DETECTION OF BLACK HEART DISEASE IN SEED POTATO BASED ON TRANSMISSION SPECTROSCOPY TECHNIQUE
/
基于透射光谱技术的马铃薯种薯黑心病检测研究

Xianhe WANG[1], Min HAO[1,*], Xingtai CAO[1], Yutao ZHANG[1]
[1]College of Electromechanical Engineering, Inner Mongolia Agricultural University,
Hohhot, Inner Mongolia Autonomous Region 010018, China
*Corresponding authors, Email: hhhthaomin@163.com
DOI: https://doi.org/10.35633/inmateh-73-42

Keywords: *Neural networks, Transmission spectroscopy, Black heart, Seed Potatoes*

ABSTRACT
Black heart disease is one of the screening indicators of seed potatoes, which has a serious impact on the quality and yield of potato, and at present there are fewer non-destructive testing methods for internal defects of seed potatoes. This paper aims to utilize non-destructive testing technology to quickly identify qualified and black hearted seed potatoes, and then to protect yield and food security. In this paper, transmission spectroscopy system was utilized to collect the spectral data of 104 qualified seed potatoes and 104 black hearted seed potatoes in 450~940 nm band. Subsequently, four algorithms, namely Savitzky-Golay (SG), Standard Normal Variate (SNV), Multiplicative Scatter Correction (MSC) and First-order Derivative (FD), were utilized to pre-process the seed potatoes spectral data to improve the data noise ratio. Feature wavelength extraction was made using Competitive Adaptive Reweighted Sampling (CARS) and Successive Projections Algorithm (SPA) to enhance the sample data characteristics and improve the model interpretability. The construction of classification models for qualified and black hearted seed potatoes relied on two deep learning techniques, Convolutional Neural Networks (CNN) and Recurrent Neural Network (RNN), which were trained and tested for the feature bands respectively. The experimental results showed that SG-CARS-CNN was the optimal combination of classification algorithms, and the classification accuracies of both the training set and the test set reached 100%, which improved the accuracy of the test set by 3.85% compared with that of the traditional machine learning algorithms, and provided an accurate method for the rapid screening of qualified seed potatoes.

摘要
黑心病是马铃薯种薯筛选指标之一，对马铃薯的品质和产量有严重的影响，而目前针对马铃薯种薯内部缺陷无损检测方法较少，本文旨在利用无损的检测技术快速识别合格与黑心马铃薯种薯，进而保障马铃薯产量和粮食安全。本文利用透射光谱系统采集 104 个合格种薯和 104 个黑心种薯 450~940nm 波段光谱数据，随后利用 SG 卷积平滑（SG）、标准正态变换法（SNV）、多元散射校正法（MSC）和一阶导数（FD）4 种算法对马铃薯种薯光谱数据预处理，以提高数据信噪比；利用竞争性自适应重加权采样法(CARS)和连续投影算法(SPA)进行特征波长提取，以强化样本数据特征和提高模型可解释性；合格与黑心马铃薯种薯分类模型的构建依赖于卷积神经网络(CNN)和循环神经网络(RNN) 2 种深度学习技术，分别对特征波段进行训练与测试。试验结果表明，SG-CARS-CNN 为最优分类算法组合，训练集和测试集分类准确度均达到 100%，相比于传统机器学习算法测试集准确率提高了 3.85%，为快速筛选合格马铃薯种薯提供准确方法。

INTRODUCTION
Potato is the fourth largest food crop in China, with high nutritional and economic value, strong adaptability, large yield, wide range of uses (*Wang et al., 2018*), and the quality of seed potatoes directly affects the level of potato production in China. No. 1 of the 2024 central document states that the production of food and important agricultural products should be grasped to ensure national food security and food security of the country with the experience of the "Ten Million Project" (*Jiang et al., 2024*). Therefore, in order to accelerate the modernization of agriculture and fight a good battle for the seed industry (*Ning et al., 2021*), efficient screening of excellent seed potatoes is one of the important guarantees for food security in China.

Xianhe Wang, Student; Min Hao, Associate professor; Xingtai Cao, Student; Yutao Zhang, Student.

Black heart disease is one of the important screening indicators of seed potatoes, and its use in seeding can lead to the spread of the disease, reduce yield and affect food safety and other hazards. Seed potatoes are usually screened by machine and by hand. According to China's seed potatoes quality standard (GB/T 29377-2012), the permissible rate of black heart disease is 0.1% (*GB/T 29377-2012, 2012*). Manual screening is overly dependent on manual judgment, low detection and sorting precision (*Tiwari et al., 2020)*, and unable to screen out black heart disease seed potatoes, while machine screening is mainly for potatoes weight and shape screening, and there is a certain limitation on the screening of black heart disease seed potatoes. In the field of agriculture, spectral detection technology is usually used in plant disease detection and internal quality detection of agricultural products, with the advantages of being non-destructive and accurate. Since black heart disease of seed potatoes belongs to the internal defects of agricultural products, it is feasible to use spectral detection technology to screen out seed potatoes containing black heart disease.

At present, most of the potato quality detection uses traditional machine learning algorithms to extract features and construct quality detection models. *Zhou et al. (2012)* used diffuse reflectance spectra by the second-order derivative and normalized combination of the model built on the test set of samples identified the correct rate of 92.31%. *Gao et al. (2013)* used partial least squares discriminant model (PLS-DA) to establish a discriminant model for potato black heart disease, and verified the feasibility of the method. *Han et al. (2021)* used principal component analysis combined with partial least squares discriminant model (PCA-PLS-DA) to establish a discriminant model of potato black heart disease, and its accuracy was 96.73%. Foreign studies on potato black heart disease are fewer and still remain in the detection of potato nutrients. *Abukmeil et al. (2022)* estimated the nutrients in potato plants by detecting the reflectance of potato leaves, thus replacing the destructive method of chemical experiments.

