Original Article
Food Science and Technology

ISSN 0101-2061 (Print)
ISSN 1678-457X (Online)

DOI: https://doi.org/10.1590/fst.100122

Hyperspectral image-based measurement of total flavonoid content of leaf-use Ginkgo biloba L.

Xiaoyang XING[1], Maocheng ZHAO[1]*, Xiwei WANG[1], Yuweiyi TANG[1]

Abstract

Ginkgo biloba leaf is a kind of food raw material with high concentration of total flavonoid content (TFC). Since conventional measurements of TFC are time-consuming and destructive, the approach of hyperspectral imaging-based prediction was investigated in this study. Hyperspectral images were collected in visual near-infrared (VIS-NIR) and short-wave near-infrared (SW-NIR). Initial comparative analysis on the full spans of the two wavelengths using partial least squares regression (PLSR) and support vector regression (SVR) showed that the TFC of the leaves could be barely captured only when PLSR working on short-wave infrared, with R_p^2 and $RMSE_p$ on testing set being respectively 0.5496 and 0.6384 mg/g. To further improve prediction performance meeting industrial requirements, another comparative study on feature selection was conducted to fine-tune the PLSR on SW-NIR using a genetic algorithm (GA) and successive projections algorithm (SPA). Results showed that GA-PLSR with 50 characteristic wavelengths mostly from the ranges of 1100-1200 nm and 1400-1500 nm had a significant improvement in performance, giving 0.8482 and 0.2967 mg/g for R_p^2 and $RMSE_p$, respectively. Therefore, an approach, GA-PLSR modeling on SW-NIR hyperspectral images, was established for the ginkgo leaves industry to rapidly and non-invasively predict the leaf concentration from the total flavonoid of ginkgo biloba seedling leaf.

Keywords: hyperspectral imaging; leaf-use ginkgo; ginkgo biloba leaves; total flavonoid content; characteristic wavelength.

Practical Application: This manuscript is relevant for the ginkgo biloba leaf production industry, especially for the online testing of the determination of leaves. This study provides online testing data on leaf-use ginkgo biloba leaves and establishes an online test method of leaf quality. The method provides technical support and a theoretical basis for precision production, management, and cultivation of leaf-use ginkgo.

1 Introduction

Gingko is the only gymnosperms in the ginkgo family, and its history can be traced back to the Carboniferous period (Ye et al., 2022), which has been widely cultivated in the world for its medical, ecological, and ornamental functions (Yang et al., 2021a, b). The leaf of ginkgo is a kind of raw food material and common herb, and flavonoids are the important content in it, which has the functions of protecting cardiovascular and cerebrovascular diseases, anti-aging, and so on (Liang et al., 2020, 2022; Wang et al., 2022a; Zhang et al., 2022a). Even though there are a lot of ginkgo trees in the world, the TFC in ginkgo biloba leaves decreases sharply with the increase of tree age, which results in that normal ginkgo leaves can hardly be used as herb ingredients. For high-quality leaves for pharmaceutical purposes, the leaf of specially cultivated ginkgo biloba seedlings under 5 years in eastern China is harvested from nurseries, with high TFC quantifying its commercial values, which promotes the development of ginkgo biloba leaf-related industries (Zou et al., 2019; Guo et al., 2020). However, during the planting and purchasing process, an obstacle exists since there is no effective means of testing leaf quality. To tackle this issue, a rapid, efficient, and non-invasive testing method is urgently needed to improve the quality and yield of ginkgo leaves.

Due to the important value of ginkgo leaves, a large number of studies on the content of flavonoids in ginkgo leaves have been carried out (Zhao et al., 2018). Nowadays, many effective detection methods have been developed. Spectrophotometer, high-performance liquid chromatography, and near-infrared spectrometer are conventional testing methods for predicting TFC in Ginkgo biloba L., which have been verified and implemented in the laboratory (Wu et al., 2021; Wang et al., 2022b; Zhang et al., 2022b). Even if those methods have high accuracy to meet the detection requirements, in addition to the near-infrared spectrometer, other methods need to be preprocessed before they can be detected. This process is not only time-consuming and not cost-effective but can cause damage to objects as well. The near-infrared spectrometer fails to meet the needs of online rapid detection for pipelines since it is spectrally sampled in local areas with a high cost and insufficient information. Therefore, the traditional detection methods are difficult to meet the needs of rapid non-destructive testing of ginkgo leaves.

