NUTRIENT DEFICIENCY DIAGNOSIS IN WHOLE HYDROPONIC LETTUCE BASED ON RANDOM FOREST
/
基于随机森林算法的整株水培生菜缺素诊断

Xinyu ZHANG[1,2]; Dandan CAO[1,2]; Minghui WANG[1,3]; Gongpei CUI[1]; Yinggang SHI[1]; Yongjie CUI *[1,3] [1]
[1] College of Mechanical and Electronic Engineering, Northwest A&F University, Yangling / China;
[2] Key Laboratory of Agricultural Internet of Things, Ministry of Agriculture and Rural Affairs, Yangling / China;
[3] Shaanxi Key Laboratory of Agricultural Information Perception and Intelligent Service, Yangling / China
Tel: +86 13720581232; E-mail: agriculturalrobot@nwafu.edu.cn
DOI: https://doi.org/10.35633/inmateh-68-08

Keywords: *hydroponic lettuce, machine learning, feature extraction, deficiency diagnosis*

ABSTRACT
The phenotypic information of lettuce leaves can well reflect its health. In order to diagnose the nutrient deficiency types of hydroponic lettuce accurately, non-destructively and quickly in the mid-growth stage, a method for diagnosis of whole lettuce based on random forest algorithm (RF) was proposed. The images of lettuce under four different conditions, K-deficiency, Ca-deficiency, N-deficiency and Normal, were collected and segmented by Extra-green algorithm. Then, features of color, texture and shape were extracted. A RF classification model for the hydroponic lettuce nutrient deficiency diagnosis was constructed and compared with support vector machine (SVM) and back propagation neural network (BP). RF had the best classification effect among the three methods. The overall classification accuracy was 86.32%, Kappa coefficient was 0.82, and it can provide a basis for the prevention and remedies of lettuce deficiency and the scientific management of nutrient solutions.

摘要
生菜叶片的表型信息可以很好地反映其健康状况。为了准确、无损、快速地诊断水培生菜生长中期的缺素类型，以整株水培生菜叶片图像为研究对象，提出一种基于随机森林算法(RF)的缺素诊断方法。采集缺钾、缺钙、缺氮以及正常 4 种生长条件下的生长中期的水培生菜缺素图像，利用超绿算法分割得到整株水培生菜叶片图像，并提取其颜色、纹理和形状的特征。基于 RF 建立水培生菜缺素诊断模型，并与支持向量机(SVM)和 BP 神经网络(BP)进行对比试验。三种方法中 RF 的分类效果最好,总体分类准确率为 86.32%，Kappa 系数为 0.82，可为水培生菜缺素症防治及采取补救措施以及营养液的科学管理提供依据。

INTRODUCTION
Nutrient solution is the key to the health of crops in the nutrient solution hydroponic plant factory *(He, 2018; Germer et al., 2011; Fang, 2016)*. The difference in its concentration can lead to nutrient deficiency or excess, which will cause the quality and yield of the crop to decline, and even cause disease. In actual agricultural production, the lack of nutrient leads to abnormal growth of crops or even diseases, which is called deficiency disease *(Gillespie et al., 2020; Zhao, 2014)*. Lettuces are one of the most widely planted crops in the hydroponic mode of plant factories; it has high nutritional value and economic benefits. Nutrient deficiency will seriously affect the harvest and sale of lettuce *(Yang et al., 2015)*. Therefore, a timely and accurate method for the diagnosis of nutrient deficiency can provide a basis for the prevention and treatment of hydroponic lettuce nutrient deficiency, and it is of great significance in actual production.

It is a direct and effective method to judge the health status of crops through their leaf feature information. Since the 1980s, many scholars have diagnosed the nutrition and disease of crops by technology of image processing and machine learning, and have made some achievements.

Many scholars have used machine vision to judge the nutritional status of crop leaves by extracting features such as color, texture, and shape *(Mao et al., 2003; Xu, 2013; Flores et al., 2020; Zhang and You, 2017)*. Lu, B. et al. proposed a diagnosis method that combines hyperspectral technology and image color and texture features to identify the disease type and period of lettuce accurately *(Lu et al., 2018)*.

[1] *Xinyu Zhang, M.S. Stud. Eng.; Dandan Cao, M.S. Stud. Eng.; Minghui Wang, M.S. Stud. Eng.; Gongpei Cui, Ph.D. Stud. Eng.; Yinggang Shi, Assoc. Prof. M.S. Eng.; Yongjie Cui, Prof. Ph.D. Eng.*

Yu. et al. established a regression model using hyperspectral technology combined with deep learning to quickly detect the nitrogen content in rape *(Yu et al., 2018)*. Wei, W. S. et al. focused on the development of non-destructive online detection system for leaf vegetables quality detection based on machine vision *(Wei et al., 2018)*. *Mao. et al.* established a predictive model for the nitrogen content of lettuce based on the canopy and spectral reflectance, image features and variable wavelength information by using machine learning *(Mao et al., 2015). Zhang, B. et al.* collected color images of lettuce leaves with different nitrogen levels, extracted their texture and color features, filtered and optimized the image feature vectors, and established a lettuce nitrogen content prediction model with an average error rate of less than 13% *(Zhang et al., 2018)*.