However, traditional machine learning algorithms have low detection accuracy and weak generalization ability, resulting in decreased performance in practical production. In recent years, deep learning algorithms have been developed, which are fast, accurate and efficient, and can effectively extract features to improve production efficiency and automation. *Khorramifar et al. (2022)* predicted potato soluble solids pH using artificial neural network (ANN) with an accuracy of 92%. *Gupta et al. (2023)* used convolutional neural network (CNN) for potato leaf disease classification with 90% accuracy.

In summary, domestic and foreign research on potato defect detection mainly stays in the research of potato skin, dry matter, nutrient elements, ring rot and so on. Our country has a very low permission rate for black heart disease seed potatoes, and there is less research on the detection of black heart disease defects in potato, and the detection accuracy is lower. Therefore, for the current problem of low screening accuracy of seed potatoes containing black heart disease, this paper proposes a study on black heart disease detection in seed potatoes based on transmission spectroscopy technology, using visible-near infrared spectrometer to collect transmission spectroscopy data of seed potatoes, and establishing a dichotomous classification model of seed potatoes with black heart disease by comparing and analyzing multiple algorithmic combination models to meet the requirements of screening black heart disease in seed potatoes in China.

MATERIALS AND METHODS

Sample collection and preparation

Ulanchabu is the potato capital, located in the central part of the Inner Mongolia Autonomous Region. The study varieties used were field-harvested Ulanchabu purple-flowered white seed potatoes, which are characterized by thin skin, large tubers, good shape, and few pests and diseases. Since seed potatoes containing black heart disease could not be directly observed, this paper used the field sample harvesting method and laboratory sample preparation method for sample collection, with 104 qualified seed potatoes and 104 black hearted seed potatoes, totaling 208. The cross section of qualified and black heart potato samples are shown in Fig. 1a, 1b and 1c. Black heart potatoes are completely different from certified potatoes. Blackheart potatoes are rotten, soft and smelly on the inside.

Potato black heart disease is caused by high temperature and hypoxic conditions. The laboratory sample preparation method is: field-harvested seed potatoes were vacuum-sealed and transferred to a thermostat and stored at 40° C for 27 h. During this period, the temperature was adjusted to 3° C at 6-h intervals to minimize the respiration rate and stored for 1 h. The temperature was adjusted to 3° C at 6-h intervals to minimize the respiration rate.

a. Qualified potato *b - Slight black heart potato* *c - Preparation of black heart potato in laboratory*

Fig. 1 - Cross section of qualified and black heart potato samples

Transmission Spectroscopy Data Acquisition System

Transmission spectroscopy data acquisition system mainly includes: DELL laptop, Ocean View spectroscopy acquisition software, Ocean Optics USB4000 micro-spectrometer, optical fiber, collimator mirror, twelve 60 W halogen lamps. Ocean Optics USB4000 spectrometer has spectral band range of 450 nm ~ 940 nm and a total of 3151 bands; halogen lamp is added in the bulb halogen gas at high temperatures to achieve luminescence, with high luminous efficacy, long life, it will not affect the quality of the samples, the transmittance system of twelve 60 W halogen lamps is placed in the experimental samples of the upper-left 45° and the upper-right 45°. The structure of transmission spectrum data acquisition system is shown in Fig. 2, the light passes through the sample, and is transmitted to the spectrometer through the collimating mirror and optical fiber, and then the optical signal is converted into an electrical signal.

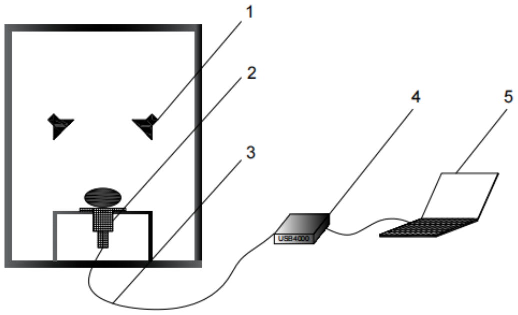


Fig. 2 - Structure of transmission spectrum data acquisition system
1 - Halogen lamp; 2 – Collimator; 3 - Fiber optic; 4 - USB4000 Spectrometer; 5 - Laptop

In order to ensure the stability of the test and its accuracy, the spectrometer needs to be switched on and warmed up for 30 min before the test, and after the spectrometer and the light source are stabilized, the reference correction is carried out, respectively, by using the blackboard and the collection of standard cellophane (yellow) to cover the collimating mirror, to obtain the dark-field light intensity and the reference light intensity, with the aim of eliminating the systematic error and maintaining the consistency of the measurement results.

$$T = \frac{T_s - T_d}{T_r - T_d} \tag{1}$$

In the formula, T is the transmissivity, T_s is the original light intensity, T_r is the reference light intensity, T_d is the dark field light intensity.

Data preprocessing

During the experimental process, the transmission spectroscopy data acquisition system will generate electronic noise, ambient temperature fluctuations and other interferences, and the differences in the size and shape of the experimental samples will affect the consistency of the spectral data. Therefore, in order to reduce the influence of environmental noise, improve the signal-to-noise ratio of spectral data, and optimize the model performance (*Li et al., 2024; Liu et al., 2024)*, it is necessary to preprocess the transmission spectral data of the black heart seed potatoes before building the model. In this paper, the preprocessing method uses the most commonly used SG convolutional smoothing, standard normal transform method, multiple scattering correction (MSC) and first-order derivatives (FD) four algorithms, and the optimal preprocessing algorithms are derived from the comparative analysis of traditional machine learning models using support vector machine (SVM).

Characteristic Wavelength Extraction
Competitive Adaptive Reweighted Sampling (CARS)

The Competitive Adaptive Reweighted Sampling (CARS) method aims to select the most competitive band combinations (*Zhao et al., 2023)*. Monte Carlo sampling is used to select some samples for modeling, and the remaining samples are used as a prediction set to build a partial least squares (PLS) model to obtain the regression coefficients for the first wavelength:

$$|K_i|(i=1,2,\dots,p) \tag{2}$$

The combinations of wavelengths with higher weights are filtered as a new subset using the exponential decay function:

$$r_i = \mathrm{a}e^{-\mathrm{b}j}(j=1,2,\dots,N) \tag{3}$$

In the formula, j denotes the j^{th} Monte Carlo sampling; N denotes the total number of Monte Carlo sampling; a and b are constants.