Research on the TFC in Ginkgo biloba leaves by hyperspectral imaging technology is very limited. Thus, the Ginkgo biloba industry lack basic guidance. Compared with the spectrometer, hyperspectral imaging technology is capable of obtaining the

Received 26 Sept., 2022
Accepted 18 Nov., 2022
[1] *College of Mechanical and Electronic Engineering, Nanjing Forestry University, Nanjing, China*
*Corresponding author: mczhao@njfu.edu.cn

spectral and spatial information of the sample at the same time (Osco et al., 2020; Frey et al., 2020; Tang et al., 2020; Malmir et al., 2020). It is a detection method for analyzing composition inside and outside the sample by measuring the spectral signal. With the continuous development of chemometrics and machine learning technology, several new methods were applied to build prediction models for the composition of samples, and the ability of hyperspectral imaging technology has improved for better performance (Xin et al., 2020; Koirala et al., 2020; Ma et al., 2019). Nowadays, hyperspectral imaging camera has been extended to many fields, numerous research has been carried out and proved their feasibility (Ma et al., 2019; Deng et al., 2020; Qu et al., 2020; Kang et al., 2020). It has been reported that a variety of components in foods and plants were detected by hyperspectral imaging technology. A method using modified Adaboost RT based on an AOTF hyperspectral imaging system was employed to test the Brix and pH values and a method using PLSR based on a short-wave near-infrared hyperspectral imaging system was used to test the TVB-N value of pork (Wang et al., 2020). Thus, it seems to be feasible that hyperspectral imaging technology could be used to test the quality of leaves of leaf-use ginkgo.

This study aimed to apply hyperspectral imaging technology to test the quality of leaves of leaf-use ginkgo. Before image capturing, a new hyperspectral imaging system, including VIS-NIR and SW-NIR bands, was built and tested. Prediction models were established on the data of images captured by the hyperspectral imaging system, and TFC values were obtained by high-performance liquid chromatography (HPLC) to explore the method of testing TFC in Ginkgo biloba leaves.

2 Materials and methods

2.1 Hyperspectral imaging system

A non-destructive hyperspectral imaging system for predicting TFC of Ginkgo biloba leaves was set up (Figure 1), comprising of a push-broom VIS-NIR hyperspectral camera (GaiaField-V10E-AZ4, Jiangsu Dualix Spectral Image Technology Co. Ltd, China), a push-broom SW-NIR hyperspectral camera (GaiaField-N17E-HR, Jiangsu Dualix Spectral Image Technology Co. Ltd, China), two self-made dome light source systems, an uninterrupted power supply (UPS) (C3K, Shante, China), a transmission desk, a dark chamber and a computer (T570, Lenovo, China). Each camera was equipped with the same dome light source system, which includes 12 halogen lamps (Halogen 12V, Philips, China), with UPS providing a stable power supply. The whole system was surrounded by the dark chamber to prevent external light interference.

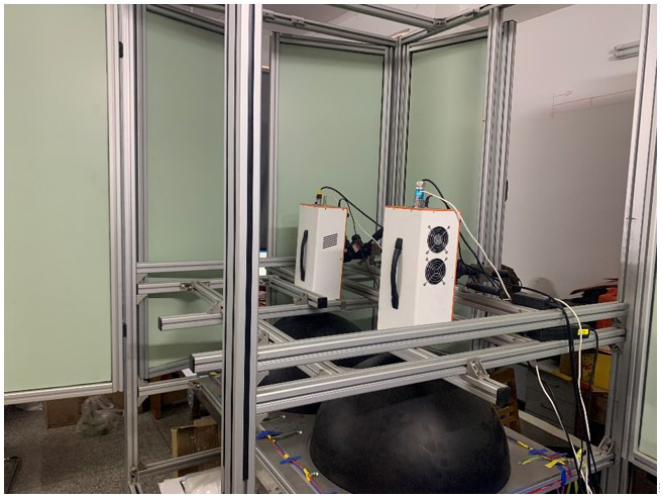

Figure 1. Hyperspectral imaging system.

2.2 Ginkgo biloba leaves samples

To obtain the leaves in the same state as the harvest time, the ginkgo biloba leaves used in this study were harvested in a Ginkgo biloba nursery in Pizhou City, Jiangsu Province in September, 2020. The leaves were transported to the nondestructive testing laboratory in a 4 °C incubator for image acquisition and total flavonoid content detection.

2.3 Determination of total flavonoid in leaves

After weighing the leaves, they were put in the oven until reaching constant leaf weight. The dried leaves were placed into a round bottom flask after crushing into powder, and 5 mL of 25% hydrochloric acid and 30 mL methanol were added to the same flask. After 1.5 hours of water bath refluxing, filtrating the residue, the filtrate was taken in the volumetric flask, and the methanol was added to 50 mL after achieving a constant temperature.

The standards used in this study were quercetin, kaempferol, and isorhamnetin. 5.7 mg of quercetin, 5.7 mg of kaempferol, and 2.1 mg of isorhamnetin were weighed and dissolved in 100 mL of methanol solution using a ten-thousand-point balance. 50 mL, 20 mL, 10 mL, 5 mL, 2 mL, and 1 mL were diluted to 50 mL, respectively, and then filtered through a membrane. Standard preparation of three parts was made to perform three repeated experiments to ensure the reliability of experimental data.

High-performance liquid chromatography (alliance e2695, Waters, USA) was used to determine the content of total flavonoids in standard and experimental samples. The following conditions were also employed: the column temperature, 30 °C; the detection wavelength, 254 nm; the mobile phase, methanol-0.1% formic acid solution (1:1); the flow rate, 1 mL/min; and the injection volume, 10 μL.

