R^2=1-\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2} \quad (14)$$

$$RPD=\sqrt{\frac{n\sum_{i=1}^{n}(y_i-\bar{y})^2}{(n-1)\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}} \quad (15)$$

where n represents the number of samples, y_i represents the actual value of physical and chemical experiments of samples, $\hat{y}_i$ represents the predicted value of the sample model, and represents the average value of y_i.

Results and Discussion

Analysis on the change law of leaf pigment content

The concentration of nitrogen in nutrient solutions affects the absorption and assimilation potential of crops. The effects of different concentrations of nutritional solution nitrogen on the pigment content of tomato leaves are significantly different, as shown in Figure 3A. When the nitrogen concentration in the nutrient solution (N100) was 302.84 mg·L^{-1}, the pigment content in the leaves reached its maximum. When the concentration was less than 302.84 mg·L^{-1}, with the increase in concentration, the maximum growth rates of chlorophyll a, chlorophyll b, chlorophyll and carotenoid reached 14.29%, 20.75%, 17.45%, and 17.85%, respectively. When the concentration was greater than 302.84 mg·L^{-1}, with the increase of concentration, the maximum decrease rates of chlorophyll a, chlorophyll b, chlorophyll and carotenoid reached

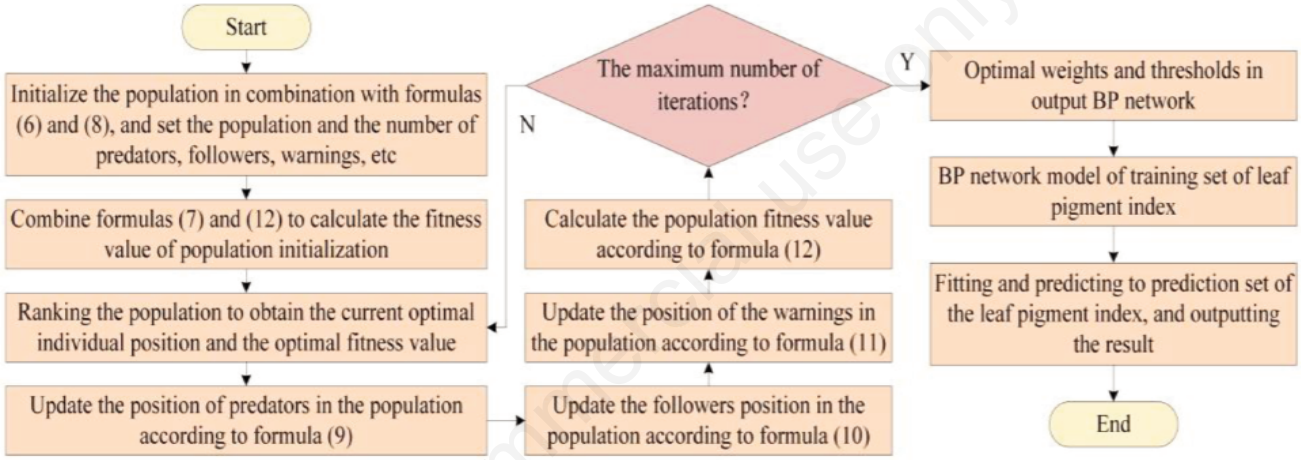

Figure 2. Flow chart of logistic-sparrow search algorithm-back propagation neural network.