After several calculations, the evaluation of weights is used for variable selection:

$$\omega_i = |K_i| / \sum_{i=1}^{p} |K_i| (i=1,2,\dots,p) \tag{4}$$

Finally, the subset with the smallest root mean square error of validation (RMSECV) is chosen as the characteristic wavelength of black heart disease in seed potatoes by selecting cross validation.

Successive Projection Algorithm (SPA)

Successive projection algorithm (SPA) is a forward iterative method of feature variable screening, by projecting one wavelength onto another wavelength to compare the size of the projection vector, the wavelength with larger projection is used as the feature wavelength; with each iteration, SPA selects the wavelength with the least covariance with the screened feature wavelength as the new feature wavelength, which minimizes the covariance feature; the cycle repeats itself and the wavelengths obtained subset contains most of the data variability and is used as the final feature wavelength. Therefore, SPA is less sensitive to noise and outliers, and feature wavelength extraction can effectively reflect the essential characteristics of the data (*Arshaghi et al., 2021)*. Transmission spectroscopy data acquisition system is susceptible to noise interference, so it is chosen to utilize SPA for feature wavelength extraction of the data pre-processed by SG.

Neural network model
Convolutional neural network (CNN)

Convolutional neural network (CNN) is one of the typical deep learning algorithms, the core of which is to use convolutional and pooling layers to extract data features and classify them through fully connected layer (*Al-Adhaileh et al., 2023; Huang et al., 2022)*. The model has two convolutional layers with 16 and 32 convolutional kernels, respectively. Since the ReLU function has an input value equal to itself in the positive interval and an output of 0 in the negative interval, it makes the network nonlinear; since the ReLU function always has a gradient of 1 on the positive semiaxis, which does not decrease with the number of layers, it is a better solution compared to other activation functions such as Sigmoid or Tanh activation functions that can better mitigate the gradient vanishing problem. Therefore, utilizing the ReLU function as the activation function of the model is more helpful for model training.

$$ReLu = \max(0, x) \tag{5}$$

In the formula, x denotes the feature input in the network.

The two pooling layers reduce the data dimensionality and prevent overfitting, with kernel sizes set to 1 and 2, step sizes of 2 for both. After combining with the ReLU activation function, the nonlinear expression ability of the network can be increased to improve the model generalization ability. The fully connected layer unfolds the feature wavelengths of different samples into one-dimensional vectors for integration, and uses the Softmax function to map the feature representations learned from the convolutional and pooling layers to the samples (*Rogers et al., 2023)*, and ultimately realizes the role of a classifier that transforms from the feature wavelengths to the qualified and the black heart seed potatoes. The structure simplification diagram of CNN classification model is shown in Fig.3.

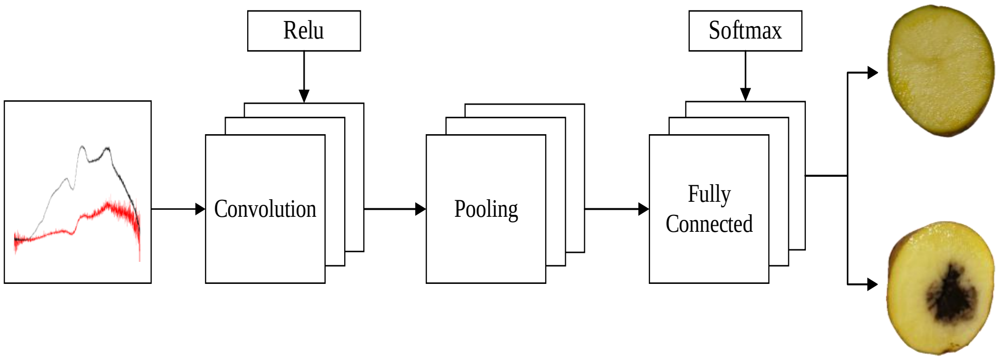


Fig. 3 - Structure simplification diagram of CNN classification model

Recurrent neural network (RNN)

Recurrent neural network (RNN) consists of input layer, hidden layer and output layer, with weight sharing property and memory function. The core of RNN is the hidden layer, which optimizes the model and improves the classification ability through the way of neuron cyclic connection. The weight sharing property is manifested in the fact that the recurrent network uses the same parameters to process the data at each moment, which makes the model reduce the parameters during learning, and reduces the risk of overfitting. The recurrent network has a certain memory ability by processing sequential data, which is mainly manifested in the fact that the network takes into account the influence of inputs from other moments every time it processes data or outputs, and transmits and saves the key information through the memory units in the network, which gives the network a "memory" function. The network is endowed with the function of "memory". The weight sharing and memory function can effectively reduce the complexity of the model and give the dependence of each neuron in the network, which can more powerfully explain the relationship between the feature wavelength and the classification results in the classification of qualified and black heart disease seed potatoes. The structure of the RNN classification model is shown in Fig. 4.

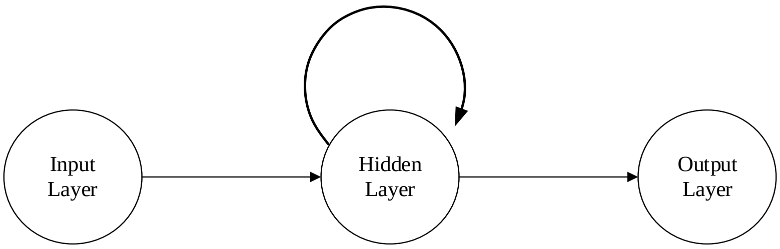


Fig. 4 - Structure simplification diagram of RNN classification model

RESULTS AND ANALYSIS

Data preprocessing

In order to avoid oversaturation, the integration time of the spectrometer was set to 100 ms, and the acquisition band was 450 ~ 940 nm, 104 transmission spectra of qualified seed potatoes and 104 transmission spectra of black-centered seed potatoes were collected, totaling 208 samples of data. The average curve of wavelength-transmission of qualified and black-cored potatoes are shown in Fig. 5.

