Identification of Disease Type of Tobacco Leaves Based on Near Infrared Spectroscopy and Convolutional Neural Network

Liang Ying,[a] Ma Kun,[a] Zhang Xinyu,[a] Yang Qifu,[a] Wu Jiaquan*,[a] and Yang Shuangyan*,[b]

[a]*College of Science, Kunming University of Science and Technology, Kunming 650500, China*

[b]*Yunan Tobacco Biological Technology Co., Ltd., Kunming 650500, China*

It is important to identify the types of tobacco diseases accurately and take effective control measures in time to improve the efficiency of tobacco planting. In this paper, a hand-held near-infrared spectrometer was used to collect the spectral data of different types of tobacco disease samples. The training models were established via convolutional neural network algorithm. Meanwhile, the traditional classification algorithms support vector machine and back propagation neural network were also compared. The results showed that the prediction accuracy of convolutional neural network algorithm was the highest and the overall performance of the model was the best. The rapid detection method based on a hand-held near-infrared spectrometer and convolutional neural network algorithm could identify tobacco leaf disease species efficiently, non-destructively, quickly and accurately, which provided a new technical reference for tobacco leaf disease species detection and identification.

Keywords: hand-held near-infrared, convolutional neural network, tobacco leaf, disease identification

Introduction

Tobacco is an important economic crop in China. Tobacco diseases have a great influence on the quality of tobacco. Besides, it also affects the economic development of the tobacco industry and the income of tobacco farmers. Therefore, effective identification of tobacco diseases is essential to ensure the physiological health of the leaves and improve their quality. There is a wide variety of tobacco leaf diseases and the pathological mechanism is complex. The main diagnosis methods of tobacco diseases are manual identification and laboratory tests. Manual identification is poor in accuracy and low in efficiency.[1,2] Laboratory tests have high accuracy, but it is expensive and has long periods of analysis. As a result, it is essential to develop a novel disease identification method which is fast and low-cost.

Near-infrared (NIR) spectroscopy technique is widely used in varied fields, such as agriculture, petrochemical industries, medicine, etc.[3-6] NIR has the advantages of being rapid, non-destructive, green and has low cost for sample analysis.[7] It has already been used in tobacco area. Li *et al.*[8] have identified the producing areas of the flue-cured tobacco leaves rapidly and non-destructively by using a (NIR) spectrometer and a multi-layer-extreme learning machine (ML-ELM) algorithm. Lu *et al.*[9] have proposed a non-destructive discrimination method based on NIR spectroscopy to evaluate the quality of raw intact tobacco leaves and explore the application of near-infrared technology. Jianqiang *et al.*[10] have classified the grades of tobacco leaves by using the near-infrared spectroscopy device. Zhang *et al.*[11] used a hand-held NIR spectrometer to detect the deep green infected of the tobacco leaf. However, it only provides a single disease detection method for tobacco leaf via NIR technology. The tobacco leaf has about 10 different types of diseases.[12] Therefore, it is of great importance to assess the value of this tool for more plant disease diagnosis with novel approaches.

As one of the representative algorithms of deep learning, convolutional neural network (CNN) is a kind of algorithm including convolution calculation and deep structure feedforward neural network. CNN has been developed rapidly, especially in the field of image classification, action recognition, satellite remote sensing and atmospheric science, etc.[13-16] Nevertheless, CNN has insufficiently been

used in the field of spectrum analysis. The dimension of NIR spectral data is huge, and the measuring is low. It also has large relative error and noise. CNN algorithm can extract higher dimensions feature information automatically, has fewer parameters and is more robust. Hence, NIR combined with CNN can identify disease types of tobacco leaf effectively.

This paper proposed a novel method that can identify 5 different diseases of tobacco leaves rapidly by using NIR technology and CNN algorithm. The method can identify the disease type of tobacco leaves in the field accurately, rapidly and nondestructively.