At present, researches on crop nutrition diagnosis mainly focus on cash crops such as rice and rape. There are few studies on the diagnosis of lettuce deficiency. In the existing leaf vegetable disease identification, the leafy vegetables at the mature stage or single leaves are mostly used, which may cause destructive sampling, and is not conducive to remedies. In this study, images of the whole hydroponic lettuce leaves in the mid-growth stage were taken as the research object, collecting images of hydroponic lettuce under K-deficiency, Ca-deficiency, N-deficiency, and Normal. And then, the color, texture and shape features were extracted and randomized. The random forest constructed a diagnosis model, and compared the effects of different methods. The study provided a fast, accurate and non-destructive method for diagnosing the nutrient deficiency of hydroponic lettuce.

MATERIALS AND METHODS

Image acquisition

In November 2020, the planting environment of a plant factory was simulated in an intelligent artificial climate room, and a planting experiment of hydroponic lettuce was designed to obtain images of lettuce in different nutrient deficiency conditions (Shown in Figure 1). The experiment selected the butter lettuce seeds produced by Chinese Vegetable Seed Technology (Beijing) Co.Ltd. Sow the seeds in a seedling tray covered with ceramics and place 2 seeds in each seedling hole. When the lettuce grows to 6 leaves and 1 heart, pull out the seedlings and wrap the roots with a sponge, then, plant them in the cultivation trough. The knop hydroponic formula was used to cultivate 4 types of hydroponic lettuce, which are K-deficiency, Ca-deficiency, N-deficiency and Normal. Throughout the experiment, all environmental parameters of the intelligent artificial climate room were installed referring to the actual conditions of the plant factory, and the pH, EC and liquid temperature of the nutrient solution were monitored continuously to ensure the best level of the planting conditions, so that the growing environment for hydroponic lettuce had always been best. The specific parameters' range is shown in Table 1.

Fig. 1 - Planting experiment of hydroponic lettuce in intelligent artificial climate room

Table 1

Parameters of hydroponic lettuce planting condition

Type	Name	Ranges
Solution Parameters	pH	5.5-6.5
	EC	2-3 dS/m
	Solution Temperature	15-20 ºC
Environmental Parameters	Environmental Temperature	Day 22 ºC Night 18 ºC
	Humidity	70-80% RH
	Illumination Time	24h
	Illumination Type	Red and Blue
	Concentration of CO_2	800 ppm

The hydroponic lettuce images for the experiment were collected by the Lifecam Studio produced by Microsoft. The camera uses a CMOS sensor and it can be connected to a laptop through a USB interface with a maximum resolution of 1280x720. The photo box was built with PVC boards to simulate the shading environment, and that can avoid the influence of red and blue light. The camera was fixed directly above the inside of the photo box, and the vertical distance from the lettuce plant is 50 cm. The white LED light strip was used as a supplementary light source to provide suitable brightness. The entire growth cycle of hydroponic lettuce after planting is generally 30-40 days. In this experiment, there were 36 lettuces in 4 categories including K-deficiency, Ca-deficiency, N-deficiency, and Normal. 1440 lettuce images were collected for follow-up research in the mid-growth stage, that is, within 10-20 days after planting.

It can be seen from Figure 2 that, compared with normal hydroponic lettuce, the overall growth of hydroponic lettuce slowed down in the nutrient-deficient state. The color of K-deficient leaves turns yellow and lighter, the leaves are slightly shrunken, and the tip is accompanied by brown spots; the Ca-deficient leaves have dark brown edges, and the leaves are curled; the N-deficient leaves are smaller, the overall color is lighter, and the stems are slender and soft.

(a) K-deficiency *(b) Ca-deficiency* *(c) N-deficiency* *(d) Normal*

Fig. 2 - Some hydroponic lettuce samples of different nutrient-deficient types under the same growth period

Four kinds of lettuce leaves with different nutrient-deficient types showed different characteristics in color, texture, and shape. Afterwards, pre-treatment and multi-feature extraction were performed on the images of hydroponic lettuce, and recognizing the images of different kinds of lettuce based on the three methods of back propagation neural network (BP), support vector machine (SVM) and random forest algorithm (RF). The classification accuracy of the three models was compared and the best one was selected to realize the diagnosis of nutrient deficiency types. The overall image processing flow is shown in Figure 3.