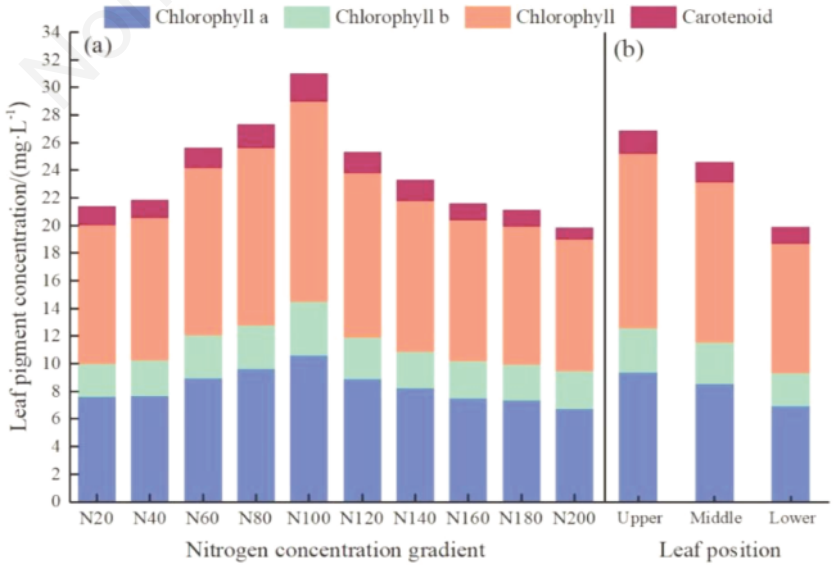

Figure 3. Effects of different nitrogen concentration on different positions of tomato leaves.

16.35%, 21.74%, 17.78% and 27.40%, respectively. This is because the optimum concentration of nitrogen can maintain the balance of nutrients in plants. Too high or too low will upset this balance and inhibit the growth and development of crops. The results showed that the inhibitory effect of high concentrations was stronger than that of low concentrations. At the same time, nutrients are easily transferred during the process of crop growth and development, resulting in differences in pigment contents in tomato leaves, as shown in Figure 3B. When the leaf was from the upper to the middle, the decrease rates of chlorophyll a, chlorophyll b and chlorophyll were the same, about 7.92%, while the decrease rate of carotenoid content was 13.5%. When the leaf position was from the middle to the low, the decrease rates of chlorophyll a, chlorophyll b, chlorophyll and carotenoid were the same, reaching about 19.12%. This indicates that the plant is in a vigorous growth stage.

Extraction of spectral data and spectral response characteristics of leaves

As there are many pixels in the image collected by the spectral imager, it is difficult to extract and process these data. To simplify this process, batch extraction and processing software of HSI was developed for SpectralView software. Use the ellipse model to select the center coordinate, x/y semi-axis length, and x/y axis interval of the image region of interest (ROI). Generate a coordinate matrix of the ROI on the target image. Importing images through SpectralView follows the principle of "from left to right, from top to bottom", and actively extracts ROI information from images. Then, the spectral information of different bands of each pixel was output in batches as an arithmetic average, as shown in Figure 4. Because the resolutions of the hyperspectral imager are different, the number of pixels taken is also different. At 430-900 nm and 950-1650 nm, 18810 and 2130 pixels were extracted from each leaf, *i.e.*, 112860 (18810×6) and 12780 (2130×6) pixels were extracted from each sample, and the average spectral reflectance was output in batches. Different nitrogen concentrations of nutrient solutions and pigment contents in different leaves are quite different, which leads to great differences in the spectral reflectance of samples treated with different treatments. Figure 4 shows the response curves of leaves at 430-900 nm and 950-1650 nm. There is a strong absorption area of chlorophyll and carotenoid at 490 nm (Wang *et al.*, 2019). There is a strong reflection region of chlorophyll at 550 nm. There is a strong absorption area of chlorophyll at 680 nm. There are obvious wave point fluctuations in 750-900 nm, that are related to the absorption of water or oxygen (Schmilovitch *et al.*, 2014). There is a second harmonic generation (SHG) of free O-H groups stretching vibration at 980 nm (Kostin *et al.*, 2015). There is the harmonic generation and combined spectral band of the C-H groups stretching vibration at 1200 nm (Sankar *et al.*, 2010). There is the SHG of the O-H bond in water molecules at 1450 nm (Huang *et al.*, 2013).

Dataset partition

The SPXY algorithm, which comprehensively considers the differences between the spectral characteristics and physicochemical properties of the samples, was used to divide the dataset (Galvao *et al.*, 2005). As shown in Table 2, the sample mean values of the training set and the prediction set were almost the same, indicating the rationality of the dataset partition.