Due to the cell death of the pith tissue in the center of black heart seed potatoes tubers, the central part of the tubers became black and partly cracked and hollow, which was manifested as the loss of water and hardening, and as shown in Fig. 1b and Fig. 1c, the transmittance of black-centered potatoes was obviously lower than that of normal potatoes, and the two presented the maximum difference at 717 nm, and at 652 nm, and the maximum difference at 652 nm, and at 652 nm. The maximum difference between the two was at 717 nm, and the absorption peaks were obvious near 652 nm, 717 nm and 814 nm.

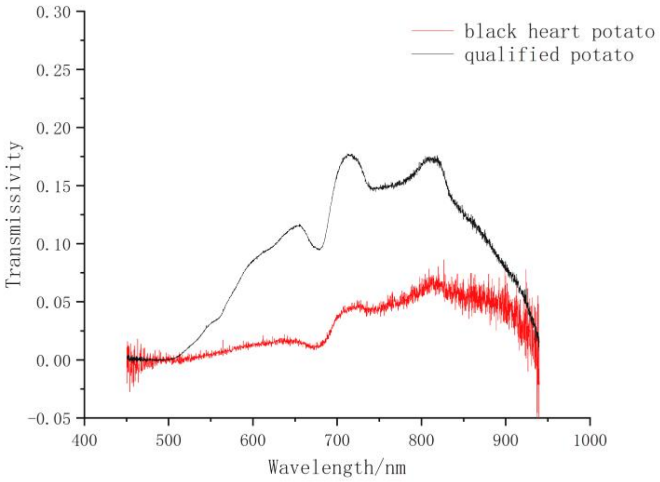


Fig. 5 - Wavelength-transmissivity curve of qualified potato and black heart potato

Table 1 shows the support vector machine discriminant model (SVM) using different preprocessing methods.

Table 1

Classification models of different preprocessing methods based on SVM

Algorithmic	Training set		Test set	
	Number of Errors	Correctness / %	Number of Errors	Correctness / %
Raw	10	93.59	4	92.31
SG	4	97.44	2	96.15
SNV	9	94.23	3	94.23
MSC	6	96.15	4	92.31
FD	11	92.95	5	90.38

The classification models of different preprocessing methods based on SVM are shown in Table 1 ,by comparing and analyzing the number of misclassification and correct rate of different preprocessing methods, the number of misclassification in the correction set of first-order derivative (FD) method is 11, with a correct rate of 92.95%, and the number of misclassification in the prediction set is 5, with a correct rate of 90.38%; the number of misclassification in the correction set of Savitzky-Golay is 4, with a correct rate of 97.44%, and the number of misclassification in the prediction set is 2, with a correct rate of 96.15%; among the four preprocessing methods first order derivative (FD) has the lowest classification accuracy and S-G convolutional smoothing has the highest classification accuracy.

In summary, the Savitzky-Golay will be used to preprocess the transmission spectral data of seed potatoes. However, this preprocessing method does not meet the black heart disease allowable rate of seed potatoes quality standard (GB18133-2000) formulated in China, and the spectral band dimension is large, which is weak for model interpretation, so the spectral data preprocessed by Savitzky-Golay will be downgraded to extract the most relevant features.

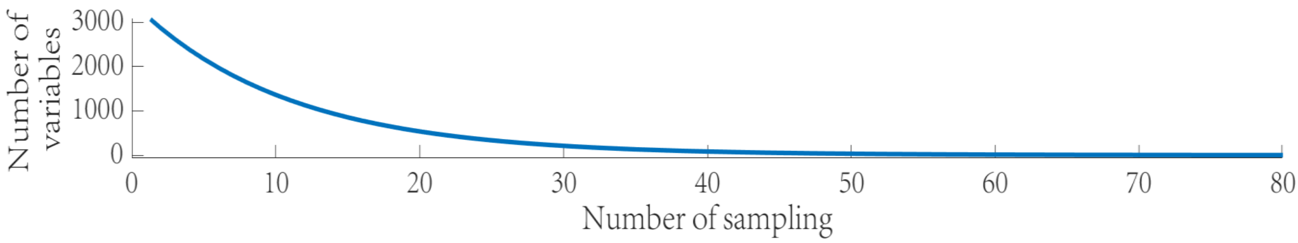
Characteristic Wavelength Extraction

The transmission spectral data acquisition system is used to obtain 3151 feature information; however, the number of these features is huge and contains irrelevant noise, not all features are beneficial to model building, irrelevant features will lead to a reduction in the accuracy of the model, and the huge feature latitude cannot meet the real-time, accuracy, and interpretability of the model detection (*Lakshmanan et al., 2023)*. So, in order to meet the requirements of fast and accurate modeling, improve the model interpretation ability, reduce the consumption of computational calculations, and enhance the model performance. In this paper, Competitive Adaptive Reweighted Sampling (CARS) and Successive Projection Algorithm (SPA) were utilized for feature wavelength extraction, and were combined with Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN) to build a binary classification model.

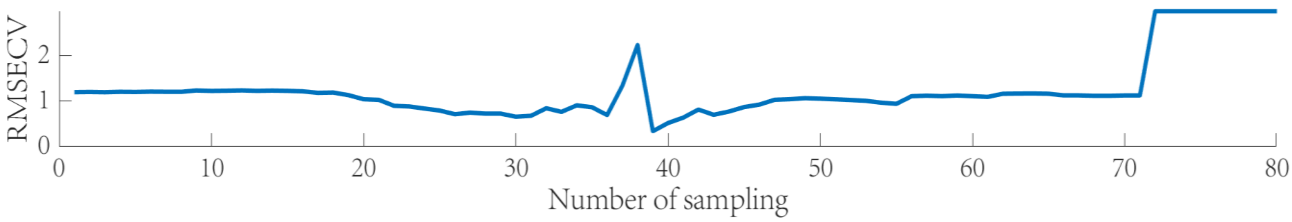
Competitive Adaptive Reweighted Sampling (CARS)

In this paper, the CARS algorithm was utilized to downscale the spectral data pre-processed by SG respectively, using 80 times Monte Carlo sampling and 5-fold cross-validation, and its running results are shown in Fig. 6.