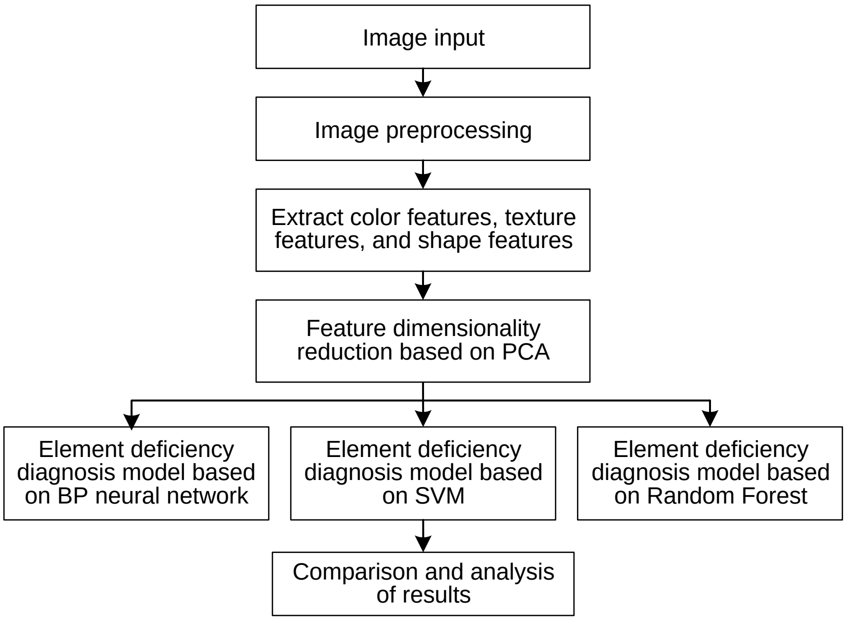


Fig. 3 – The overall flow of image processing

Image feature extraction and optimization

Before the nutrient deficiency diagnosis of the hydroponic lettuce, image enhancement, image segmentation and other pretreatment were performed on the images, and then extracted the color, texture, and shape of the whole lettuce image.

The principal component analysis (PCA) was used to analyze and optimize the all feature data, and the effective parameters were used for the establishment of deficiency diagnosis model.

Image pre-treatment

The pre-treatment of the image can reduce or eliminate irrelevant information of the original image, while retaining useful information, so as to ensure the information reliability of hydroponic lettuce feature extraction, recognition and classification. The specific process is shown in Figure 4. First, the original color image was enhanced by median filtering and histogram equalization. According to the color features of hydroponic lettuce, the Extra-green algorithm was selected to realize the effective segmentation of the hydroponic lettuce leaves, and its gray image and binary image were obtained.

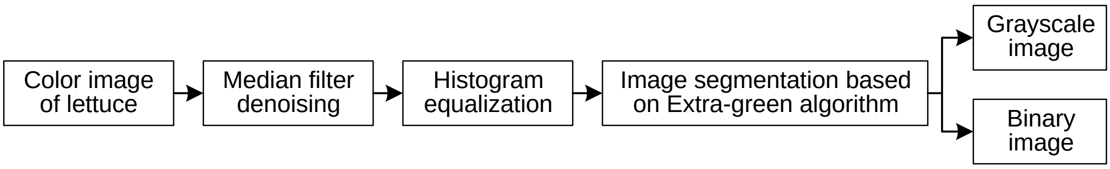


Fig. 4 – Process of image pretreatment

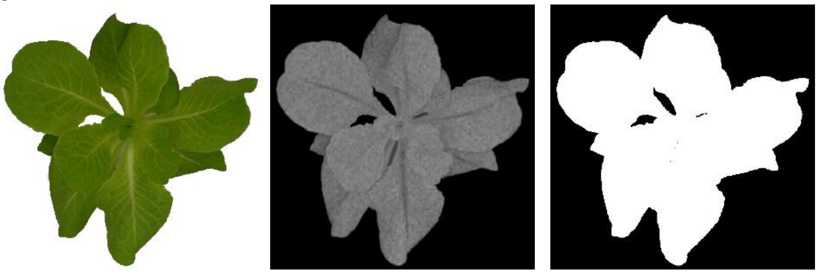
The hydroponic lettuce image segmentation method based on the Extra-green algorithm, which converted the color image in the RGB space into three channels of R, G, and B. When extracting the weight of the G channel, the contrast between the foreground of the green part and the background of the non-green part in the picture was increased. The pixel in the image data was defined as (x, y), and R (x, y), G (x, y), B (x, y) respectively represented the value of the three channels in the RGB color model. Then set the different gray values of the pixels according to equation (1).

$$EXG(x,y)=\begin{cases} & 2G(x,y)-R(x,y)-B(x,y) \\ 255, & 2G(x,y)-R(x,y)-B(x,y)>255 \\ 0, & 2G(x,y)-R(x,y)-B(x,y)<0 \end{cases} \tag{1}$$

For the gray image obtained, the threshold segmentation algorithm was also used to extract the target of hydroponic lettuce leaves from the background. For the image f(x,y) to be segmented, the image segmentation was performed by selecting an appropriate threshold value T and the image g(x,y) was output. The segmentation relationship is shown in equation (2).

$$g(x,y)=\begin{cases} 255, & f(x,y)<T \\ 0, & f(x,y)>T \end{cases} \tag{2}$$

Finally, the image of the hydroponic lettuce leaves can be segmented more accurately, and the result is shown in Figure 5.



