Supplemental File

Zhicheng Jia [1,*], Qifeng Duan [1], Yue Wang [1], Ke Wu [2] and Hongzhe Jiang [1]

1 College of Mechanical and Electronic Engineering, Nanjing Forestry University, Nanjing 210037, China;

2 College of Science, Nanjing Forestry University, Nanjing 210037, China

* Correspondence: jzcnfu@njfu.edu.cn; Tel.: +86 13705176940

1. Elements used by preprocessing combinations

Table S1. Preprocessing Methods for Hyperspectral Data Analysis in Detecting Poplar Tree Diseases

Experiment	Smoothing	Baseline correction	Scatter correction	Scaling
1	Savitzky-Golay(SG)	airPLS	Variable Sorting for Normalization(VSN)	Mean-centering
2	Gaussian-weighted moving average	Continuous Wavelet Transform(CWT)	multiplicative scatter correction (MSC)	Mean-centering

2. All model results based on 27 preprocessing combinations

From the preprocessed data, we modelled and evaluated the accuracy and F1-score. A total of 108 models were evaluated using eight metrics across 27 preprocessed datasets.

(1) The classification accuracy and F1-score of RF, SVM, LSTM, and 1DCNN on the original dataset are less than 0.8. However, after applying the preprocessing tech-niques of Gaussian+CWT+VSN, Gaussian+CWT+MSC, and Gaussian+No pro-cessing+MSC, the accuracy and F1-score increased to 0.84 or higher. These results demonstrate that data preprocessing leads to an improvement in model performance.

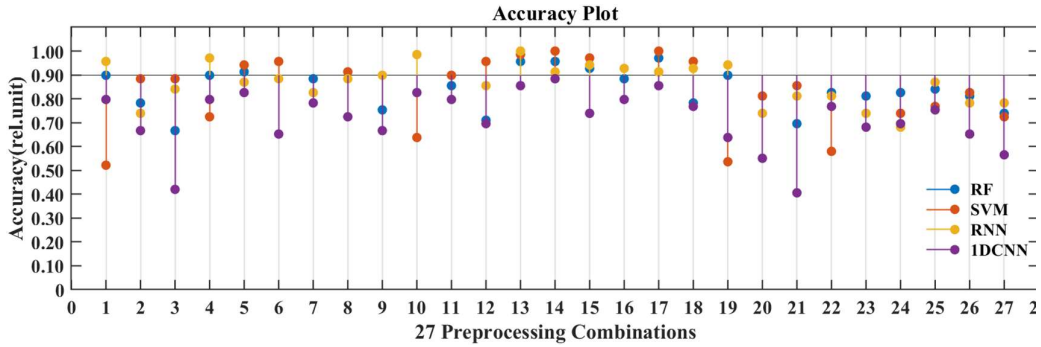
(2) Following SG+airPLS+VSN preprocessing, the SVM classification accuracy is only 0.52. However, after applying Gaussian+CWT+MSC and Gaussian+No pro-cessing+MSC preprocessing, the classification accuracy improves to 1. These results demonstrate the importance of selecting appropriate preprocessing combinations for optimal model accuracy.

(3) The overall accuracy and F1-score of RF, SVM, and LSTM are all greater than 0.9 when using the preprocessing combinations Gaussian+CWT+VSN, Gaussi-an+CWT+MSC, and Gaussian+CWT+No processing. These results indicate that Gauss-ian+CWT is a practical improvement for the model performance of the preprocessing combination.

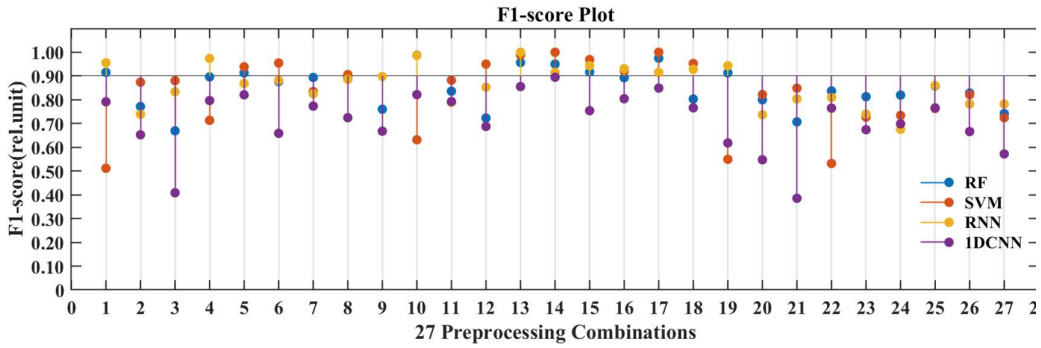
Table S2. Accuracy and F1-score for 108 models. The results of best model are shown in red.