Prediction of leaf pigment content

For chlorophyll a, chlorophyll b, chlorophyll and carotenoid, Figure 5 and Table 3 show the changes in iterations and fitness values of logistic-SSA and SSA. From the iterative results, it can be seen that logistic-SSA obviously improves the quality of the initial solution of the population, and effectively increases the diversity of the population. With the increase in iterations, the early convergence rate was obviously improved, which indicates that the global searching capability of the population was optimized. Although the number of optimization iterations has increased, the optimization accuracy has been improved; that is, the optimal fitness value is obviously lower than that of SSA. The chlorophyll a, chlorophyll b, chlorophyll and carotenoid decreased by 0.03, 0.02, 0.01, and 0.02, respectively. The results show that the logistic chaotic mapping strategy can effectively improve the searching performance of

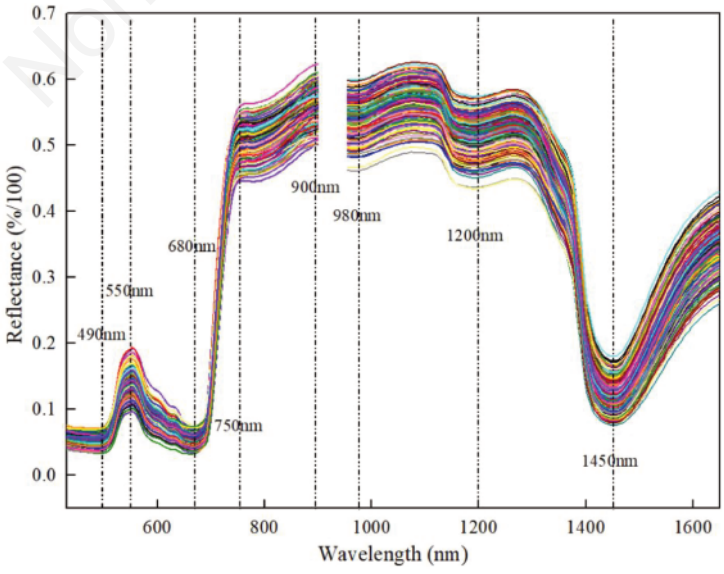

Figure 4. Spectral response curve of tomato leaves.

the SSA population. In the process of spectral scanning, factors such as leaf state, measurement environment, instrument response, *etc*. will cause the spectral baseline to shift, thus affecting the prediction accuracy of the model. The Savitsky-Golay (S-G), standard normal variable (SNV), and multiple scattering correction (MSC) were used to eliminate multiple linear errors caused by leaf surface scattering, optical path change, and interference among components. Through logistic-SSA iterative optimization, the weights and thresholds of the BP neural network from the input layer to the hidden layer and from the hidden layer to the output layer are optimized, respectively, to obtain the optimal solution for optimal fitness output. The contents of chlorophyll a, chlorophyll b, chlorophyll, and carotenoid were predicted through the reconstruction of the BP network. The prediction accuracy results of different pretreatment methods for the raw data are shown in Table 4. For chlorophyll a, the optimal prediction was obtained by combining