2.4 Dada acquisition and processing

Hyperspectral images acquisition

An illuminance meter was placed in the center of the camera field of view after opening the light source system to monitor the light source status, and the image acquisition was conducted after the numerical stability of the illuminance meter. Each sample was placed on the black rubber material platform, and the image was captured when the sample passed the camera line scanning area. Due to the diffuse reflection illumination of the dome light source system, leaves did no lose much water during the image acquisition process.

The VIS-NIR hyperspectral camera collected images at a full resolution of 669 × 800 pixels in the range of 336.20-1092.50 nm with an increment of 5.5 nm, producing a spectral cube with a total of 150 bands for each scan, and the SW-NIR hyperspectral camera collected images at full resolution of 550 × 640 pixels in the range of 874.00-1731.00 nm with an increment of 1.7 nm, producing a spectral cube with a total of 512 bands for each scan.

Spectral calibration

In addition to sample images, 7 kinds of standard reflectivity panels (Labspere, USA) and a dark image were captured, before and after samples image acquisition, including the panel with reflectivity of 2%, 5%, 10%, 20%, 50%, 75%, and 99%. Due to the unstable of the light source and external interference, it was necessary to conduct a calibration on the images, which converted hyperspectral gray value images to hyperspectral reflectivity images. A relative reflectivity image cube was calculated as follows (Equation 1):

$$I_c = \frac{I_{raw} - I_{dark}}{I_{calib} - I_{dark}} \times C \quad (1)$$

Where I_c is the result of the relative reflectivity image, I_{raw} is the raw gray value image, I_{dark} is the dark hyperspectral image, I_{calib} is the white hyperspectral image and C is the correction coefficient. In this study, dark and white images were not strictly identified respectively as images captured without removing the lens cover and images of 99% reflectivity standard panel, but with low reflectivity value standard panel images and high reflectivity value standard panel images, respectively. And C was different when making reflectivity calibration with different reflectivity standard panels.

Spectra extraction

Hyperspectral image, with a huge amount of data, is hard to be used to build a spectral prediction model directly. To reduce the amount of calculation, the common method is to extract the spectral curves of samples by converting three-dimensional data into two-dimensional data, which greatly reduces the calculation difficulty. With the continuous development of computer vision technology, it has been possible to obtain spectral information representing the region of interest of the complete sample in various ways. A simple threshold algorithm was used to obtain the mask of samples from images. The background of the image is a black rubber material platform, its spectral characteristics are very different from the sample. It is easy to get a mask of the sample by setting the threshold in particular wavebands, which can be used to conduct on all datasets in both VIS-NIR and SW-NIR wavebands automatically. The mean spectra data were extracted from the region of interest of hyperspectral images, by computing the mean value in the region of interest at each channel.

2.5 Establishment of TFC prediction model

Prediction model

Partial least square regression (PLSR) is a high-efficiency regression model in spectral analysis that combines the functions of multiple linear regression (MLR) and principal component regression (PCR), which can extract features among variables and analyze correlations among variables (Zhang et al., 2021; Chen et al., 2021). Support vector regression (SVR) is an important branch of support vector machine, that can transform a nonlinear problem into a linear problem (Tan et al., 2022; Lu et al., 2019). By minimizing support vector spacing, SVR can find the relationship between spectral information and target values. In this paper, these two methods were adopted to build regression models on both VIS-NIR and SW-NIR wavebands.

The performance of models was assessed by root mean square error (*RMSE*) of the training set ($RMSE_c$) and the test set ($RMSE_p$) and coefficient of determination (R^2) of the training set (R_c^2) and test set (R_p^2). In general, the closer the R^2 to 1 and the closer the *RMSE* to 0, indicating the stability and accuracy of the model (Wang et al., 2022c).

Optimal wavelength selection

There were 150 and 512 wavebands in VIS-NIR and SW-NIR hyperspectral cameras, respectively. Even though the number of images to curves was minimized to reduce the amount of data, there was still a lot of information that needed to be used in prediction models. Meanwhile, hyperspectral images are highly correlated between adjacent wavebands, which leads to collinearity and redundancy problems. Optimal wavelength selection algorithm needed to be used to solve the former problem, in order to shorten the time of building prediction models, reduce the dimension of spectral data and promote the performance of prediction models (Zou et al., 2022a, 2022b).

Successive projections algorithm (SPA) is a forward feature variable selection method, that can find the most representative characteristic wavebands by analyzing collinearity between bands (Ma et al., 2022). The genetic algorithm (GA) is a method of searching for an optimal solution by simulating the natural evolution process (Wan et al., 2021). This method established the corresponding fitness function after encoding each channel, and the characteristic wavebands were obtained by methods of selection, crossover, and mutation. SPA and GA were used in this study to select characteristic wavebands, which would be used to build light models for predicting TFC in Ginkgo biloba leaves.