Experimental

Experimental samples

The experimental samples were collected from the tobacco field in Longtankou village, Yunnan province, China, which was located between 102°21'-102°47' E and 25°08'-25°36' N. There were a total of 460 samples and it included health and infected leaves (powdery mildew, deep green, mosaic virus, and brown-spot). The numbers of the five different types of samples were 80, 100, 80, 120 and 80 for health, powdery mildew, deep green, mosaic virus, and brown-spot, respectively. All the infection samples were sent to the laboratory to determine the disease type. The test stands were as follows: powdery mildew (YC/T 341.7-2010),[17] deep green (YCT 161-2002),[18] mosaic virus (DB53/T 348-2011),[19] brown-spot (YC/T 341.1-2010).[20]

Spectral acquisition

The MicroNIR (VIAVI, Beijing, China) is equipped with a 128-pixel detector array, which records data. The system is composed by two small tungsten light bulbs as the radiation source and a linear-variable filter (LVF) directly connected to a linear indium gallium arsenide (InGaAs) array detector. Spectral data from tobacco samples were collected using a MicroNIR handheld near-infrared spectrometer. The wavelength range and the
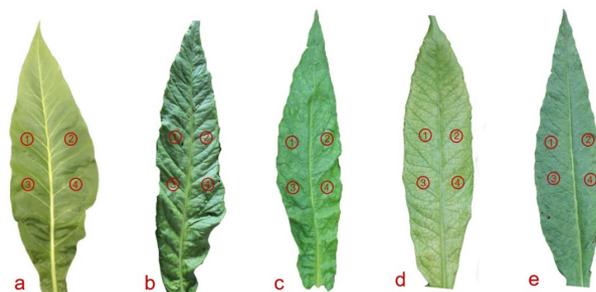
Figure 1. Positions where the near-infrared spectrum scans were performed on the tobacco leaf samples (a) healthy, (b) powdery mildew, (c) deep green, (d) mosaic virus (e) brown-spot.

spectral resolution of the device were 908-1676 nm and 4 cm⁻¹, respectively. The data sample interval was set as 6 nm and the integration time was 9.6 µs. A 99% diffuse reflective white board was placed under the tobacco leaf sample. 4 reflectance spectral data of each sample were collected as shown in Figure 1 and the average of the 4 spectral data was set as the final data. The samples were randomly divided into training set and test set. The ratio of training set and test set was 6:4. The details are shown in Table 1.

Theory of algorithms

Support vector machine algorithm

Support vector machine (SVM) is a binary classification model and the basic model is defined as a linear classifier with the largest interval on the feature space.[21] The radial basis function (RBF) kernel is the most commonly used kernel function in support vector machine classification. RBF based on support vector machines,[22] which a supervised learning algorithm used to solve classification problems.[23] It can solve high-dimensional problems, has high generalization ability, and does not need to rely on the entire data. The following interpolation conditions:

$$F(X) = d^P, P = 1, 2, 3, …, p \qquad (1)$$

where, the N dimensional space have P input vectors P = 1, 2, 3, …, p, and the corresponding target values d^P, P = 1, 2,

Table 1. Samples of different types of tobacco leaf diseases

Type	Year	Regional sources	Spectral dimension / nm	Total samples	Training samples	Test samples
Health	2022	Longtankou Village, Fumin County, Yunnan Province	908-1679	80	44	36
Powdery mildew	2022	Longtankou Village, Fumin County, Yunnan Province	908-1679	100	64	36
Deep green	2022	Longtankou Village, Fumin County, Yunnan Province	908-1679	80	48	28
Mosaic virus	2022	Longtankou Village, Fumin County, Yunnan Province	908-1679	120	76	44
Brown-spot	2022	Longtankou Village, Fumin County, Yunnan Province	908-1679	80	44	36

3, …, p, P in the output space constitutes the training sample set, and F(X) is nonlinear mapping function.

In the formula, the function F describes an interpolation surface that must pass through all the training data points. P basis functions are selected as the training data, and each basis function is of the form:

$$\varphi(\|X - X^P\|), P = 1, 2, 3, …, p \qquad (2)$$

where, the basis function φ is a nonlinear function and the training data point X^P is the center of φ. The basis function takes the distance between the point X of the input space and the center X^P as the independent variable of the function. The difference function based on the radial basis function technique is defined as a linear combination of basic functions:

$$F(x) = \sum_{P=1}^{P} \omega_P \varphi\left(\left\|x - x^P\right\|\right) \qquad (3)$$

where w_p is an uncertainty factor about P, the $\left\|x - x^P\right\|$ is norm, X^P is center point, x is data.

