Application of machine learning methods for the differentiation of fungal diseases in strawberry based on hyperspectral image analysis

Anna Cheshkova*

Siberian Federal Scientific Center of AgroBioTechnology of the Russian Academy of Sciences, Krasnoobsk, Novosibirsk region, 630501, Russia

Abstract. Fungal diseases have a significant negative impact on strawberry yield. Their detection and differentiation using hyperspectral measurements is a possible alternative to traditional methods. In this study, strawberry leaves infected with *Ramularia Tulasnei, Marssonina potentillae* and *Dendrophoma obscurans* with visible symptoms of the disease were used for hyperspectral analysis. The reflection spectrum of leaves was recorded with a Photonfocus hyperspectral camera (wavelength range 475–900 nm, 149 channels) under laboratory conditions using the line scanning method. This research has aimed to compare four machine learning methods: spectral angle mapper (SAM), support vector machine (SVM), k-nearest neighbors (KNN) and linear discriminant analysis (LDA). Classification models were built based on the full spectrum, as well as on 12 vegetation indices (VI) as spectral features. The results demonstrated that the SVM model based on full spectra reached highest classification accuracy 94%. The KNN model performed slightly worse with 91% accuracy. The performance of models based on VIs was lower than that of models based on full spectra with an accuracy range of 78–85%.

1 Introduction

Fungal diseases cause large economic damage to strawberry production**.** For the targeted application of plant protection measures in horticulture, early detection and identification of diseases is essential. Technical advances in the production of digital hyperspectral cameras contributed to the development of plant diseases diagnosing methods based on the analysis of their optical properties in various regions of the electromagnetic spectrum [1, 2]. An analysis of the reflectance spectrum of plants makes it possible to discover the disease, assess its severity, differentiate types of pathogens, and detect the biotic stresses at early stages, including during the incubation period [3].

Currently, for the hyperspectral data exploration, the machine learning methods, neural networks, cluster and discriminant analysis are successfully used. [4].

Thus, in the study of T. Rumpf et al. [5], the support vector machine and vegetation indices were applied for the early detection and differentiation of sugar beet diseases based

* Corresponding author: cheshanna@yandex.ru

on the analysis of hyperspectral images. A classification accuracy of up to 97% has been achieved in distinguishing between healthy and diseased sugar beet leaves. The multiple classification accuracy between healthy leaves and leaves with symptoms of the three diseases (cercosporosis, leaf rust, and powdery mildew) exceeded 86%. Besides, the possibility of presymptomatic detection of plant diseases was demonstrated.

In study of Jiang Q. et al. [6] six classification models for identification of gray mold and anthracnose in strawberries were compared. Most classification models have demonstrated excellent accuracy (100%), recognizing classes of fungus infections even during the incubation period.

The objectives of our study were: (a) to explore the feasibility of hyperspectral imaging for differentiation between healthy strawberry leaves and leaves innoculated with pathogens *Ramularia Tulasnei* Sacc, *Marssonina potentillae* Desm. and *Dendrophoma obscurans* Anders; (b) to compare accuracy of four models (SAM, SVM, KNN, LDA) for classifying hyperspectral images.

2 Materials and methods

Strawberry leaves used in the experiment were taken from plants grown in the field of Siberian Federal Scientific Center of AgroBioTechnology (Russia). In July 2022, 20 healthy plants were selected, as well as 20 plants each infected with *Ramularia Tulasnei, Marssonina potentillae* and *Dendrophoma obscurans* with visible symptoms of the disease.

Identification of the disease was carried out through visual expertise [7, 8]. Leaves affected by *Ramularia Tulasnei* Sacc. had purple or reddish-brown round spots with a white centre (fig. 1a). Leaves affected by *Marssonina potentillae* (Desm.) were set with small dark purple rounded spots without a border (fig. 1b). *Dendrophoma obscurans* was characterized by the appearance of light brown spots along the edge of the leaves (fig. 1c).