3 Results and discussion

3.1 Reference TFC values

The reference TFC values of the 138 Ginkgo biloba leaf samples measured by HPLC were obtained and shown in Table 1. The TFC values range from 0.7204 mg·g^{-1} to 4.9748 mg·g^{-1}, which showed a similar value to Wang's data in the same city and demonstrated that it was valid to establish a TFC prediction model on this dataset (Zou et al., 2019). The rule of splitting datasets was to sort all TFC values from the smallest to the largest, taking one in four randomly into the test set and the remaining three putting into the training set. There were 104 samples in the training set and 34 samples in the test set, which satisfied the normal training and test split rate.

3.2 Hyperspectral data analysis

The spectral images are shown in Figure 2 with the pixel spectra at the sample region and background. Owing to the low reflectivity of the background platform, samples showed great differences from the background in the 90th and 50th channel, on VIS-NIR and SW-NIR wavebands, respectively, and the thresholds were set to 0.1 and 0.2, respectively. The masks of the region of interest on VIS-NIR and SW-NIR are shown in Figure 2c and Figure 2d, respectively.

By the mask generated, the spectra data were computed, and the bad channels were cut on SW-NIR wavebands, remaining the wavebands range from 1100- 1600 nm, the spectra curves are shown in Figure 3. It can be seen that the variation trend of the overall mean spectrum of the leaves was consistent, indicating

Table 1. Statistical results of TFC of samples.

Sample set	Number of samples	Minimum/mg·g^{-1}	Maximum/mg·g^{-1}	Mean/mg·g^{-1}	SD /mg·g^{-1}
Training set	104	0.7204	4.9748	2.4242	0.8143
Test set	34	1.0084	4.1766	2.4338	0.7728
All	138	0.7204	4.9748	2.4266	0.8015

Note: SD: Standard Deviation.

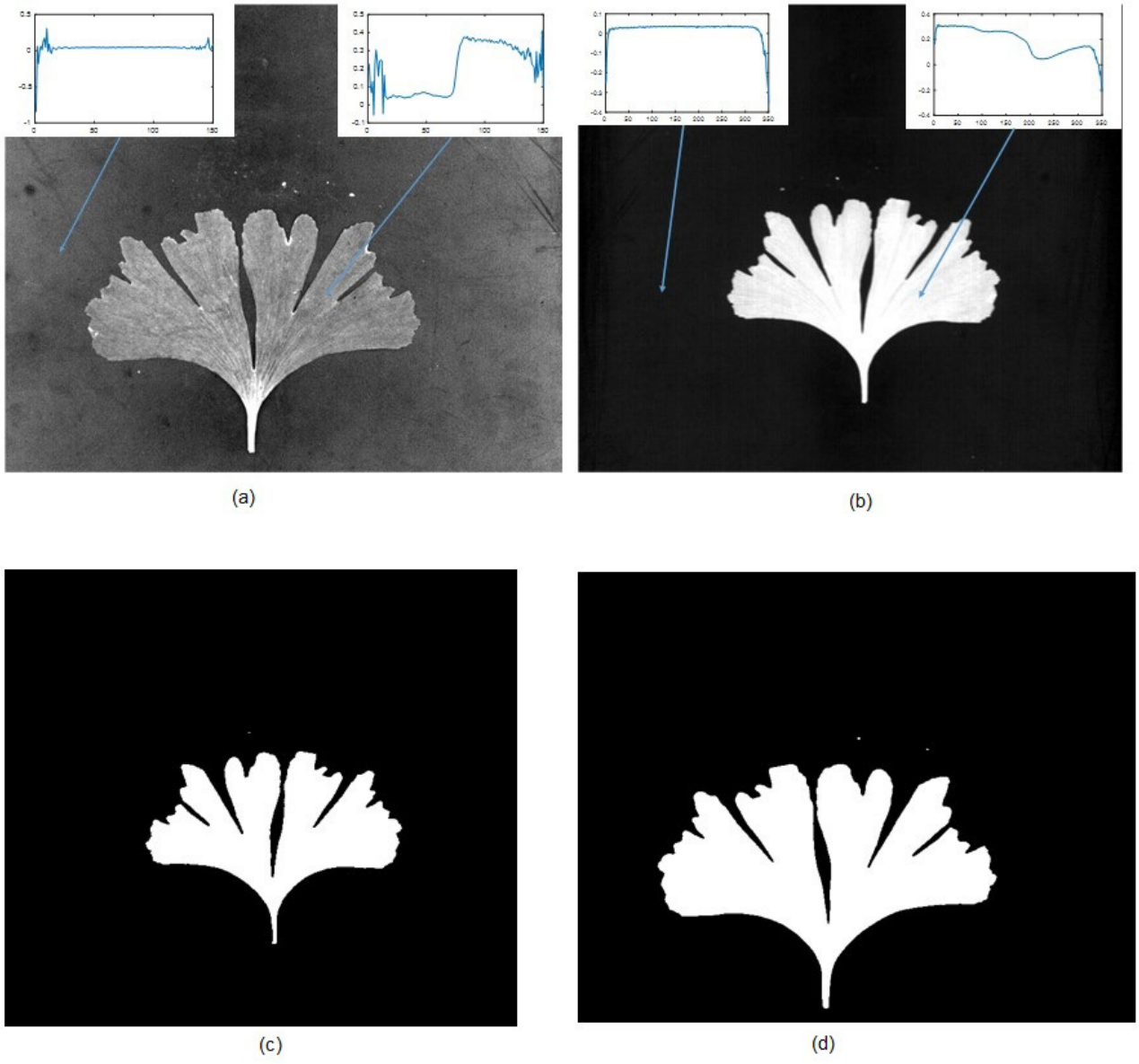

Figure 2. Spectral image and region of interest (ROI) image (a) Spectral image of the 90th band of VIS/NIR (b) Spectral image of the 50th band of NIR (c) ROI image of VIS/NIR (d) ROI image of NIR.