Back propagation neural network algorithm

Back propagation (BP) neural network is a multilayer feedforward network trained by error back propagation.[24] The form is:

$$g(x) = \frac{1}{1 + e^{-x}} \qquad (4)$$

where, g(x) is the excitation function.

(*i*) The output of the implicit layer H_j :

$$H_j = g\left(\sum_{i=1}^{n} w_{ij} x_i + a_j\right) \qquad (5)$$

where, a_j is the bias of the input layer to the implicit layer, w_{ij} is the weight of the input layer to the implicit layer.

(*ii*) Output layer output:

$$O_k = \sum_{j=1}^{l} H_j w_{jk} + b_k \qquad (6)$$

where, w_{jk} is the weight of the implicit layer to the output layer, b_k is the bias of the implicit layer to the output layer, l is hidden layer.

(*iii*) Error calculation:

$$E = \frac{1}{2}\sum_{k=1}^{m}(Y_K - O_K)^2 \qquad (7)$$

where m are the numbers of node in the input layer, Y_K is the desired output and $Y_k - O_K = e_k$, E can be expressed as:

$$E = \frac{1}{2}\sum_{k=1}^{m}e_k^2 \qquad (8)$$

where $k = 1… m$.

(*iv*) Weight update:

$$\begin{cases} w_{ij} = w_{ij} + \eta H_j\left(1 - H_j\right)x_i \sum_{k=1}^{m} w_{jk} e_k \\ w_{jk} = w_{jk} + \eta H_j e_k \end{cases} \qquad (9)$$

where, η is the learning rate, w_{ij} is the weight of the input layer to the implicit layer, w_{jk} is the weight of the implicit layer to the output layer, H_j is the hidden output.

Convolutional neural network algorithm

CNN is a supervised learning algorithm, and it is based on end-to-end supervised learning by backpropagation and removes the focus from built-in invariance mechanisms, using pooling not as a way to tolerate small shifts but as a regularization tool that decreases model complexity.[25] The CNN is a class of feedforward neural networks that include convolutional computation and have a deep structure, and is one of the representative algorithms of deep learning.[26,27] The CNN contains input layer, convolutional layer, pooling layer, fully connected layer, and output layer of softmax activation function.

(*i*) Input layer propagates forward to the convolutional layer;

The dimensionality of the convolution kernel and the number of submatrices of the input tensor are the same. The algorithm for forward propagation is:

$$a^2 = \sigma\left(z^2\right) = \sigma\left(a^1 * W^2 + b^2\right) \qquad (10)$$

where the superscript is the layer, a is the input of the layer, W is the weight matrix of the layer, and b is the bias of the layer, z is the intermediate output of the layer's input, σ is the activation function, and * is the convolution.

(*ii*) Forward propagation of the hidden layer to the convolutional layer;

All convolutional and pooling layers including fully connected layers form the hidden layer, the formula is:

$$a^1 = \sigma\left(z^1\right) = \sigma\left(a^{l-1} * W^1 + b^1\right) \qquad (11)$$

J. Braz. Chem. Soc. **2024**, *35*, 5, e-20230181

3 of 8

where, a^l is the input to layer l, σ is the activation function, z^l is the intermediate output of the l layer's input, a^{l-1} is the input of the $l-1$ layer, W^l is the weight matrix of the l layer, and b is the bias of the layer.

After convolution of M submatrices:

$$a^l = \sigma\left(z^l\right) = \sigma\left(\sum_{K=1}^{M} z_k^l\right) = \sigma\left(\sum_{K=1}^{M} a_k^{l-1} * W_k^l + b^l\right) \tag{12}$$

where, M is the number of matrices, a^l is the input to layer l, σ is the activation function, z^l is the intermediate output of the l layer's input, a^{l-1} is the input of the $l-1$ layer, W^l is the weight matrix of the l layer, and b is the bias of the l layer, k is the size of the convolution kernel.