Processing group	RF		SVM		RNN		1DCNN	
	Accuracy	F1-score	Accuracy	F1-score	Accuracy	F1-score	Accuracy	F1-score
1	0.8986	0.9155	0.5217	0.5117	0.9565	0.9556	0.7971	0.7917
2	0.7826	0.7718	0.8841	0.8739	0.7391	0.7389	0.6667	0.6527
3	0.6667	0.6693	0.8841	0.8807	0.8406	0.8339	0.4203	0.4089
4	0.8986	0.8965	0.7246	0.7135	0.9710	0.9740	0.7971	0.7968
5	0.9130	0.9129	0.9420	0.9391	0.8696	0.8688	0.8261	0.8209
6	0.8841	0.8769	0.9565	0.9547	0.8841	0.8814	0.6522	0.6590
7	0.8841	0.8940	0.8261	0.8346	0.8261	0.8254	0.7826	0.7734
8	0.8841	0.8952	0.9130	0.9061	0.8841	0.8856	0.7246	0.7247
9	0.7536	0.7602	0.8986	0.8980	0.8986	0.8980	0.6667	0.6678
10	0.9855	0.9875	0.6377	0.6318	0.9855	0.9866	0.8261	0.8219
11	0.8551	0.8361	0.8986	0.8824	0.7971	0.7889	0.7971	0.7926
12	0.7101	0.7230	0.9565	0.9502	0.8551	0.8528	0.6957	0.6878
13	0.9565	0.9565	0.9855	0.9880	1	1	0.8551	0.8556
14	0.9565	0.9502	1	1	0.9130	0.9146	0.8841	0.8949

15	0.9275	0.9170	0.9710	0.9692	0.9420	0.9433	0.7391	0.7542
16	0.8841	0.8928	0.9275	0.9220	0.9275	0.9306	0.7971	0.8045
17	0.9710	0.9748	1	1	0.9130	0.9150	0.8551	0.8492
18	0.7826	0.8032	0.9565	0.9527	0.9275	0.9283	0.7681	0.7661
19	0.8986	0.9134	0.5362	0.5496	0.9420	0.9434	0.6377	0.6184
20	0.8116	0.8002	0.8116	0.8219	0.7391	0.7371	0.5507	0.5475
21	0.6957	0.7071	0.8551	0.8486	0.8116	0.8039	0.4058	0.3854
22	0.8261	0.8372	0.5797	0.5321	0.8116	0.8096	0.7681	0.7649
23	0.8116	0.8130	0.7391	0.7271	0.7391	0.7396	0.6812	0.6741
24	0.8261	0.8198	0.7391	0.7344	0.6812	0.6759	0.6957	0.6989
25	0.8406	0.8582	0.7681	0.7633	0.8696	0.8602	0.7536	0.7658
26	0.8116	0.8283	0.8261	0.8216	0.7826	0.7824	0.6522	0.6663
27	0.7391	0.7420	0.7246	0.7247	0.7826	0.7819	0.5652	0.5721



(a)



(b)

Figure S1. Performance of models constructed based on 27 preprocessing combinations: (a) accuracy of all models; (b) F1-score of all models. The X-axis 1-27 represents 27 preprocessing combinations