(a) Color image *(b) Grayscale image* *(c) Binary image*

Fig. 5 - Image segmentation result based on Extra-green algorithm

<u>Color feature</u>

Color information can be used as an important feature for judging the types of deficiency in hydroponic Lettuce. As a global descriptor to describe the color features of an image, color moments are often used to represent the correlation properties of colors on image. The effective information of an image is generally concentrated in its low-order moment. The Mean reflects the brightness of the image. The Variance describes the distribution range of colors in an image, and the Skewness shows the symmetry of the image color distribution. The above three can effectively express the color distribution of the image. The mathematical definition is expressed in equation (3).

$$\begin{cases} \mu_i = \dfrac{1}{N}\sum_{j=1}^{N} p_{i,j} \\ \sigma_i = \left(\dfrac{1}{N}\sum_{j=1}^{N}\left(p_{i,j}-\mu_i\right)^2\right)^{\frac{1}{2}} \\ S_i = \left(\dfrac{1}{N}\sum_{j=1}^{N}\left(p_{i,j}-\mu_i\right)^3\right)^{\frac{1}{3}} \end{cases} \tag{3}$$

Where:

$P_{i,j}$ indicates the i-th color component of the j-th pixel of the color image, and N is the number of pixels in the image.

In order to avoid the image of hydroponic lettuce being disturbed by light, the HSV color model, which has little influence on the change of light intensity, was selected to collect low-order moment under the channels of H, S and V. At the same time, combined with the actual situation that the color in the image of hydroponic lettuce is not complex, only the Mean and Variance were used as the feature data to describe the color changes of the image. Partial data are shown in Table 2.

Table 2

HSV color feature of some hydroponic lettuce

Sample name	Mean of H channel	Mean of S channel	Mean of V channel	Variance of H channel	Variance of S channel	Variance of V channel
K-deficiency	0.2496	0.5485	0.0878	0.1289	0.0827	0.1427
Ca-deficiency	0.2041	0.5934	0.1110	0.0622	0.0800	0.1542
N-deficiency	0.2626	0.5373	0.1375	0.1218	0.0787	0.1613
Normal	0.2408	0.5699	0.1560	0.1227	0.0942	0.1548

Texture feature

According to the previous experimental observation, the texture depth and roughness of different types of hydroponic lettuce leaves are different. Therefore, the texture feature is an important indicator for the diagnosis of hydroponic lettuce deficiency. Gray level co-occurrence matrix is usually used to describe texture features. It describes the texture by studying the spatial correlation characteristics of gray levels in the image, and it can intuitively reflect the roughness, depth and similarity of the image. Generally, the texture features of images are usually quantified by calculating the Energy, Entropy, Contrast and Correlation of the grayscale co-incidence matrix in different directions.

Energy can indicate the uniformity of gray distribution and the roughness of texture on the image of hydroponic lettuce. Contrast shows the clarity of hydroponic lettuce leaves and the depth of texture grooves, which can more realistically reflect the image texture features. Entropy expresses the size of texture information on the image of hydroponic lettuce. When the gray level of the image of hydroponic lettuce shows greater randomness, the more complex the image is, the greater the entropy is. Correlation mainly reflects the similarity degree of gray level of hydroponic lettuce image in row direction and column direction. It is calculated by the equation (4).

$$\begin{cases} ASM = \sum_i \sum_j p(i,j)^2 \\ CON = \sum_i \sum_j (i-j)^2 p(i,j) \\ ENT = -\sum_i \sum_j p(i,j)\log(i,j) \\ COR = \left[\sum_i \sum_j (ij)p(i,j) - \mu_i\mu_j\right] / \sigma_i\sigma_j \end{cases} \tag{4}$$

Where: $\mu_i = \sum_i \sum_j i \cdot p(i,j)$ $\mu_j = \sum_i \sum_j j \cdot p(i,j)$ $\sigma_i^2 = \sum_i \sum_j p(i,j)(i-\mu_i)^2$ $\sigma_j^2 = \sum_i \sum_j p(i,j)(j-\mu_j)^2$

(i, j) is the gray value of a pixel pair in the image, and $p(i, j)$ is the probability of the gray value appearing in the whole image.

In this study, the values in the four directions of 0°, 45°, 90° and 135° were calculated to represent the texture features of the hydroponic lettuce image, meanwhile, the mean and variance were calculated as the feature information for subsequent research. Partial data are shown in Table 3.