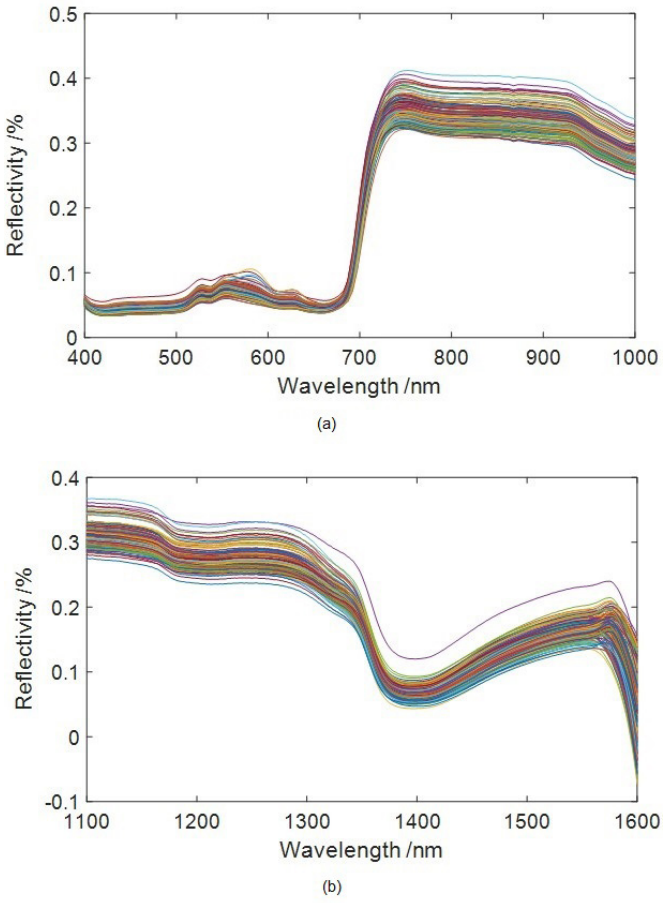

Figure 3. Mean spectral curve (a) Mean spectral curve of VIS/NIR (b) Mean spectral curve of SW-NIR.

that the collected spectral images had a good consistency. And the spectral mean of each sample showed some differences, indicating that there were differences between the collected samples, and the data exhibitedstrong generalization.

3.3 Establishment of the prediction model

Models on full spectrum

PLSR was implemented to find the relationship between spectral data and TFC values in Ginkgo biloba leaves, by inputting the whole spectral matrix and TFC values. To reduce the possibility of model over-fitting, the k-fold cross-validation method was used on the training set during the process of model building, and k value was set to 5. The PLSR prediction model with principal components from 1 to 50 was established, and the best model for different wavebands was selected by comparing the performance of different principal component models. Several models are shown in Table 2, whose principal components are around the principal components of the best model. On both VIS-NIR and SW-NIR bands, the performance of models on the training set increased with the increase of principal components, and the performance of models on the test set increased to the peak and then decreased. Model on VIS-NIR wavebands achieved the best performance with the principal

Table 2. Performance of all-band prediction model.

model	wavebands	R_c^2	$RMSE_c$ / mg·g^{-1}	R_p^2	$RMSE_p$ / mg·g^{-1}
PLSR_3	VIS-NIR	0.4212	0.6921	0.2914	0.9390
PLSR_4		0.4329	0.6771	0.3013	0.8758
PLSR_5		0.4719	0.6523	0.2869	0.8862
PLSR_6		0.4871	0.6277	0.2538	0.9413
PLSR_9	**SW-NIR**	0.4914	0.7058	0.5023	0.6637
PLSR_11		0.5511	0.6762	0.5292	0.6425
PLSR_13		**0.6180**	**0.6371**	**0.5496**	**0.6384**
PLSR_15		0.7201	0.5623	0.3857	0.6671
SVR	ViS-NIR	0.5322	0.5879	0.2312	0.9354
SVR	SW-NIR	0.5567	0.5063	0.3113	0.6836

components number of 4, it can be seen that the R_c^2 value and $RMSE_c$ value of the best PLSR model on VIS-NIR wavebands were 0.4329 and 0.6771 mg·g-1, respectively, and the R_p^2 value and $RMSE_p$ value were 0.3013 and 0.8758 mg·g-1, respectively. Model on SW-NIR wavebands exhibited the best performance with the principal components of 13, it can be seen that the R_c^2 value and $RMSE_c$ value of the best PLSR model on SW-NIR wavebands were 0.6180 and 0.6371 mg·g-1, respectively, and the R_p^2 value and $RMSE_p$ value were 0.5496 and 0.6384 mg·g-1, respectively. It can be seen on VIS-NIR wavebands that, when principal components reached 4 and had poor performance, then the model tended to be over-fitting. This phenomenon showed that the PLSR model could hardly predict the TFC of Ginkgo biloba leaves on VIS-NIR wavebands. Compared to models on VIS-NIR wavebands, models on SW-NIR wavebands showed better performance on both the training set and test set, which means that SW-NIR wavebands data were more suitable for building the prediction models by PLSR.