3. The structure of 1DCNN

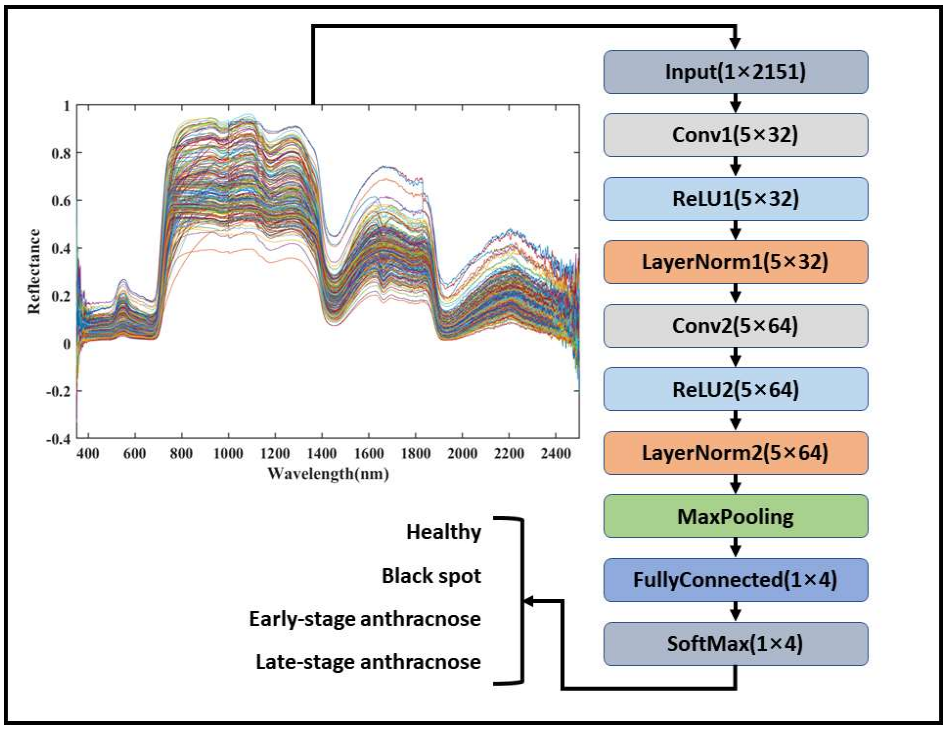


Figure S2. Structure of 1DCNN: the reflectance of all samples is used as the input of 1DCNN. Through convolution, pooling and full connection, four types of responses are finally obtained to achieve four-category classification.

3. Sensitivity, the ratio of th e average spectral reflectance of diseased samples to the average spectral reflectance of healthy samples in the same waveband.

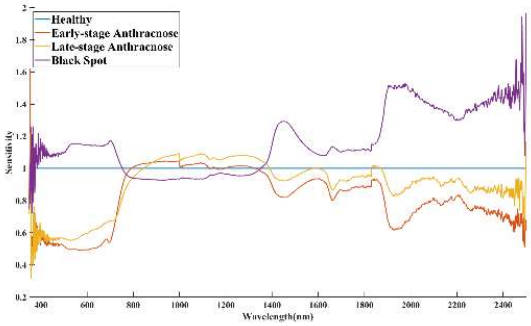
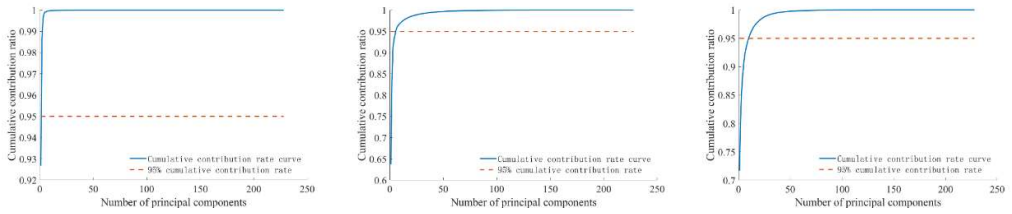


Figure S3. shows the samples' average reflectance and sensitivity curves in different physiological states.

Figure S3. In the near infrared (NIR) 1374-2400 nm and visible bands, anthracnose and black spot sensitivities were easily distinguishable. The sensitivity range for early-stage anthracnose was 0.49 to 0.96; for late-stage anthracnose, it was 0.53 to 1.17; and for black spot, it was 0.98 to 1.96. The reflectance spectra of anthracnose and black spot showed significant differences.

4. The interpretation bands of PC1 and PC2 through PCA

PC1 (71.69%) and PC2 (9.13%) explained 80.82% of the data variation, including regions 712-716 nm, 725-744 nm, and 762-776 nm.



(a) **(b)** **(c)**

Figure S4. Shows the cumulative principal component contributions

Table S3. Interpreted bands of PC1 and PC2 after PCA treatment

Data	Principal Component	Contribution Rate	Wavelength(nm)
VIS	PC1	92.68%	725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744
	PC2	5.88%	705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 757 758 759 760
NIR	PC1	63.75%	761 762 1653 1654 1655 1656 1657 1658 1659 1660 1661 1662 1663 1664 1665 1666 1667 1668 1669 1670
	PC2	19.80%	761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780
ALL	PC1	71.69%	725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744
	PC2	9.13%	712 713 714 715 716 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776