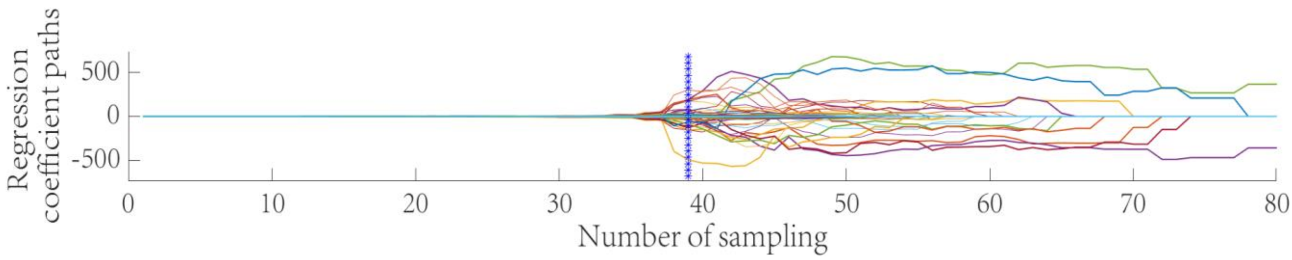
From Fig. 6b, it can be seen that the cross-validation RMS error gradually decreases in the first 30 samples, indicating that the feature variables with less relevance to the black hearted seed potatoes are eliminated; the fluctuation of the RMS error of the cross-validation is larger in the 30th to 39th samples, in which the 38th sampling eliminates some important parameters; then the minimum cross-validation RMS error is obtained in the 39th sampling, which achieves the optimal extraction of feature wavelengths, and 83 feature wavelengths are obtained at this time, accounting for 2.63% of the whole band.



a - Number of variables



b - Root mean square error of cross-validation



c - Regression coefficient paths

Fig. 6 - Variables selected by SG-CARS method

Successive Projection Algorithm (SPA)

The minimum value of the number of wavelengths is set to 1, and the maximum value is set to 50, and the spectral data after S-G preprocessing are downgraded, and as the number of selected feature wavelengths increases the root mean square error decreases, and finally 30 feature wavelengths are extracted, accounting for 0.95% of the full wavelength range, and at this time, the root mean square error is 0.168. The SPA selects this point because the feature wavelengths already contain the feature information of the original spectral data and has interpretability. The variables selected by SG-SPA method is shown in Fig. 7.

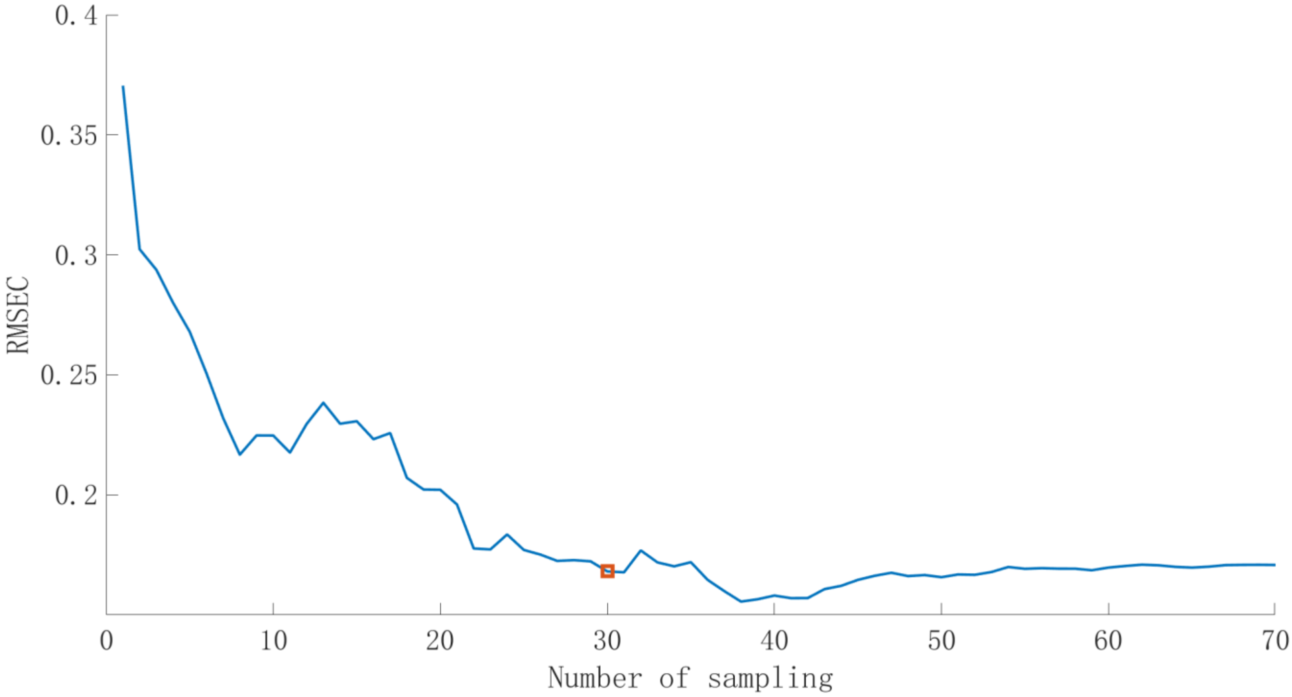


Fig. 7 - Variables selected by SG-SPA method

Modeling the classification of black heart and qualified potatoes

Data set segmentation

The KS algorithm (Kennard-Stone, KS) is used to divide the dataset, the core of which is to iteratively select the features that are least similar to those in the current training set, so that the samples in the training set are dispersed as much as possible as a way to cover the sample space more comprehensively and strengthen the interpretability of the model structure. The feature distribution of seed potatoes spectral data is complex, and the use of the KS algorithm can effectively capture the overall structure of seed potatoes spectral data and improve the stability and classification ability of the model. Using the KS algorithm to divide the data set according to the ratio of 3:1, 152 and 56 sample data were obtained for the training set and test set, respectively, and qualified seed potatoes were set as class 1 samples, and black hearted seed potatoes were set as class 2 samples.