(a) (b) (c)

Fig. 1. Strawberry leaves infected with: (a) *Ramularia Tulasnei,* (b) *Marssonina potentillae,* (c) *Dendrophoma obscurans.*

Spectral reflectance was measured using a Photonfocus MV1-D2048x1088-HS05-96-G2-10 hyperspectral camera, with an IMEC CMV2K-LS150-VNIR sensor (wavelength range 470-900 nm, 149 channels, a spatial resolution of 2048x1088 pixels). Line scanning was performed using a movable platform controlled by a computer.

According to the results of scanning 80 strawberry leaves, three-dimensional data arrays (hypercubes) were formed, containing two spatial data dimensions and one spectral data dimension.

Image pre-processing included reflection calibration with dark and white reference spectra. The dark spectrum was recorder by covering the camera lens with an opaque cap. The white spectrum was obtained using a standard white surface. Then, the original hyperspectral image was calibrated using the formula:

$$R = \frac{I_o - I_D}{I_W - I_D},$$

where R is the calibrated image, I_o is the original image, I_D, I_W are the dark and white reference images.

The data hypercube was formed and the images were calibrated using a program specially developed in Python.

The next step was to select regions of interest (ROI) and extract the data. These operations were performed in the ENVI 5.2 program (ITT Visual Information Solutions). First of all, binary "masks" were built and applied to remove the background of images. Masks were built based on the threshold values of the NDVI index, followed by aggregation of the inner regions. Further, on each strawberry leaf, ROIs were identified corresponding to healthy leaf tissue, central and marginal areas for *Ramularia Tulasnei*, color spots for *Marssonina potentillae* and *Dendrophoma obscurans* (fig.2). Pixels were randomly selected from each ROI. As a result, a training set of 8500 spectrum values was formed (1700 for each type of region). The test set was created in a similar way.

Fig. 2. ROIs corresponding to: 1 healthy, 2 *Ramularia* (center), 3 *Ramularia* (edge), 4 *Marssonina,* 5 *Dendrophoma.*

Spectral data were smoothed using a second-order Savitzky-Golay filter.

The following machine learning methods were used for classification [9].

Support Vector Machine (SVM) is a supervised binary classification algorithm that utilizes a linear division of the feature space using a hyperplane. The radial Gaussian basis function was taken as the classifier kernel.

K-Nearest Neighbors (KNN) is the simplest metric classifier based on object similarity assessment. The object being classified is assigned to the class to which the nearest objects of the training set belong. The Euclidean distance was used as a metric.

Spectral Angle Mapper (SAM) is a classifier in which the objects similarity is evaluated by the value of the angle between the reflection spectrum of the object and the reference spectrum. The smaller angle is, the more similar the two spectra are. Hence, the object will be assigned to the class that has the smallest angle value with some reference spectrum.

Linear Discriminant Analysis (LDA) is a machine learning technique used to find linear combinations of features that best separate two or more classes of objects. The found combination can be used as a linear classifier.

Due to the large dimensionality and multicollinearity of the original hyperspectral data, before applying machine learning methods, dimensionality reduction is often carried out by extracting features. In our study, the values of 12 vegetation indices (VI) presented in Table 1 were used as features. These indices are algebraic combinations calculated from reflectance values for selected wavelengths. VIs characterize the content of chlorophyll, pigments, water, nitrogen and carbon in plant tissues.

Table 1. Vegetation indices utilized in this study.