(*iii*) Forward propagation of the hidden layer to the pooling layer;

The input matrix is $N \times N$ dimensional, the pooled region is of size K × K of output matrix is $\dfrac{N}{K} \times \dfrac{N}{K}$ dimensional. The size of the CNN pooling region is K, and the pooling criterion where, the Max is max Pooling, the Averge is Average Pooling.

(*iv*) Forward propagation of hidden layers to fully connected layers

$$a^l = \sigma\left(z^l\right) = \sigma\left(w^l a^{l-1} + b^l\right) \tag{13}$$

where, σ is the activation function, z^l is the intermediate output of the l layer's input, a^{l-1} is the input of the $l-1$ layer, W^l is the weight matrix of the l layer, and b is the bias of the l layer.

Measures of classification performance

Confusion matrix is a visualization tool for supervised learning, unsupervised learning is generally called a matching matrix. Figure 2 shows the basic form of the confusion matrix. In Figure 2 and equations 14-17, TP is true positive, FN is false negative, FP is false positive and TN is true negative.

Accuracy is the proportion of total observations for which all judgments of the classification model are correct:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{14}$$

The precision rate is the ratio of the number of positive samples correctly classified to the number of all samples divided by the classifier:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{15}$$

Sensitivity is the ratio of the number of correctly classified positive samples to the total number of samples:

$$\text{Sensitivty} = \frac{TP}{TP + FN} \tag{16}$$

Specificity is the correct proportion of actual negatives measured:

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{17}$$

Confusion matrix		Predicted value	
		Positive	Negative
True value	Positive	TP	FN
	Negative	FP	TN

Figure 2. Confusion matrix of the two-category task.

The precision, recall, and specificity shown above only calculate the characteristics of a certain classification, while accuracy and balanced *F* score (*F*1-score) and subject operating characteristic curve (ROC) can evaluate the overall criteria of the classification model. The output results of the *F*1-score metric synthesize precision and sensitivity, and are the harmonized average of precision and sensitivity. The *F*1-score takes values from 0 to 1. The higher the *F*1-score, the better the model performance, which is calculated as follows:

$$F_1 - \text{score} = \frac{2 \times \text{Precision} \times \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}} \tag{18}$$

The receiver operating characteristic (ROC), which responds to the classification ability of the model is a graphical line that can visualize the classification effect of the classifier. The closer the ROC curve is to the upper left corner, the better is the classification of the classifier. The performance of the model is indicated by the area under curve (AUC), which takes values between 0.5 and 1. The larger the AUC value, the better the classification performance of the model.

Results and Discussion

Data pre-processing

The raw spectral data contained both the information of the samples and the noise. The pre-procession

operation could reduce the influence of noise and enhance the experimental ability of the model. Here, Savitzky-Golay + 1st derivatives (SG + D1), Standard Normal Variate (SNV), Multiplicative Signal Correction (MSC), and Savitzky-Golay (SG) were chosen to establish the models of tobacco disease types by using SVM algorithm. The results with different pre-processing methods are shown in Table 2. In Table 2, the accuracies of training set and test set were higher when using the SG + D1 pre-procession operation. Therefore, the SG + D1 algorithm pre-procession operation was used before building the training models for each algorithm in the follow-up study. Figure 3 shows the original spectral data and the pre-processing result after the SG + D1 operation. Here, SVM algorithm was chosen to build the training model, the parameters of SVM algorithm were set as: the kernel function was RBF, penalty parameter C = 10.0, hyper-parameter gamma = 0.01.

Table 2. Accuracies of training set and test set based on SVM algorithm with different preprocessing methods, the training set and test set include healthy, powdery mildew, black streak, mosaic virus, brown-spot

Algorithm	Preprocessing method	Training set accuracy / %	Test set accuracy / %
	SNV	63.76	64.13
SVM	MSC	69.92	67.93
	SG	70.12	64.52
	SG + D1	74.27	67.93

SVM: support vector machines; SNV: Standard Normal Variate; MSC: Multiplicative Signal Correction; SG: Savitzky-Golay; SG+D1: Savitzky-Golay + 1st derivatives.