CNN classification modeling

The model sets the initial learning rate of the CNN model to 0.01, the regularization parameter to 0.01, and the number of iterations to 100, and the SG preprocessed data are used to extract features and build the CNN classification model using CARS and SPA, respectively.

Table2

Combined classification model of different algorithms based on CNN

Algorithmic combination	Training set accuracy/%	Test set accuracy/%
RAW-CARS-CNN	90.38	82.69
RAW-SPA-CNN	83.97	88.46
SG-CARS-CNN	100	100
SG-SPA-CNN	98.72	100

The combined classification model of different algorithms based on CNN are shown in Table 2, the classification accuracy of the CNN model built after extracting the feature wavelengths of the unprepared data is lower than that of the CNN model after SG preprocessing, and the combination of SG-CARS-CNN algorithms has the highest classification accuracy, with the classification accuracy of 100% in both the training set and the test set. The SG-CARS-CNN training set and test set confusion matrix are shown in Fig. 8 and Fig. 9.

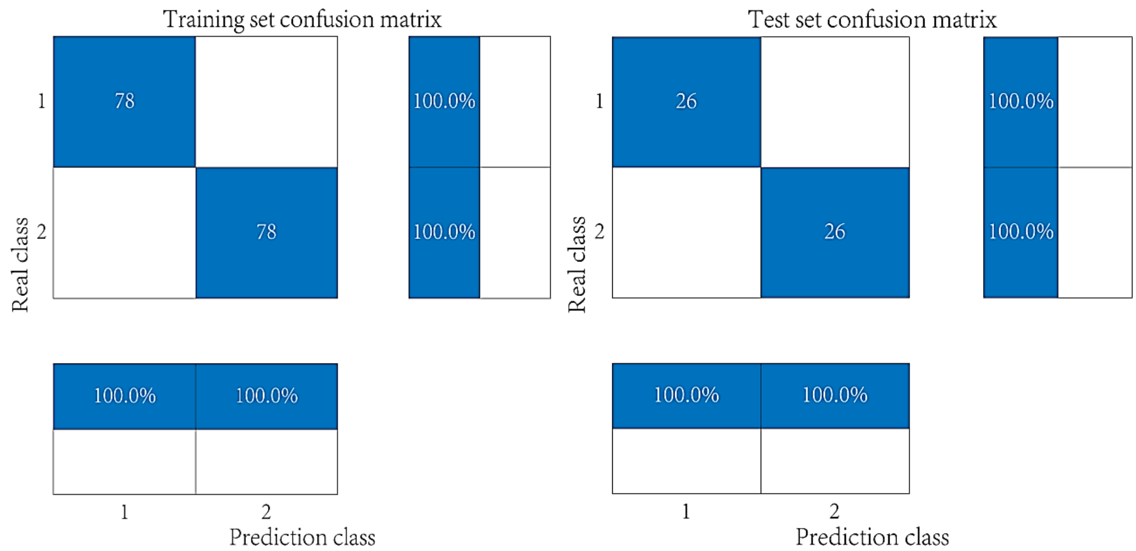


Fig. 8 - SG-CARS-CNN training set and test set confusion matrix

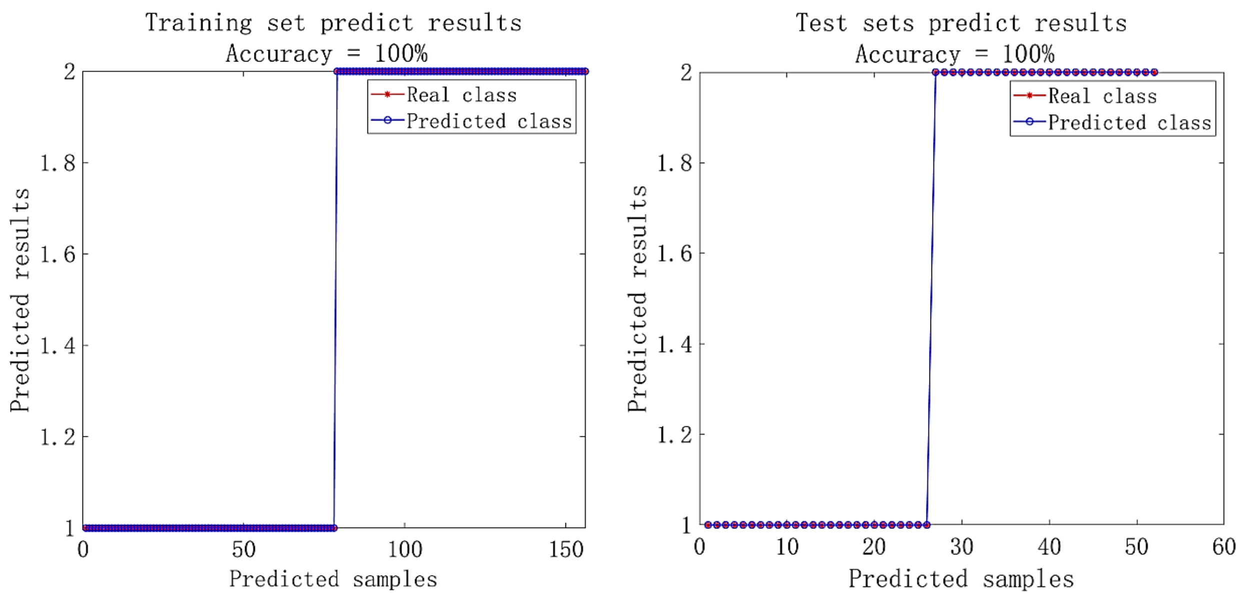


Fig. 9 - SG-CARS-CNN training set and test set accuracy

RNN classification modeling

The dataset division and model parameter settings are the same as CNN, and the RNN binary classification model is constructed using the SG preprocessed data after the feature wavelengths are extracted by CARS and SPA, respectively.