SVR used the same method to cross-validate on the training set as PLSR, and the best hyper-parameters were founded by grid search. In Table 2, the best model is presented on both SW-NIR and VIS-NIR wavebands, with the hyper-parameters founded by grid search. The R_c^2 value and $RMSE_c$ value of the SVR model on VIS-NIR wavebands were 0.5322 and 0.5879 mg·g-1, respectively, and the R_p^2 value and $RMSE_p$ value were 0.2312 and 0.9354 mg·g-1, respectively. The R_c^2 and $RMSE_c$ values of the SVR model on SW-NIR wavebands were 0.5567 and 0.5063 mg·g-1, respectively, as compared to the R_p^2 and $RMSE_p$ values being 0.3113 and 0.6836 mg·g-1, respectively. Comparing those two models, it could be found that the model built on VIS-NIR wavebands has lower performance. By comparing those models shown in the table, SVR models had very poor performance than PLSR ones in predicting TFC. Even though SVR was a powerful model, it could not be used to build models for predicting TFC.

In summary, spectral data of VIS-NIR wavebands showed less correlation with TFC in Ginkgo biloba leaves than SW-NIR wavebands, and PLSR models had better performance than SVR models. As a result, models built by PLSR based on spectral data of SW-NIR wavebands had the potential to predict TFC in Ginkgo biloba leaves.

Models on characteristic wavebands

According to the results of the models built before, the prediction model was optimized based on the PLSR model built on the spectral data of SW-NIR wavebands. In this study, SPA and GA were used to select the characteristic wavebands, and the target number of characteristic wavebands was set to 20, 30, 40, 50, and 60. The performance of the optimal models is shown in Table 3.

SPA is a method to eliminate redundant information between wavebands. By setting the target number, the characteristic wavebands were obtained. From the table, it can be shown that the 5 SPA-PLSR models with different characteristic wavebands number had similar performance, which showed a small increase in performance over the PLSR models. The R_c^2 and $RMSE_c$ values of the model with the best performance in the table were 0.6533 and 0.5857 mg•g-1, respectively, as compared to the R_p^2 and $RMSE_p$ values being 0.6286 and 0.6944 mg•g-1, respectively.

Compared with SPA, GA is a supervised learning algorithm, that selects the characteristic wavebands by minimizing loss function, and the iteration of optimizing was 2000. From the table, it can be found that, when the target number of characteristic wavebands was 20, the performance of GA-PLSR seemed to be the worst among the 5 GA-PLSR models, and with the increase of the target number, the performance had a significant improvement. It could be concluded that, a good model with high performance should have enough spectral data, 20 wavebands seemed insufficient to build a stable model for predicting TFC in ginkgo biloba leaves. When the number of characteristic wavebands was 50, the performance of GA-PLSR reached a peak, and the R_c^2 and $RMSE_c$ values of the model with the best performance in the table were 0.8532 and 0.5403 mg•g-1, respectively, while the R_p^2 value and $RMSE_p$ value were 0.8482 and 0.2967 mg•g-1, respectively. 50 characteristic wavebands selected by GA are shown in Figure 4.

Discussions

In this study, prediction models were built on full-wavebands and characteristic wavebands spectral data, the performance of these models is shown in Table 2 and Table 3. By comparing the result of performance, it could be found that it was difficult to build a stable and accurate model for predicting TFC in Ginkgo biloba leaves on the full wavebands data, the performance could hardly be employed in testing leaves. However, the full-wavebands models were not completely useless, which showed that SW-NIR wavebands had better performance on both PLSR and SVR than VIS-NIR. This phenomenon, was probably, due to the fact that, spectral data of VIS-NIR wavebands had poor correlation with flavonoids, and spectral data of SW-NIR wavebands had more information that is invisible and maybe had a stronger correlation with flavonoids.

The GA-PLSR models showed a better performance than SPA-PLSR and had a more significant improvement than single PLSR. GA is a kind of algorithm that simulates biogenetics, which scans the whole wavebands to find the characteristic bands to build a stable model. SPA is an algorithm aimed to find the most representative information in spectral data, without using the real TFC values of each sample. Naturally, the time to explore a stable GA-PLSR is much longer than SPA-PLSR. Fortunately, the time-consuming is meaningful, and the performance of GA-PLSR built on 50 characteristic wavebands is acceptable.