Construction of the qualitative model

SVM, BP, CNN algorithms were used to establish the qualitative model of the types of tobacco diseases. The accuracy, sensitivity and specificity were used as the model evaluation indices. Table 3 shows the performance of the training model for different types of tobacco leaf disease using different modeling approaches. Here, the data was normalized in the input layer during the training process of the CNN algorithm. The method could strengthen the training effect of the model and improve the generalization ability of the model. Besides, regularization parameter was used in the CNN model and it was set as 1×10^{-4}. The loss function did not rise in the process of running the CNN algorithm. It indicated that the CNN model was not overfitting. The classification precision, sensitivity, specificity and accuracy of CNN training model were all 100%, which were higher than those of SVM and BP training models.

The prediction performance of SVM, BP and CNN algorithms is shown in Table 4 in the form of a confusion matrix. It can be seen from Table 4 that the average prediction accuracies of SVM, BP and CNN algorithms were 67.93, 85.87 and 98.91%, respectively. The prediction accuracy of CNN algorithm was 30.98 and, 13.04% higher than that of SVM and BP, respectively. Besides, the precision, sensitivity and *F*1-score contained in each base classifier's confusion matrix, the CNN algorithm occupied the highest accuracy. It meant CNN model had a higher recognition capability and a lower misdiagnosis rate. The reason was that the CNN algorithm could extract the features automatically and reduce the number of parameters. The convolutional layer could extract higher level spectral features.[28] Deep learning could obtain the multiscale feature of crop diseases, and realize the characteristic expression of different diseases.[29] In summary, the CNN algorithm could improve the robustness and accuracy of the model. The above results showed that the CNN algorithm had a better performance to build the models for the different types of tobacco disease with NIR spectral data.

Model evaluation

The classification performance of the three algorithms
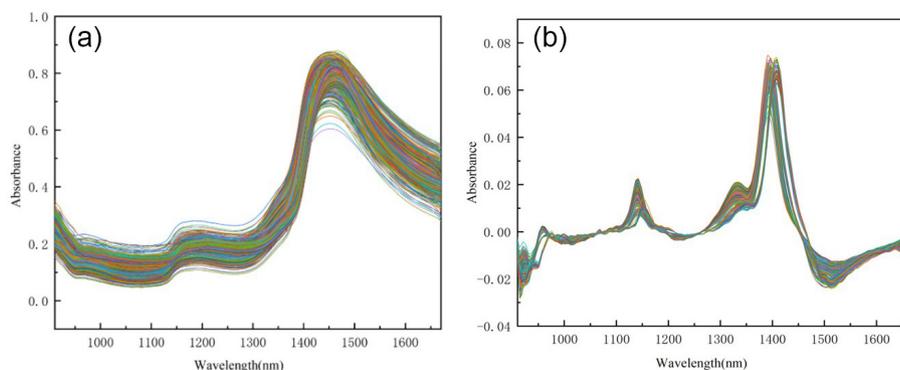
Figure 3. The original near infrared spectroscopy data and after pre-processing near infrared spectroscopy data, including healthy, powdery mildew, deep green, mosaic virus, brown-spot total data. (a) The original spectral data and (b) pre-processing results.

J. Braz. Chem. Soc. **2024**, *35*, 5, e-20230181

5 of 8

Table 3. Comparison model effects of training models for different types of tobacco leaf disease