The SG-CARS-RNN training set and test set accuracy are shown in Fig. 10, the accuracy of the training set is 98.08%, in which all of the class 1 samples are successfully recognized, and three of the class 2 samples are misclassified as class 1 samples; the accuracy of the test set is 98.08%, in which all of the class 1 samples are accurately recognized, and one of the class 2 samples is misclassified as class 1 sample.

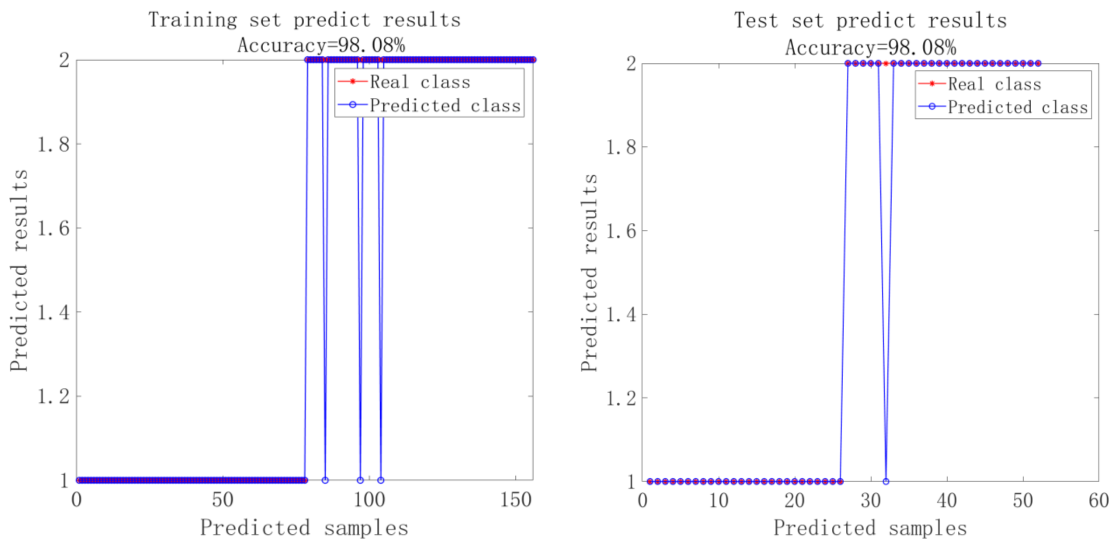


Fig. 10 - SG-CARS-RNN training set and test set accuracy

Table 3 shows the comparison of different combined classification models based on CNN and RNN, in the RNN model, the SG-CARS-RNN algorithm combination training set and test set have the highest classification accuracy, both of which are 98.08%; while in the CNN model, the SG-CARS-CNN algorithm combination training set and test set have the highest classification accuracy, both of which are 100%, which has an advantage of 1.92% over the RNN model.

Table3

Comparison of different combined classification models based on CNN and RNN

Model	Algorithmic combination	Training set accuracy/%	Test set accuracy/%
CNN	SG-CARS-CNN	100	100
	SG-SPA-CNN	98.72	100
RNN	SG-CARS-RNN	98.08	98.08
	SG-SPA-RNN	92.31	92.31

CONCLUSIONS

In this paper, the detection of black heart disease in seed potatoes was realized based on the fusion of transmission spectroscopy and neural network.

(1) Utilizing different preprocessing methods has an impact on the accuracy of the established classification model of qualified and black hearted seed potatoes, among the four preprocessing methods utilizing Savitzky-Golay preprocessing could achieve the highest classification accuracy, and the classification accuracy of the SG-SVM test set was 96.15%.

(2) Using CARS to extract the characteristic wavelengths of the seed potatoes spectral data pre-processed by Savitzky-Golay, 83 characteristic wavelengths were extracted from the original 3151 wavelengths, which accounted for 2.63% of the whole band; and using SPA to extract the characteristic wavelengths of the seed potato spectral data pre-processed by Savitzky-Golay, which finally extracted 30, which accounted for 0.95% of the whole band.

(3) After SG preprocessing, CNN and RNN classification models were built using CARS and SPA, respectively. Among the CNN models, the combined SG-CARS-CNN algorithm had the highest classification accuracy, reaching 100% in both the training and test sets; among the RNN models, the combined SG-CARS-RNN algorithm had the highest classification accuracy, with 98.08% in both its training and test sets. Although SPA extracted more concise feature wavelengths, none of the classification algorithms built using SPA had less accuracy than CARS, which is more suitable for feature wavelength extraction from seed potatoes spectral data.

The combined SG-CARS-CNN algorithm was finally used as a black heart disease detection model for seed potato, and its classification accuracy reached 100% in both the training and test sets. Therefore, this study can accurately classify qualified and black hearted seed potatoes and reach the black heart disease allowable rate of seed potato quality standard (GB18133-2000) formulated in China.

ACKNOWLEDGEMENTS

The work was supported by the National Natural Science Foundation of China (No. 62366041).