Vegetation Index	Equation
NDVI	$(R_{800}-R_{670}) / (R_{800}+R_{670})$
NRI	$(R_{570}-R_{670}) / (R_{570}+R_{670})$
RENDVI	$(R_{750}-R_{705}) / (R_{750}+R_{705})$
PSRI	$(R_{680}-R_{500}) / R_{750}$
PSSRa	R_{800} / R_{680}
PSSRb	R_{800} / R_{635}
PRI	$(R_{531}-R_{570}) / (R_{531}+R_{570})$
CRI	$1/R_{510} - 1/R_{550}$
ARI	$1/R_{550} - 1/R_{700}$
MCARI	$[(R_{700}-R_{670})-0.2\cdot(R_{700}-R_{550})](R_{700} / R_{670})$
TVI	$0.5\cdot[120\cdot(R_{750}-R_{550}) -200\cdot(R_{670}-R_{550})]$
PhRI	$(R_{550}-R_{531}) / (R_{550}+R_{531})$

All computations were made in the R program using the *kernlab, hsdar, class, MASS*, and *caret* packages.

3 Results and discussion

The averaged spectra of healthy and infected areas of strawberry leaves are shown in Figure 3. The reflection spectrum differs for different types of disease. White spots (central areas affected by *Ramularia Tulasnei*) demonstrate a relatively higher reflectivity in the visible wavelength region (400-700 nm) and lower in the near infrared region (700-900 nm) compared to other areas of the leaves. Healthy green areas show a high reflectivity in the near infrared region and exhibited a reflection peak near 550 nm. Leaf areas affected by *Dendrophoma obscurans* (brown spots) have a reflection peak at 630 nm. An insignificant difference in the spectra was found in areas affected by *Marssonina potentillae* and *Ramularia Tulasnei* (red and dark purple spots).

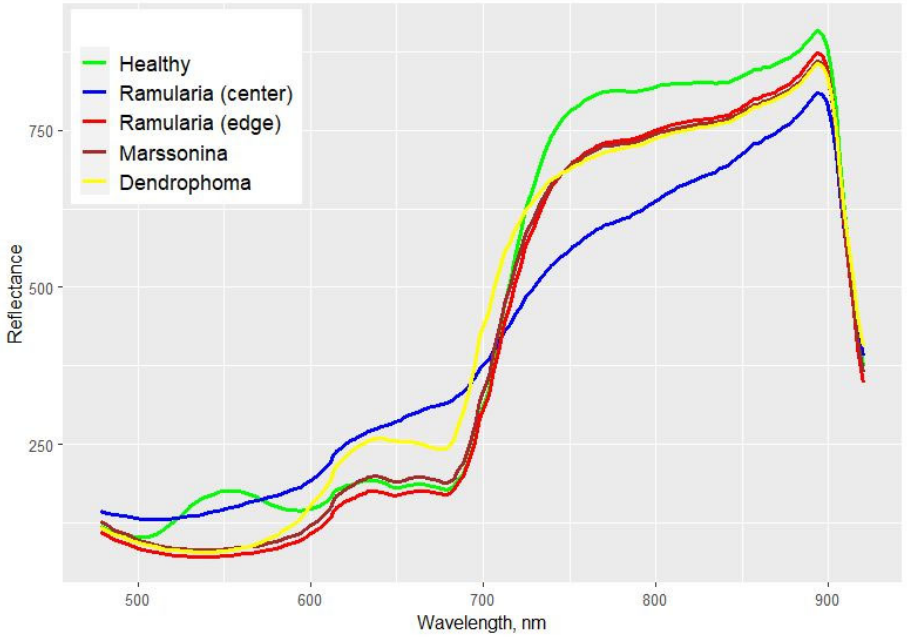

Fig. 3. Spectral curves of infected and healthy areas of strawberry leaves.

Four classification methods have been used to differentiate diseases. First, the SAM, SVM and KNN models were built for the full spectral range of 149 wavelengths. The LDA method is not suitable for use on the full spectrum as the data has a high degree of collinearity. Then, to reduce the dimensionality of the original data and speed up calculations, 12 vegetation indices were calculated. They were used as features in the LDA_{VI}, SVM_{VI}, KNN_{VI} models. Using the SVM method, a model was additionally built based on the spectrum values for three selected wavelengths (R_{670}, R_{550}, R_{480}) corresponding to the *RGB* range.