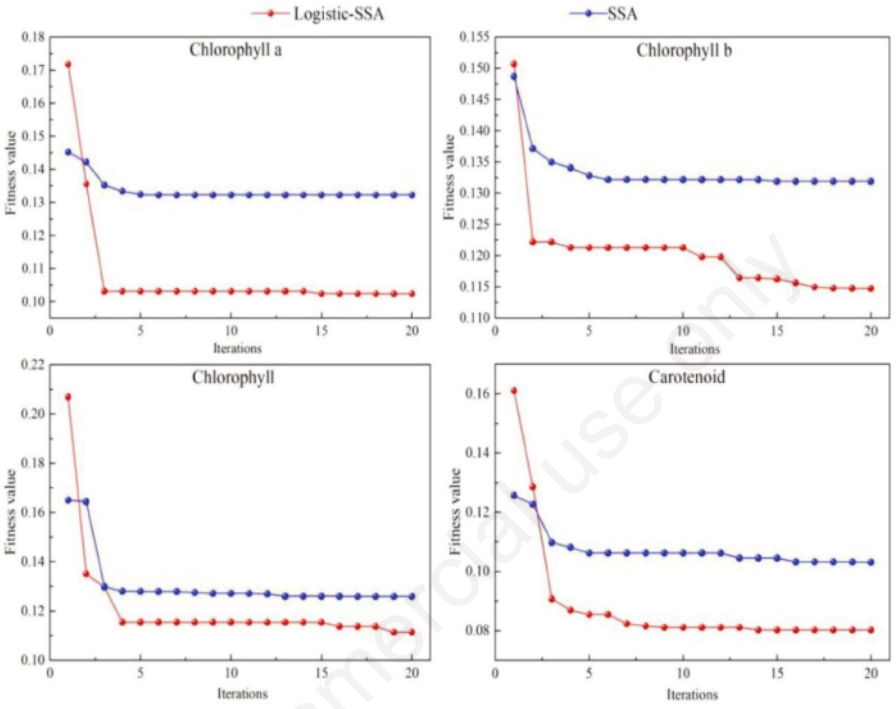

Figure 5. Comparison of iterative curves between logistic-sparrow search algorithm and sparrow search algorithm.

Table 2. Statistical results of leaf pigment content dataset.

Sample category	Pigments	Number	Maximum	Minimum	Mean	SD
Training set	Chlorophyll a	213	14.72	3.37	8.32	2.18
	Chlorophyll b		4.89	1.24	2.86	0.74
	Chlorophyll		19.62	4.61	11.15	2.89
	Carotenoid		3.23	0.60	1.42	0.44
Prediction set	Chlorophyll a	72	12.43	5.26	8.31	1.69
	Chlorophyll b		4.38	1.80	2.90	0.57
	Chlorophyll		16.63	6.96	11.31	2.22
	Carotenoid		2.14	0.90	1.40	0.34

SD, standard deviation.

Table 3. Optimal iteration times and fitness value of logistic-sparrow search algorithm and sparrow search algorithm for leaf pigment index.

Pigments	Logistic-SSA			SSA		
	1st	Optimal number of iterations	Best fitness value	1st	Optimal number of iterations	Best fitness value
Chlorophyll a	0.17	15	0.10	0.15	6	0.13
Chlorophyll b	0.15	17	0.11	0.14	15	0.13
Chlorophyll	0.21	19	0.11	0.17	13	0.13
Carotenoid	0.16	14	0.08	0.13	16	0.10

SSA, sparrow search algorithm; 1st, initial optimal fitness value of the population.

Table 4. Comparison of prediction accuracy results of different pretreatment methods.

Pigments	Pretreatment method	RMSE	R^2	RPD
Chlorophyll a	Raw	0.82	0.70	1.84
	S-G	0.77	0.74	1.97
	S-G+SNV+MSC	0.77	0.77	2.08
Chlorophyll b	Raw	0.34	0.57	1.52
	S-G	0.31	0.63	1.65
	S-G+MSC	0.30	0.66	1.71
Chlorophyll	Raw	0.89	0.77	2.09
	S-G	0.86	0.79	2.18
	S-G+SNV	0.88	0.81	2.28
Carotenoid	Raw	0.21	0.66	1.71
	S-G	0.18	0.73	1.92
	S-G+SNV	0.14	0.75	2.00

RMSE, root mean squared error; R^2, coefficient of determination; RPD, relative percent deviation; Raw, raw data; S-G, Savitsky-Golay; SNV, standard normal variable; MSC, multiple scattering correction.