Algorithm	Type	Training samples	Training correct number	Number of false positives	PR / %	SN / %	SP / %	*F*1-score	AC / %
SVM	healthy	51	50	1	98.1	100	99.6	0.99	
	powdery mildew	35	28	7	81.4	55.6	96.2	0.66	
	deep green	23	7	6	71.9	53.5	96.1	0.62	74.28
	mosaic virus	43	22	21	51.8	66.2	85.8	0.58	
	brown-spot	53	43	10	80.3	98.1	94.1	0.88	
BP	healthy	43	39	4	91.3	97.7	98.3	0.94	
	powdery mildew	59	57	2	96.4	91.5	99.1	0.94	
	deep green	52	45	7	85.7	80.8	96.9	0.83	91.67
	mosaic virus	70	67	3	95.6	92.9	98.5	0.94	
	brown-spot	52	45	7	87.7	96.2	96.9	0.92	
CNN	healthy	45	45	0	100	100	100	1.00	
	powdery mildew	54	54	0	100	100	100	1.00	
	deep green	52	52	0	100	100	100	1.00	100
	mosaic virus	70	70	0	100	100	100	1.00	
	brown-spot	55	55	0	100	100	100	1.00	

SVM: support vector machines; BP: back propagation; CNN: convolutional neural networks; NIR: near-infrared; PR: precision; SN: sensitivity; SP: specificity; AC: accuracy.

Table 4. Comparison model effects of test models for different types of tobacco leaf disease

Sample	Algorithm	Type	Test samples	Test correct number	Number of false positives	PR / %	SN / %	SP / %	*F*1-score	AC / %
Tobacco	SVM	healthy	29	29	0	100	100	100	1.00	
		powdery mildew	16	12	4	76.2	43.2	96.6	0.55	
		deep green	18	11	5	72.0	48.6	95.2	0.58	67.93
		mosaic virus	38	21	17	55.9	69.1	76.7	0.62	
		brown-spot	24	14	10	58.5	92.3	89.2	0.72	
	BP	healthy	37	32	5	88.1	100	96.6	0.94	
		powdery mildew	41	38	3	91.9	82.9	97.9	0.87	
		deep green	28	22	6	80.0	57.1	97.4	0.67	85.87
		mosaic virus	50	45	5	89.6	86.0	96.3	0.88	
		brown-spot	28	21	7	75.7	100	94.2	0.87	
	CNN	healthy	35	35	0	100	100	100	1.00	
		powdery mildew	45	45	0	100	97.8	100	0.99	
		deep green	27	26	1	96.4	96.4	99.4	0.96	98.91
		mosaic virus	50	49	1	98.0	100	99.3	0.99	
		brown-spot	25	25	0	100	100	100	1.00	

SVM: support vector machines; BP: back propagation; CNN: convolutional neural networks; NIR: near-infrared; PR: precision; SN: sensitivity; SP: specificity; AC: accuracy.

was evaluated by AUC values. Table 5 shows AUC values of the three different classification algorithms. It could be seen that AUC of SVM classification algorithm was 1 for healthy type and the other types were all less than 0.8. The AUC of BP algorithm was 0.97 for healthy type, 0.87 for powdery mildew, 0.62 for black streak, 0.89 for mosaic, and 0.96 for brown-spot. The AUC of CNN classification algorithm was 1 for both healthy type and brown-spot, 0.98, 0.97, and 0.99 for powdery mildew, deep green, and mosaic disease, respectively. It could be easily seen that the AUC of CNN algorithms was higher than that of SVM and BP algorithms.

The ROC curve is also known as receiver operating characteristic curve. The model classification performance of CNN was further evaluated via ROC curve. Figure 4 shows the ROC curves of CNN for different types of tobacco diseases. It could be seen that the ROC curve area was close to 1, indicating that the CNN algorithm had good performance in the recognition model of tobacco disease types.

Conclusions

This paper proposed a novel method based on NIR technology and CNN algorithm to identify different types of tobacco disease rapidly, accurately, and non-destructively.