Table 3

Texture feature of some hydroponic lettuce

Sample name	Energy		Entropy		Contrast		Correlation	
	Mean	Variance	Mean	Variance	Mean	Variance	Mean	Variance
K-deficiency	0.6484	0.0112	0.6747	0.0351	0.0469	0.0130	5.3476	0.2418
Ca-deficiency	0.5732	0.0026	0.8394	0.0226	0.0289	0.0080	2.8147	0.0321
N-deficiency	0.5103	0.0182	0.9583	0.0560	0.0812	0.0224	3.6239	0.2176
Normal	0.7892	0.0046	0.4361	0.0175	0.0233	0.0059	9.2589	0.2836

Shape feature

The combination of shape and geometry features can be used as an important basis for distinguishing different types of lettuce deficiency. Based on the boundary feature, the contours of the continuous area of hydroponic lettuce leaves were extracted from the binary image, and then five types of information such as Extent, Eccentricity, Solidity, Thinness ratio and Aspect ratio were extracted to describe the shape features of hydroponic lettuce leaves. Table 4 shows the partial data.

Table 4

Shape feature of some hydroponic lettuce

Sample name	Extent	Eccentricity	Solidity	Thinness ratio	Aspect ratio
K-deficiency	0.4833	0.4252	0.6869	0.0381	1.0441
Ca-deficiency	0.4863	0.7838	0.7182	0.0575	1.3380
N-deficiency	0.3146	0.6336	0.4727	0.0132	1.2285
Normal	0.5840	0.5963	0.7759	0.0796	0.9158

Feature selection based on PCA

PCA is a data processing method that recombines the original variables into a new set of unrelated integrated variables. It uses a few new integrated variables to contain as much of the original data as possible, which reflects the relationship between the original variables and is widely used in data dimensionality reduction processing.

In this study, the 19 original features were extracted from the image of hydroponic lettuce. To avoid a large number of input variables in the establishment of a nutrient deficiency diagnosis model, the principal component analysis was performed to calculate the characteristics of each principal component. The value and contribution rate are shown in Table 5. It can be seen from the results that the eigenvalues of the first six principal components are all greater than 1, and their cumulative contribution rate reaches at 79.376%. Therefore, comprehensively considering the principle of eigenvalue greater than 1 and cumulative contribution rate not less than 80% in the selection of principal components, the first 6 principal components were selected to comprehensively express 19 features.

Table 5

Eigenvalues and contribution rates of principal components

Component	Eigen value			Extraction sums of squared loadings		
	Total	Variance contribution/%	Cumulative variance contribution/%	Total	Variance contribution/%	Cumulative variance contribution/%
1	6.873	36.172	36.172	6.873	36.172	36.172
2	2.610	13.738	49.910	2.610	13.738	49.910
3	1.600	8.421	58.332	1.600	8.421	58.332
4	1.558	8.198	66.529	1.558	8.198	66.529
5	1.440	7.582	74.111	1.440	7.582	74.111
6	1.000	5.265	79.376	1.000	5.265	79.376
7	0.927	4.881	84.257			
8	0.840	4.419	88.675			
9	0.702	3.694	92.369			

Table 5
(continuation)

10	0.495	2.608	94.977			
11	0.287	1.511	96.488			
12	0.248	1.306	97.794			
13	0.145	.765	98.560			
14	0.138	.729	99.288			
15	0.048	.251	99.539			
16	0.044	.233	99.773			
17	0.037	.195	99.968			
18	0.004	.022	99.991			
19	0.002	.009	100.000			

By calculating the load of the 19 features and the eigenvalue of each principal component, the function expressions Y1 to Y6 of the six principal components can be obtained, as shown in equation (5). Z1 ～Z19 respectively represent the mean of H channel, mean of S channel, mean of V channel, variance of H channel, variance of S channel, variance of V channel, mean of energy, variance of energy, mean of entropy, variance of entropy, mean of contrast, variance of contrast, mean of correlation, variance of correlation, extent, eccentricity, solidity, thinness ratio, aspect ratio.