S-G, SNV and MSC, with RMSE, R^2 and RPD of 0.77, 0.77 and 2.08, respectively. For chlorophyll b, the optimal prediction was obtained by combining S-G and MSC, with RMSE, R^2 and RPD of 0.30, 0.66 and 1.71, respectively. For chlorophyll, the optimal prediction was obtained by combining S-G and SNV, with RMSE, R^2 and RPD of 0.88, 0.81 and 2.28, respectively. For carotenoid, the optimal prediction was obtained by combining S-G and SNV, with RMSE, R^2 and RPD of 0.14, 0.75 and 2.00, respectively. In summary, it can be concluded that logistic-SSA-BP can achieve better prediction for chlorophyll a, chlorophyll, and carotenoid, and to a certain extent for chlorophyll b. The model can be used to predict chlorophyll a, chlorophyll and carotenoid. It shows that the model has some generalizability.

Conclusions

In this paper, the contents of chlorophyll a, chlorophyll b, chlorophyll and carotenoid in different leaf positions of tomato seedlings under nutrient solution cultivation with different nitrogen concentrations were taken as research indexes. The prediction model of pigment content is based on hyperspectral technology and machine learning algorithms. To solve the problems of the SSA in the optimization process, the logistic chaotic mapping strategy was proposed to initialize the population. The logistic-optimized SSA was used to optimize the weights and thresholds of the BP neural network to establish the pigment content prediction model. The following conclusions were mainly obtained. Firstly, there are significant differences in the effects of different nitrogen concentrations on the pigment content of tomato leaves. When the nitrogen concentration is 302.84 $mg \cdot L^{-1}$, the pigment content in leaves is the highest, and the inhibitory effect of a high concentration is stronger than that of a low concentration. It could provide data support for precise management of fertilizer in facility agriculture and then improve the quality and output of tomatoes.

Moreover, the logistic chaotic mapping strategy was adopted to initialize the SSA population, which effectively improves the initial solution of the population. It reduces the optimal fitness of the population and improves the convergence speed and optimization efficiency of the algorithm. Thereby improving the searchability of the population.

Lastly, different combinations of pretreatment methods for different pigments significantly improved the accuracy of the logistic-SSA-BP prediction model. The RMSE, R^2 and RPD of chlorophyll a were 0.77, 0.77 and 2.08, respectively. The RMSE, R^2 and RPD of chlorophyll b were 0.30, 0.66 and 1.71, respectively. The RMSE, R^2 and RPD of chlorophyll were 0.88, 0.81 and 2.28, respectively. The RMSE, R^2 and RPD of carotenoid were 0.14, 0.75 and 2.00, respectively. The results show that the optimization algorithm has a good prediction effect and universality, which provides a theoretical basis for the application of crop or agricultural product information prediction.

References

Baglieri A., Cadili V., Monterumici C.M., Gennari M., Tabasso S., Montoneri E., Nardi S., Negre M. 2014. Fertilization of bean plants with tomato plants hydrolysates. Effect on biomass production, chlorophyll content and N assimilation. Sci. Hortic.-Amsterdam 176:194-9.

Fathy A., Alanazi T.M., Rezk H., Yousri D. 2022. Optimal energy management of micro-grid using sparrow search algorithm. Energy Rep. 8:758-73.

Flores P., Carvajal M., Cerdá A., Martínez V. 2001. Salinity and ammonium/nitrate interactions on tomato plant development, nutrition, and metabolites. J. Plant Nutr. 24:1561-73.

Fontes P.C.R., Pereira P.R.G., Conde R.M. 1997. Critical chlorophyll, total nitrogen, and nitrate□nitrogen in leaves associated to maximum lettuce yield. J. Plant Nutr. 20:1061-8.

Galvao R.K.H., Araujo M.C.U., José G.E., Coelho Pontes M.J., Silva E.C., Bezerra Saldanha T.C. 2005. A method for calibration and validation subset partitioning. Talanta 67:736-40.

Gao B., Shen W., Guan H., Zheng L., Zhang W. 2022. Research on multistrategy improved evolutionary sparrow search algorithm and its application. IEEE Access 10:62520-34.

Huang H., He R., Yao W., Lin Z. 2013. Noncentrosymmetric mixed-cation borate: crystal growth, structure and optical properties of $Cs_2Ca[B_4O_5(OH)_4]_2 \cdot 8H_2O$. J. Cryst. Growth. 380:176-81.