The confusion matrix of each analysis method was built (Table 2) and the overall classification accuracy was calculated.

The main errors in classification occur when differentiating *Ramularia Tulasnei* (edge) and *Marssonina potentillae*, since these areas have a similar reflection spectrum.

The SVM method based on full spectra achieved the highest classification accuracy (94%). Slightly lower (91%) was the accuracy of the KNN method. When vegetation indices were used as features, the classification accuracy decreased to 85% and 81%, respectively. The SAM and LDA methods provided poor performance compared with other classifiers. SVM in RGB space showed an accuracy of only 70%, which indicates the advantages of using hyperspectral images compared to RGB images.

Table 2. Confusion matrices of hyperspectral imaging analysis on strawberry diseases assessment.

Classified	**Reference**					**Overall accuracy**
	Healthy	**Ramularia (center)**	**Ramularia (edge)**	**Marssonina**	**Dendrophoma**	
SAM						73%
Healthy	89%	0%	1%	1%	0%	-
Ramularia (center)	0%	94%	2%	8%	15%	-
Ramularia (edge)	11%	1%	68%	33%	3%	-
Marssonina	0%	3%	25%	52%	20%	-
Dendrophoma	0%	2%	4%	6%	62%	-
SVM						94%
Healthy	100%	0%	0%	0%	0%	-
Ramularia (center)	0%	98%	1%	1%	1%	-
Ramularia (edge)	0%	1%	88%	10%	1%	-
Marssonina	0%	0%	9%	87%	2%	-
Dendrophoma	0%	1%	2%	2%	96%	-
KNN						91%
Healthy	100%	0%	0%	0%	0%	-
Ramularia (center)	0%	96%	1%	2%	3%	-
Ramularia (edge)	0%	1%	82%	13%	1%	-
Marssonina	0%	2%	13%	82%	3%	-
Dendrophoma	0%	1%	4%	3%	93%	-
SVM_{VI}						85%
Healthy	100%	0%	0%	0%	0%	-
Ramularia (center)	0%	96%	2%	2%	4%	-
Ramularia (edge)	0%	1%	63%	15%	2%	-

Marssonina	0%	2%	27%	77%	8%	-
Dendrophoma	0%	1%	8%	6%	86%	-
KNN_{VI}						81%
Healthy	100%	0%	1%	0%	0%	-
Ramularia (center)	0%	95%	2%	3%	7%	-
Ramularia (edge)	0%	1%	62%	23%	4%	-
Marssonina	0%	3%	26%	67%	10%	-
Dendrophoma	0%	1%	9%	7%	79%	-
LDA_{VI}						78%
Healthy	99%	1%	1%	1%	0%	-
Ramularia (center)	0%	93%	2%	2%	13%	-
Ramularia (edge)	1%	1%	58%	24%	3%	-
Marssonina	0%	4%	31%	67%	13%	-
Dendrophoma	0%	1%	8%	6%	71%	-
SVM_{RGB}						70%
Healthy	98%	2%	1%	1%	0%	-
Ramularia (center)	1%	79%	1%	2%	9%	-
Ramularia (edge)	1%	1%	78%	49%	15%	-
Marssonina	0%	4%	14%	40%	19%	-
Dendrophoma	0%	14%	6%	8%	57%	-