$$\begin{aligned}
Y_1 &= 0.053Z_1 + 0.0224Z_2 + 0.1454Z_3 + 0.1471Z_4 + 0.1423Z_5 + 0.1024Z_6 + 0.3365Z_7 - 0.3240Z_8 - \\
&\quad 0.3443Z_9 - 0.3577Z_{10} - 0.3454Z_{11} - 0.3528Z_{12} + 0.1811Z_{13} + 0.132Z_{14} + 0.2309\,Z_{15} - 0.0824Z_{16} \\
&\quad + 0.2383Z_{17} + 0.2173Z_{18} + 0.0019Z_{19} \\
Y_2 &= -0.0723Z_1 - 0.1068Z_2 - 0.0558Z_3 + 0.1485Z_4 + 0.3825Z_5 + 0.3045Z_6 + 0.1529Z_7 + 0.1755Z_8 + \\
&\quad 0.1392Z_9 + 0.1531Z_{10} + 0.1607Z_{11} + 0.1508Z_{12} - 0.1999Z_{13} - 0.1764Z_{14} + 0.4059Z_{15} - 0.0931Z_{16} \\
&\quad + 0.4151Z_{17} + 0.3948Z_{18} - 0.0219Z_{19} \\
Y_3 &= 0.3955Z_1 + 0.3514Z_2 + 0.4879Z_3 - 0.0094Z_4 - 0.2086Z_5 - 0.0056Z_6 - 0.0497Z_7 + 0.1674Z_8 + \\
&\quad 0.0605Z_9 + 0.1293Z_{10} + 0.1749Z_{11} + 0.1635Z_{12} + 0.2576Z_{13} + 0.2892Z_{14} + 0.1289Z_{15} - 0.3696Z_{16} \\
&\quad + 0.0897Z_{17} + 0.0985Z_{18} + 0.0943Z_{19} \\
Y_4 &= -0.3423Z_1 - 0.2982Z_2 - 0.0786Z_3 - 0.424Z_4 - 0.072Z_5 - 0.2841Z_6 - 0.1022Z_7 + 0.0066Z_8 + \\
&\quad 0.0869Z_9 + 0.05614Z_{10} + 0.0209Z_{11} + 0.0433Z_{12} + 0.382Z_{13} + 0.4327Z_{14} + 0.1675Z_{15} + 0.182Z_{16} \\
&\quad + 0.1916Z_{17} + 0.2359Z_{18} + 0.1072Z_{19} \\
Y_5 &= 0.0506Z_1 - 0.0425Z_2 - 0.2737Z_3 + 0.4194Z_4 + 0.3061Z_5 + 0.3756Z_6 + 0.0254Z_7 + 0.1072Z_8 + \\
&\quad 0.0063Z_9 + 0.0592Z_{10} + 0.0872Z_{11} + 0.0766Z_{12} + 0.4228Z_{13} + 0.4567Z_{14} - 0.1501Z_{15} + 0.1132Z_{16} \\
&\quad - 0.1472Z_{17} - 0.1816Z_{18} - 0.0568Z_{19} \\
Y_6 &= -0.2418Z_1 + 0.2325Z_2 + 0.0298Z_3 - 0.0601Z_4 + 0.1428Z_5 + 0.1987Z_6 - 0.0347Z_7 - 0.0594Z_8 + \\
&\quad 0.0403Z_9 - 0.0132Z_{10} - 0.0416Z_{11} - 0.0319Z_{12} - 0.0388Z_{13} - 0.0299Z_{14} - 0.0453Z_{15} - 0.0106Z_{16} \\
&\quad - 0.0715Z_{17} - 0.0772Z_{18} + 0.8933Z_{19}
\end{aligned} \tag{5}$$

Random forest

Random forest algorithm (RF) is widely used in data classification because of its advantages of high classification accuracy, fast calculation speed, strong anti-noise ability, and not easy to fall into over-fitting. Its core idea is to integrate multiple unrelated decision trees into a forest. Each decision tree randomly has a replacement sample from the original data set, completes independent learning and calculation, and finally determines the classification result of the random forest based on voting. The random selection of samples and features is the most important feature. The number of decision trees and the number of input feature variables when the nodes of the generated decision tree are split are important parameters that affect the effect of classification. Combining the extracted image feature information of hydroponic lettuce, this paper focused on constructing a random forest algorithm classification model to realize the nutrient deficiency diagnosis in hydroponic lettuce.

RESULTS AND DISCUSSION

The operating environment was Windows 10 system equipped with Intel Core i5-7200U processor, and the recognition algorithm was realized by Matlab and C mixed programming. The classification objects of the model were four types of hydroponic lettuce images including K-deficiency, Ca-deficiency, N-deficiency and Normal. After the features extracted on 1440 original images, rejected some null values, and finally 1424 sets of effective feature data were obtained. Each type of hydroponic lettuce contained 356 sets of feature information, and 200 sets of each type were randomly selected as training samples to form a training set containing 1000 sets of feature information, and 424 sets of feature information were left as a test set.

Results of Random Forest

The important parameters of the model were optimized after multiple experiments by the training set, and finally it was determined that there were 10 decision trees in the RF model, and the number of feature variables K was 6. After training, the model's classification accuracy of the training set reached 94.06%, indicating that the model fully fitted the training data. The 424 sets of test data were used to evaluate the accuracy of the classification results of the model. The classification accuracy of different types of lettuce is shown in Table 6, the confusion matrix of the classification results is shown in Table 7, and the overall classification accuracy of the model was 86.32%, Kappa coefficient was 0.82.

Table 6

Accuracy of diagnostic result on test set

Class	Accuracy /%
K-deficiency	88.68
Ca-deficiency	81.13
N-deficiency	90.57
Normal	84.91

The model had the highest classification accuracy of N-deficient lettuce, which was 90.57%. The classification of Ca-deficient and Normal lettuce was not accurate enough, and the classification accuracy were 81.13% and 84.91%, respectively. Through the confusion matrix, it could be concluded that the misclassified samples of Ca-deficient lettuce were mainly concentrated in K-deficiency and Normal lettuce. The reason may be that in the mid-growth, the leaf curling features of some Calcium-deficient lettuce were not obvious. The dark brown edge of the leaf was long and thin, and might be blurred during leaf segmentation. In the same way, some Ca-deficient images were highly similar to Normal images, which resulted in a large proportion of Normal lettuce that was incorrectly classified as Ca-deficient lettuce.