Kanso A., Smaoui N. 2009. Logistic chaotic maps for binary numbers generations. Chaos Soliton. Fract. 40:2557-68.

Kostin G.A., Borodin A.O., Mikhailov A.A., Kuratieva N.V., Kolesov B.A., Pishchur D.P., Woike T., Schaniel D. 2015. Photocrystallographic, spectroscopic, and calorimetric analysis of light-induced linkage NO isomers in $[RuNO(NO_2)_2$(pyri-

dine)$_2$OH]. Eur. J. Inorg. Chem. 2015:4905-13.
Li J., Shi Y., Veeranampalayam-Sivakumar A.N., Schachtman D.P. 2018. Elucidating sorghum biomass, nitrogen and chlorophyll contents with spectral and morphological traits derived from unmanned aircraft system. Front. Plant Sci. 9:1406.
Ouyang Q., Wang L., Park B., Kang R., Chen Q. 2021. Simultaneous quantification of chemical constituents in matcha with visible-near infrared hyperspectral imaging technology. Food Chem. 350:129141.
Sankar S., Manikandan M.R., Gopal Ram S.D., Mahalingam T., Ravi G. 2010. Gel growth of α and γ glycine and their characterization. J Cryst. Growth 312:2729-33.
Schmilovitch Z., Ignat T., Alchanatis V., Gatker J., Ostrovsky V., Felföldi J. 2014. Hyperspectral imaging of intact bell peppers. Biosyst. Eng. 117:83-93.
Tuerxun W., Chang X., Hongyu G., Zhijie J., Huajian Z. 2021. Fault diagnosis of wind turbines based on a support vector machine optimized by the sparrow search algorithm. IEEE Access 9:69307-15.
Wang S., Guan K., Wang Z., Ainsworth E.A., Zheng T., Townsend P.A., Li K., Moller C., Wu G., Jiang C. 2021. Unique contributions of chlorophyll and nitrogen to predict crop photosynthetic capacity from leaf spectroscopy. J. Exp. Bot. 72:341-54.
Wang W., Li Z., Wang C., Zheng D., Du H. 2019. Prediction of available potassium content in cinnamon soil by hyperspectral imaging. Spectrosc. Spect. Anal. 39:1579-85.
Wang Y., Hu X., Jin G., Hou Z., Ning J., Zhang Z. 2019. Rapid prediction of chlorophylls and carotenoids content in tea leaves under different levels of nitrogen application based on hyperspectral imaging. J Sci Food Agric. 99:1997-2004.
Wellburn A.R. 1994. The spectral determination of chlorophylls a and b, as well as total carotenoids, using various solvents with spectrophotometers of different resolution. J. Plant Physiol. 144:307-13.
Xu H.R., Ying Y.B., Fu X.P., Zhu S.P. 2007. Near-infrared spectroscopy in detecting leaf miner damage on tomato leaf. Biosyst. Eng. 96:447-54.
Xue J., Shen B. 2020. A novel swarm intelligence optimization approach: sparrow search algorithm. Syst. Sci. Control Eng. 8:22-34.
Yan S., Yang P., Zhu D., Zheng W., Wu F. 2021. Improved sparrow search algorithm based on iterative local search. Comput. Intell. Neurosci. 2021:6860503.
Yao J.S., Yang H.Q., He Y. 2009. Nondestructive detection of rape leaf chlorophyll level based on Vis/NIR spectroscopy. J. Zhejiang Univ.-Sc A. 35:433-8.
Yuan J, Zhao Z, Liu Y, He B., Wang L., Xie B., Gao Y. 2021. DMPPT control of photovoltaic microgrid based on improved sparrow search algorithm. IEEE Access 9:16623-9.
Zhang D., Lou S. 2021. The application research of neural network and BP algorithm in stock price pattern classification and prediction. Future Gener. Comp. Sy. 115:872-9.
Zhang S., Zhang L., Gai T., Xu P., Wei Y. 2022. Aberration analysis and compensate method of a BP neural network and sparrow search algorithm in deep ultraviolet lithography. Appl. Opt. 61:6023-32.