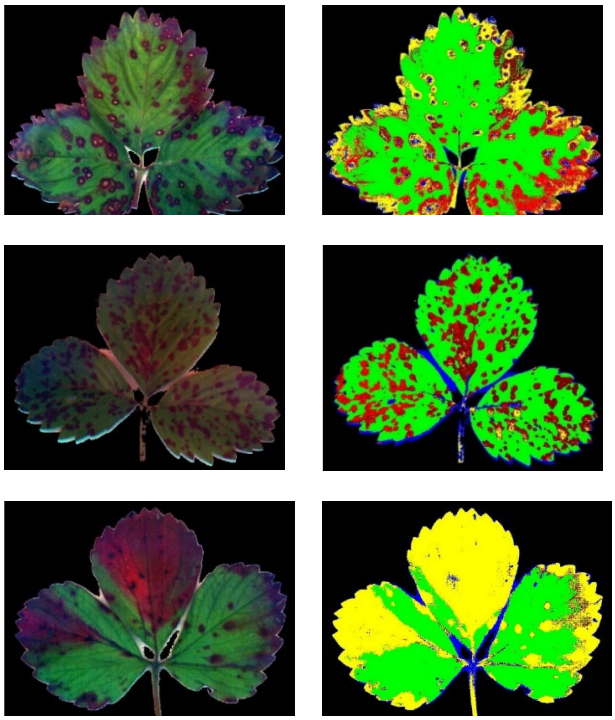

Fig. 4. Visualization of classification results by Support Vector Machines: green – healthy; blue - *Ramularia* (center); red - *Ramularia* (edge); brown – *Marssonina;* yellow - *Dendrophoma*

4 Conclusions

This study investigated the possibility of using hyperspectral imaging technique combined with machine learning methods to detect and identify fungal diseases on strawberry leaves caused by *Ramularia Tulasnei, Marssonina potentillae* and *Dendrophoma obscurans*. The reflection spectrum differed for different types of disease. In order to identify the strawberry leaves disease effectively, various classifiers (SAM, SVM, KNN, LDA) were applied using the full spectrum and the 12 vegetation indexes as features.

The SVM and KNN methods on full spectra achieved a satisfying accuracy 94% and 91% respectively. Compared with the classifiers developed with full spectrum, the overall accuracy in models based on vegetation indexes decreased to 85% and 81%, respectively. However, the use of vegetation indices as features can significantly speed up the classification to meet actual real-time disease detection needs.

References

1. A. F. Cheshkova, Vavilov J. Gen. Breed. **26(2)**, 202-213 (2022). https://doi.org/10.18699/VJGB-22-25
2. N. Zhang, G. Yang, Y. Pan, X. Yang, L. Chen, C. Zhao, Remote Sens. **12**, 3188 (2020). https://doi.org/10.3390/rs12193188
3. A. K. Mahlein, M. T. Kuska, J. Behmann, G. Polder, A. Walter, Ann. Rev. Phytopath **56**, 535–558 (2018). https://doi.org/10.1146/annurev-phyto-080417-050100
4. A. Lowe, N. Harrison, A.P. French, Plant Meth. **13**, 80-91 (2017). https://doi.org/10.1186/s13007-017-0233-z
5. T. Rumpf, A. K. Mahlein, U. Steiner, E. C. Oerke, H. W. Dehne, L. Plumer, Comp. Electr. Agric. **74**, 91-99 (2010). https://doi.org/10.1016/j.compag.2010.06.009
6. Q. Jiang, G. Wu, C. Tian, N. Li, H. Yang, Y. Bai, B. Zhang, Infr. Ph. Tech. **118** 103898 (2021). https://doi.org/10.1016/j.infrared.2021.103898
7. G. F. Govorova, D. N. Govorov, *Fungal diseases of garden strawberries, selection for immunity and other methods of protection* (MSHA, Moscow, 2015)
8. C. Garrido, M. Carbu, J. F. Fernandez-Acero, V. E. Gonzalez-Rodriguez, J. M. Cantoral, *New Insights in the Study of Strawberry Fungal Pathogens*. In: Husaini AM & Mercado JA (Eds). Genomics, Transgenics, Molecular Breeding and Biotechnology of Strawberry. Global Science Books, UK, 24-39 (2011)
9. A. Singh, B. Ganapathysubramanian, A. K. Singh, S. Sarkar, Trends Plant Sci. **21(2)** 110-124 (2016). http://dx.doi.org/10.1016/j.tplants.2015.10.015