Table 7

Confusion matrix of diagnostic result on test set

	Predicted Class							
Actual Class	**K-deficiency**		**Ca-deficiency**		**N-deficiency**		**Normal**	
	Number	**Percentage /%**	**Number**	**Percentage /%**	**Number**	**Percentage /%**	**Number**	**Percentage /%**
K-deficiency	94	22.17	6	1.42	3	0.71	3	0.71
Ca-deficiency	7	1.65	86	20.28	4	0.94	9	2.12
N-deficiency	3	0.71	5	1.18	96	22.64	2	0.47
Normal	5	1.18	11	2.59	0	0	90	21.23

Comparison of different methods

In order to further evaluate the classification effect and performance of RF model, it was compared with the SVM and BP models, and the same data were used to judge different types of hydroponic lettuce. The radial basis function was selected as the kernel function to establish the SVM model, and its important parameters such as gamma function g and penalty variable c were optimized through cross-validation. The BP model was constructed with 6 neurons in input layer, 10 neurons in hidden layer and 4 neurons in output layer, the learning rate was 0.02, and the number of iterations was 5000. The functions of hidden layer and output layer were softmax and tradingdx respectively, and the comparison results of the three methods are shown in Figure 6.

The RF model had the best classification effect among the three classification methods. The overall classification accuracy was 86.32%, which was 2.59 and 7.07 percentage points higher than that of SVM and BP, respectively. Kappa coefficient was 0.82, which was 0.04 and 0.10 higher than that of SVM and BP, respectively (Shown in Figure 6). Comprehensive comparison showed that the RF model proposed in this study could diagnose the nutrient deficiency types of hydroponic lettuce more accurately, and had good applicability.

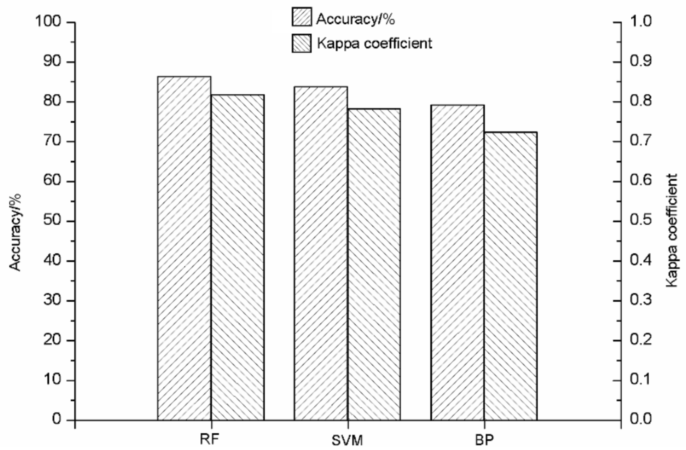


Fig. 6 - Comparison of classification results of different methods

CONCLUSIONS

A method for diagnosing nutrient deficiency in whole hydroponic lettuce based on RF was proposed. Images of the whole hydroponic lettuce in the mid-growth stage were used in the study, the 19 features of three types of leaf color, texture, and shape were extracted, and were reduced to 6 principal component variables by PCA. Then, the diagnosis models were established with RF, SVM and BP. The results showed that RF had the highest classification accuracy, 86.32%, which was 2.59 and 7.07 percentage points higher than that of SVM and BP, respectively. Kappa coefficient was 0.04 and 0.10 higher than that of SVM and BP, respectively.

So, the RF model constructed in this study can diagnose different deficiency types of hydroponic lettuce in the mid-growth stage effectively, and can provide a basis for the prevention and control of hydroponic lettuce deficiency and the efficient management of nutrient solutions in plant factories.

ACKNOWLEDGEMENT

This work was supported by research grants from Shaanxi Key Research and Development Program of China, grant no. 2019ZDLNY02-04.

REFERENCES

[1] Fang, W. (2016). Plant factory industry development and business model in Taiwan (台湾植物工厂产业发展与商业模式). *Agricultural Engineering Technology*, 36(13):17-20. https://doi.org/10.16815/j.cnki.11-5436/s.2016.13.002

[2] Germer, J., Sauerborn, J., Asch, F., de Boer, J., Schreiber, J., Weber, G., &Muller, J. (2011). Skyfarming an ecological innovation to enhance global food security. *Journal of Consumer Protection and Food Safety*, 6(2):237-251. https://doi.org/10.1007/s00003-011-0691-6

[3] Gillespie, D. P., Kubota, C., &Miller, S. A. (2020). Effects of Low pH of Hydroponic Nutrient Solution on Plant Growth, Nutrient Uptake, and Root Rot Disease Incidence of Basil. *Hortscience*, 55(8):1251-1258. https://doi.org/10.21273/HORTSCI14986-20