Table 5. AUC values of SVM, BP and CNN classification algorithms

	Algorithm	Health	Powdery mildew	Deep green	Mosaic virus	Brown-spot
	SVM	1.00	0.75	0.74	0.69	0.73
AUC	BP	0.97	0.87	0.62	0.89	0.96
	CNN	1.00	0.98	0.97	0.99	1.00

AUC: area under curve; SVM: support vector machines; BP: back propagation neural network; CNN: convolutional neural network.
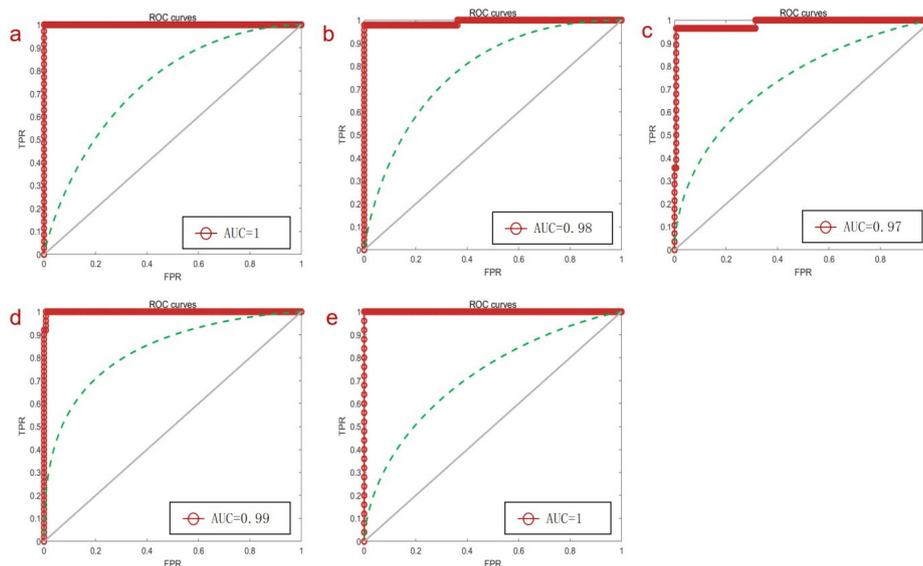
Figure 4. ROC curves of different disease types based on convolutional neural network algorithm. (a) Healthy, (b) powdery mildew, (c) deep green, (d) mosaic virus, (e) brown-spot.

This method is helpful for farmers to make appropriate decisions in precisely controlling the types of tobacco disease in the field. As the method is rapid, simple and can be used directly in the field, it provides a new technology reference for the diagnosis of diseases in various plant species.

Acknowledgments

Author Contributions

Liang Ying wrote and revised the original manuscript; Zhang Xinyu, Wu Jiaquan and Ma Kun provided theoretical guidance and overall planning; Yang Qifu proofread the manuscript; Yang Shuangyan provided theoretical guidance and overall planning.

References

1. Duraisamy, K.; Ha, A.; Kim, J.; Park, A. R.; Kim, B.; Song, C. W.; Song, H.; Kim, J. C.; *J. Plant Pathol.* **2022**, *38*, 182. [Crossref]

2. Yang, Y.; Zhou, Q.; Zahr, K.; Harding, M. W.; Feindel, D.; Feng, J.; *Eur. J. Plant Pathol.* **2021**, *159*, 583. [Crossref]

3. Tsuchikawa, S.; Ma, T.; Inagaki, T.; *Anal. Sci.* **2022**, *38*, 635. [Crossref]

4. Yu, H.; Wang, X.; Shen, F.; Long, J.; Du, W.; *IEEE Trans. Instrum. Meas.* **2022**, *316*, 123101. [Crossref]

5. Zheng, Q.; Huang, H.; Zhu, S. P.; Qi, B. H.; Tang, X.; *J. Near Infrared Spectrosc.* **2023**, *31*, 63. [Crossref]

6. Jamrógiewicz, M.; *J. Pharm. Biomed. Anal.* **2012**, *66*, 1. [Crossref]

7. Watanabe, A.; Furukawa, H.; Miyamoto, S.; Minagawa, H.; *Constr. Build. Mater.* **2019**, *196*, 95. [Crossref]

8. Li, R.; Huang, W.; Shang, G.; Zhang, X.; Wang, X.; Liu, J.; Wang, Y.; Qiao, J.; Fan, X.; Wu, K.; Zi, W.; *J. Braz. Chem. Soc.* **2022**, *33*, 251. [Crossref]