From the characteristic wavebands selected by the model, it could be found that the wavebands clustered in the range of 1100-1200 nm and 1400-1500 nm. Thus, it can be concluded that

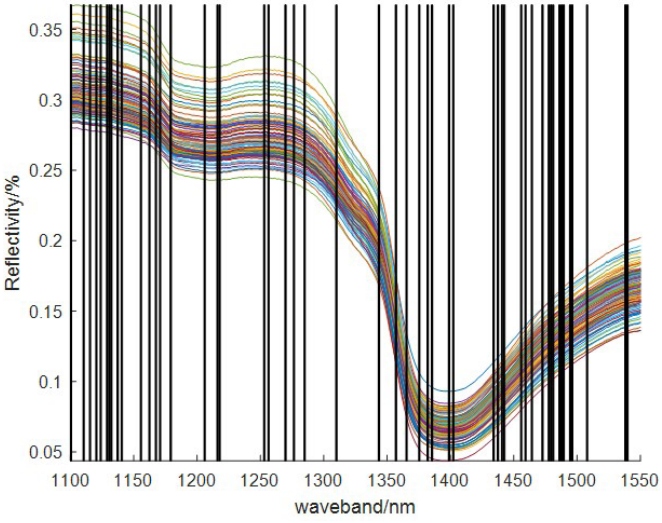

Figure 4. 50 characteristic wavebands selected by GA-PLSR.

Table 3. Performance of characteristic wavebands prediction model.

model	Numbers of characteristic wavebands	R_c^2	$RMSE_c$ /mg·g^{-1}	R_p^2	$RMSE_p$ /mg·g^{-1}
SPA-PLSR	20	0.6198	0.6470	0.5422	0.7031
	30	0.7341	0.5902	0.5759	0.6328
	40	0.6533	0.5857	0.6286	0.6944
	50	0.6953	0.5771	0.6111	0.6520
	60	0.7133	0.6013	0.5991	0.6298
GA-PLSR	20	0.8412	0.5539	0.6742	0.4397
	30	0.8319	0.6115	0.7649	0.3691
	40	0.8215	0.5833	0.7969	0.3431
	50	**0.8532**	**0.5403**	**0.8482**	**0.2967**
	60	0.8322	0.5507	0.7967	0.3378

these wavebands had a powerful correlation with the flavonoids in Ginkgo biloba leaves. In the future, a new multispectral camera system could be established by these wavebands to test the quality of Ginkgo biloba leaves at a faster speed, which will in turn improve industrial efficiency significantly.

4 Conclusion

This research demonstrated that it is feasible to estimate TFC values in Ginkgo biloba leaves using hyperspectral imaging technology. The following conclusions were made by the process of building prediction models: (1) SW-NIR wavebands data had more correlation with flavonoids in Ginkgo biloba leaves on both SVR and PLSR models than VIS-NIR wavebands, this may imply that the invisible SW-NIR wavebands had more information of flavonoids. (2) PLSR prediction models exhibited better performance than SVR on predicting TFC in ginkgo leaves, whose R_p^2 value and $RMSE_p$ value were 0.5496 and 0.6384 mg•g-1, respectively. (3) Models built on characteristic wavebands selected by GA and SPA were useful in improving the performance of the prediction model. The spectral data of SW-NIR wavebands seemed to be redundant, it was necessary to select wavebands before building prediction models. In addition, the GA-PLSR model with 50 characteristic wavebands had the best performance, with the R_p^2 and $RMSE_p$ values being 0.8482 and 0.2967 mg•g-1, respectively. This model showed a significantly higher performance than the other models. (4) The characteristic wavebands clustered at the range of 1100-1200nm and 1400-1500nm, those wavebands showed a high correlation with TFC in ginkgo leaves and could help promote the establishment of a multispectral testing system to test the quality of leaves at a fast speed meeting industrial needs.

As a result, a new method was established for rapid and non-invasive detection of leaves of leaf-use ginkgo, which provides technical support and a theoretical basis for precision production, management, and cultivation of leaf-use ginkgo.

References

Chen, S., Ma, L., Hu, T., Luo, L., He, Q., & Zhang, S. (2021). Nitrogen content diagnosis of apple trees canopies using hyperspectral reflectance combined with PLS variable extraction and extreme learning machine. *International Journal of Agricultural and Biological Engineering*, 14(3), 181-188. http://dx.doi.org/10.25165/j.ijabe.20211403.6157.

Deng, H., Konopka, C. J., Cross, T. W. L., Swanson, K. S., Dobrucki, L. W., & Smith, A. M. (2020). Multimodal nanocarrier probes reveal superior biodistribution quantification by isotopic analysis over fluorescence. *ACS Nano*, 14(1), 509-523. http://dx.doi.org/10.1021/acsnano.9b06504. PMid:31887006.

Frey, L. A., Baumann, P., Aasen, H., Studer, B., & Kölliker, R. (2020). A non-destructive method to quantify leaf starch content in red clover. *Frontiers in Plant Science*, 11, 569948. http://dx.doi.org/10.3389/fpls.2020.569948. PMid:33178239.