[4] Gu, Z. F. S., Li, Y. N., Ji, F., &He, D. X. (2019). The sustainable development prospects of artificial light type plant factories (人工光型植物工厂的可持续发展前景). *Agricultural Engineering Technology*, 39(34):22-34. https://doi.org/10.16815/j.cnki.11-5436/s.2019.34.003

[5] Lu, B., Sun, J., Mao, H. P., Yang, N., & Wu, X. H. (2018). Disease recognition of lettuce with feature fusion based on hyperspectrum and image (高光谱和图像特征融合的生菜病害识别). *Jiangsu Journal of Agricultural Sciences*, 34(6):1254-1259. https://doi.org/10.3969/j.issn.1000-4440.2018.06.008

[6] Mao, H. P., Xu, G. L., &Li, P. P. (2003). Extracting and selecting features of leaf images for diagnosing nutrient deficiency diseases in tomatoes (番茄缺素叶片的图像特征提取和优化选择研究). *Transactions of the Chinese Society Agricultural Engineering*, 19(2):133-136. http://www.tcsae.org/nygcxb/ch/reader/view_abstract.aspx?file_no=20030230&flag=1

[7] Mao, H. P., Gao, H. Y., Zhang, X. D., &Kumi, F. (2015). Nondestructive measurement of total nitrogen in lettuce by integrating spectroscopy and computer vision. *Scientia Horticulture*, 184:1-7. https://doi.org/10.1016/j.scienta.2014.12.027

[8] Paulo, F., Zhang, Z., Jithin, M., Nusrat, J., &John, S. (2020). Distinguishing Volunteer Corn from Soybean at Seedling Stage Using Images and Machine Learning. *Smart Agriculture*, 2(03): 61-74. https://doi.org/10.12133/j.smartag.2020.2.3.202007-SA002

[9] Wei, W. S., Xing, Y. Y., Li, Y. Y., Peng, Y. K., &Zhang, W. P. (2018).Online detection and classification system of external quality of leaf for dining hall and family (适于餐厅与家庭的叶菜外部品质在线检测与分级系统). *Transactions of the Chinese Society Agricultural Engineering*, 34(05):264-273. https://doi.org/10.11975/j.issn.1002-6819.2018.05.035

[10] Xu, H. X. (2013). *Study on Nondestructive Detection of Freshness of Post-harvest Spinach Based on Machine Vision and Electronic Nose (基于机器视觉和电子鼻技术的菠菜新鲜度无损检测研究)* [Master dissertation, Jiangsu University]. CNKI. https://kns.cnki.net/KCMS/detail/detail.aspx?dbcode=CMFD&dbname=CMFD201602&filename=1016727760.nh&v=MDExNjBlWDFMdXhZUzdEaDFUM3FUcldNMUZyQ1VSN3VmWU9SdUZ5bm1WTDNQVkYyNkdMUzZHZGJLcjVFYlBJUjg=

[11] Yang, L. Q., Qiu, S. F., Tang, F., Huang, D. F., &Tang, D. Q. (2015). Comprehensive Evaluation for Ornamental Lettuce Based on Principal Component Analysis (基于主成分分析的观赏生菜品质综合评价). *Journal of Shanghai Jiaotong University (Agricultural Science)*, 33(3):53-60. https://doi.org/10.3969/j.issn.1671-9964.2015.03.008

[12] Yu, X. J., Lu, H. D., &Liu, Q. Y. (2018). Deep-learning-based regression model and hyperspectral imaging for rapid detection of nitrogen concentration in oilseed rape (Brassica napus L.) leaf. *Chemometrics and Intelligent Laboratory Systems*, 172(10): 48-55. https://doi.org/10.1016/j.chemolab.2017.12.010

[13] Zhao, D. M. (2014). Common symptoms of vegetable deficiency and prevention methods (蔬菜常见缺素症状及防治方法). *Vegetables*, (11):67-68. https://kns.cnki.net/KCMS/detail/detail.aspx?dbcode=CJFD&dbname=CJFDLAST2015&filename=SCZZ201411032&v=MTY5MzNvOUdab1I4ZVgxTHV4WVM3RGgxVDNxVHJXTTFGckNVUjd1ZllPUnVGeS9uVnIzSk5pN1JkTEc0SDlYTnI=

[14] Zhang, B., Sun, J., Jin, X. M., Li, W. H., Gao, H. Y., &Xiao, L. (2013). Prediction Model of Lettuce Nitrogen Content Based on Color Images. *Information Technology Journal*, 12 (23):7833-7838. https://doi.org/10.3923/itj.2013.7833.7838

[15] Zhang, S. W., Wu, X. W., You, Z. H., &Zhang, L. Q. (2017). Leaf image based cucumber disease recognition using sparse representation classification. *Computers and Electronics in Agriculture*, 134:135-141. https://doi.org/10.1016/j.compag.2017.01.014