9. Lu, M.; Zhou, Q.; Chen, T. E.; Li, J.; Jiang, S.; Gao, Q.; Chen, D.; *J. Spectrosc.* **2021**, *2021*, ID 8807199. [Crossref]

10. Jianqiang, Z.; Panpan, Y.; Weijuan, L.; Yanmei, Y.; Tianjun, Y.; Ying, H.; Changyu, L.; *J. Braz. Chem. Soc.* **2019**, *30*, 1927. [Crossref]

11. Zhang, J. Q.; Liu, Y.; He, Y. F.; Hu, G. Y.; Bai, N. N.; *Anal. Lett.* **2020**, *53*, 2266. [Crossref]

12. Zhao, R. H.; Han, Z. F.; Jia, H. W.; Wang, J. G.; *Plant Doctor.* **2022**, *23*, 5 (in Chinese). [Crossref]

13. Chen, G.; Chen, Q.; Long, S.; Zhu, W.; Yuan, Z.; Wu, Y.; *Pattern Anal. Appl.* **2022**, *26*, 655. [Crossref]

14. Yang, H.; Yuan, C.; Li, B.; Du, Y.; Xing, J.; Hu, W.; Maybank, S. J.; *Pattern Recognit.* **2019**, *85*, 1. [Crossref]

15. Aires, F.; Boucher, E.; Pellet, V.; *Remote. Sens. Environ.* **2021**, *263*, 112553. [Crossref]

16. Cazeneuve, D.; Marchis, F.; Blaclard, G.; Dalba, P. A.; Martin, V.; Asencio, J.; *Astron J.* **2022**, *165*, 11. [Crossref]

17. YC/T 341.7-2010: *Rules for Investigation and Forecast of Tobacco Diseases-Part 7: Tobacco Powdery Mildew*, State Tobacco Monopoly Administration: Beijing, 2010 (in Chinese). [Link] accessed in October 2023

18. YC/T 161-2002: *Tobacco and Tobacco Products. Determination of Total Nitrogen. Continuous Flow Method*, State Tobacco Monopoly Administration: Beijing, 2002 (in Chinese). [Link] accessed in October 2023

19. DB53/T 348-2011: *Rules for Tobacco Intensive Seedling Production - Part 3: Seedling with Sand Culture*, Quality and Technical Supervision Bureau of Yunnan Province: Yunnan, 2011 (in Chinese). [Link] accessed in October 2023

20. YC/T 341.1-2010: *Rules for Investigation and Forecast of Tobacco Diseases. Part 1: Tobacco Brown Spot*, State Tobacco Monopoly Administration: Beijing, 2010 (in Chinese). [Link] accessed in October 2023

21. Nkengfack, L. C. D.; Tchiotsop, D.; Atangana, R.; Louis-Door, V.; Wolf, D.; *Biomed. Signal Process. Control.* **2020**, *62*, 102141. [Crossref]

22. Schaback, R.; *Constr. Approx.* **2005**, *21*, 293. [Crossref]

23. Mekni, N.; Coronnello, C.; Langer, T.; Rosa, M. D.; Perricone, U.; *Int. J. Mol. Sci.* **2021**, *22*, 7714. [Crossref]

24. Qian, G.; Zhang, L.; *Appl. Soft. Comput.* **2018**, *70*, 1034. [Crossref]

25. Sa-couto, L.; Wichert, A.; *Neural Comput.* **2021**, *33*, 34. [Crossref]

26. Ghiasi-Shirazi, K.; *Neural. Process. Lett.* **2019**, *50*, 2627. [Crossref]

27. Benayache, A.; Lamoudi, L.; Daoud, K.; *J. Coat. Technol. Res.* **2023**, *20*, 485. [Crossref]

28. Li, G.; Tang, H.; Sun, Y.; Kong, J.; Jiang, G.; Jiang, D.; Tao, B.; Xu, S.; Liu, H.; *Cluster Comput.* **2019**, *22*, 2719. [Crossref]

29. Huang, Z.; Xu, X.; Zhu, H.; Zhou, M. C.; *IEEE Trans. Neural Networks Learn. Syst.* **2020**, *31*, 4461. [Crossref]