REFERENCES

1. Abukmeil, R., Al-Mallahi, A. (2022). In-field scanning of the spectrum potato plant leaves for a less destructive technique to estimate nutrients. *CSBE/SCGAB 2022 Annual Conference*, Charlottetown, Prince-Edward-Island.
2. Al-Adhaileh, M. H., Verma, A., Aldhyani, T. H., Koundal, D. (2023). Potato blight detection using fine-tuned CNN architecture. *Mathematics*, Vol. 11(6), pp. 1516. https://doi.org/10.3390/math11061516
3. Arshaghi, A., Ashourin, M., Ghabeli, L. (2021). Detection and Classification of Potato Diseases Potato Using a New Convolution Neural Network Architecture. *Traitement du Signal*, Vol. 38(6), pp.1783-1791. https://doi.org/10.18280/ts.380622
4. Gao, H.L., Li, X.Y., Xu, M.M., Huang, T., Tao, H.L., & Li, X.J. (2013). Transmission hyperspectral detection method for weight and black heart of potato (马铃薯黑心病和单薯质量的透射高光谱检测方法). *Transactions of the Chinese Society of Agricultural Engineering*, Vol.29(15), pp.279-285. (In Chinese). https:10.3969/j.issn.1002-6819.2013.15.034
5. GB/T 29377-2012 (2012). *Code of practice for virus free seed potatoes grading and inspection (马铃薯脱毒种薯级别与检验规程)*. (In Chinese).
6. Gupta, H.K., & Shah, H.R. (2023). Deep Learning-Based Approach to Identify the Potato Leaf Disease and Help in Mitigation Using IOT. *SN Computer Science*, Vol.4(4), pp.333. https://doi.org/10.1007/s42979-023-01758-5
7. Han, Y.F., Lv, C.X., Yuan, Y.W., Yang, B.N., Zhao, Q.L., Cao, Y.F., & Yin, X.Q. (2021). PLS-Discriminant Analysis on Potato Blackheart Disease Based on VIS-NIR Transmission Spectroscopy (PLS-DA 优化模型的马铃薯黑心病可见近红外透射光谱检测). *Spectroscopy and Spectral Analysis*, Vol. 41(04), pp.1213-1219. (In Chinese). https: 10.3964/j.issn.1000-0593(2021)04-1213-07
8. Huang, W., Zhao, Z., Sun, L., & Ju, M. (2022). Dual-branch attention-assisted CNN for hyperspectral image classification. *Remote Sensing*, Vol.14(23), pp.6158. https://doi.org/10.3390/rs14236158
9. Jiang, C.Y. (2024). China's No. 1 central document for 2024 (2024 年中央一号文件的突出亮点). *Rural Finance Research*, Vol. 2024(02), pp.3-13. https:10.16127/j.cnki.issn1003-1812.20240209.001
10. Khorramifar, A., Sharabiani, V. R., Karami, H., Kisalaei, A., Lozano, J., Rusinek, R., & Gancarz, M. (2022). Investigating Changes in pH and Soluble Solids Content of Potato during the Storage by Electronic Nose and Vis/NIR Spectroscopy. *Foods*, Vol. 11 (24), pp.4077-4077. https://doi.org/10.3390/foods11244077
11. Lakshmanan, M. K., Boelt, B., & Gislum, R. (2023). A chemometric method for the viability analysis of spinach seeds by near infrared spectroscopy with variable selection using successive projections algorithm. *Journal of Near Infrared Spectroscopy*, Vol.31(1), pp.24-32. https://doi.org/10.1177/09670335221138955
12. Li, X.T., Zhang, F., Feng, J. (2024). Convolutional Neural Network Combined with Improved Spectral Processing Method for Potato Disease Detection (卷积神经网络结合改进光谱处理方法用于马铃薯病害检测). *Spectroscopy and Spectral Analysis*, Vol. 44(01), pp. 215-224. (In Chinese). https:10.3964/j.issn.1000-0593(2024)01-0215-10
13. Liu, W.Z., Zhou, X.J., Ping, F.J., Su, Y., Ju, Y.L., Fang, Y.L., Yang, J.H. (2024). Detection of Key Indicators of Ripening Quality in Table Grapes Based on Visible-near-infrared Spectroscopy (基于可见-近红外光谱的鲜食葡萄成熟品质关键指标检测). *Transactions of the Chinese Society for Agricultural Machinery*, Vol. 55(02), pp. 372-383. (In Chinese). https:10.6041/j.issn.1000-1298.2024.02.037
14. Rogers, M., Blanc-Talon, J., Urschler, M., & Delmas, P. (2023). Wavelength and texture feature selection for hyperspectral imaging: a systematic literature review. *Journal of Food Measurement and Characterization*, Vol. 17(6), pp. 6039-6064. https://doi.org/10.1007/s11694-023-02044-x
15. Tiwari, D., Ashish, M., Gangwar, N., Sharma, A., Patel, S., Bhardwaj, S. (2020). Potato leaf diseases detection using deep learning. *2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS)*, Vol. 2020, pp. 461-466. https:10.1109/ICICCS48265.2020.9121067
16. Vallese, Federico Danilo; Garcia Paoloni, Maria Soledad; Springer, Valeria; Fernandes, David Douglas de Sousa; Diniz, Paulo Henrique Gonçalves Dias; Pistonesi, Marcelo Fabián. (2024). Exploiting the successive projections algorithm to improve the quantification of chemical constituents and

discrimination of botanical origin of Argentinean bee-pollen [J]. *Journal of Food Composition and Analysis*, https://doi.org/10.1016/j.jfca.2023.105925.

17. Wang, F., Li, Y.Y., Peng, Y.K., Yang, B.N., Li, L., Yin, X.Q. (2018). Hand-held Device for Non-destructive Detection of Potato Quality Parameters (便携式马铃薯多品质参数局部透射光谱无损检测装置). *Transactions of the Chinese Society for Agricultural Machinery*, Vol. 49(07), pp. 348-354. (In Chinese). https: 10.6041/j.issn.1000-1298.2018.07.042
18. Zhou, Z., Li, X.Y., Gao, H.L., Tao, H.L., & Li, P. (2012). Comparison of diffuse reflection and transmission mode of visible/near infrared spectroscopy for detecting black heart of potato (漫反射和透射光谱检测马铃薯黑心病的比较). *Transactions of the Chinese Society of Agricultural Engineering*, Vol. 28(11), pp. 237-242. (In Chinese)