Guo, Y., Wang, T., Fu, F. F., El-Kassaby, Y. A., & Wang, G. (2020). Temporospatial flavonoids metabolism variation in Ginkgo biloba leaves. *Frontiers in Genetics*, 11, 589326. http://dx.doi.org/10.3389/fgene.2020.589326. PMid:33329734.

Kang, X., Duan, P., & Li, S. (2020). Hyperspectral image visualization with edge-preserving filtering and principal component analysis. *Information Fusion*, 57, 130-143. http://dx.doi.org/10.1016/j.inffus.2019.12.003.

Koirala, B., Zahiri, Z., & Scheunders, P. (2020). A machine learning framework for estimating leaf biochemical parameters from its spectral reflectance and transmission measurements. *IEEE Transactions on Geoscience and Remote Sensing*, 58(10), 7393-7405. http://dx.doi.org/10.1109/TGRS.2020.2982263.

Liang, C., Zhang, Z., Zhang, H., Ye, L., He, J., Ou, J., & Wu, Q. (2020). Ordered macroporous molecularly imprinted polymers prepared by a surface imprinting method and their applications to the direct extraction of flavonoids from Gingko leaves. *Food Chemistry*, 309, 125680. http://dx.doi.org/10.1016/j.foodchem.2019.125680. PMid:31670118.

Liang, H., Yuan, X., Sun, C., Sun, Y., Yang, M., Feng, S., Yao, J., Liu, Z., Zhang, G., & Li, F. (2022). Preparation of a new component group of Ginkgo biloba leaves and investigation of the antihypertensive effects in spontaneously hypertensive rats. *Biomedicine and Pharmacotherapy*, 149, 112805. http://dx.doi.org/10.1016/j.biopha.2022.112805. PMid:35276465.

Lu, B., Sun, J., Yang, N., & Hang, Y. (2019). Fluorescence hyperspectral image technique coupled with HSI method to predict solanine content of potatoes. *Journal of Food Processing and Preservation*, 43(11), e14198. http://dx.doi.org/10.1111/jfpp.14198.

Ma, H., Zhao, K., Jin, X., Ji, J., Qiu, Z., & Gao, S. (2019). Spectral difference analysis and identification of different maturity blueberry fruit based on hyperspectral imaging using spectral index. *International Journal of Agricultural and Biological Engineering*, 12(3), 134-140. http://dx.doi.org/10.25165/j.ijabe.20191203.4325.

Ma, X., Luo, H., Zhang, F., & Gao, F. (2022). Study on the influence of region of interest on the detection of total sugar content in apple using hyperspectral imaging technology. *Food Science and Technology (Campinas)*, 42, e87922. http://dx.doi.org/10.1590/fst.87922.

Malmir, M., Tahmasbian, I., Xu, Z., Farrar, M. B., & Bai, S. H. (2020). Prediction of macronutrients in plant leaves using chemometric analysis and wavelength selection. *Journal of Soils and Sediments*, 20(1), 249-259. http://dx.doi.org/10.1007/s11368-019-02418-z.

Osco, L. P., Ramos, A. P. M., Faita Pinheiro, M. M., Moriya, É. A. S., Imai, N. N., Estrabis, N., Ianczyk, F., Araújo, F. F., Liesenberg, V., Jorge, L. A. C., Li, J., Ma, L., Gonçalves, W. N., Marcato Junior, J., & Eduardo Creste, J. (2020). A machine learning framework to predict nutrient content in valencia-orange leaf hyperspectral measurements. *Remote Sensing*, 12(6), 906. http://dx.doi.org/10.3390/rs12060906.

Qu, D. X., Berry, J., Calta, N. P., Crumb, M. F., Guss, G., & Matthews, M. J. (2020). Temperature Measurement of Laser-Irradiated Metals Using Hyperspectral Imaging. *Physical Review Applied*, 14(1), 014031. http://dx.doi.org/10.1103/PhysRevApplied.14.014031.

Tan, F., Zhan, P., Zhang, Y., Yu, B., Tian, H., & Wang, P. (2022). Development stage prediction of flat peach by SVR model based on changes in characteristic taste attributes. *Food Science and Technology (Campinas)*, 42, e18022. http://dx.doi.org/10.1590/fst.18022.

Tang, Y., Cheng, Z., Miao, A., Zhuang, J., Hou, C., He, Y., Chu, X., & Luo, S. (2020). Evaluation of cultivar identification performance using feature expressions and classification algorithms on optical images of sweet corn seeds. *Agronomy (Basel)*, 10(9), 1268. http://dx.doi.org/10.3390/agronomy10091268.

Wan, X., Li, B., Chen, D., Long, X., Deng, Y., Wu, H., & Hu, J. (2021). Irrigation decision model for tomato seedlings based on optimal photosynthetic rate. *International Journal of Agricultural and Biological Engineering*, 14(5), 115-122. http://dx.doi.org/10.25165/j.ijabe.20211405.6148.

Wang, L. T., Huang, H., Chang, Y. H., Wang, Y. Q., Wang, J. D., Cai, Z. H., Efferth, T., & Fu, Y. J. (2022a). Biflavonoids from Ginkgo biloba leaves as a novel anti-atherosclerotic candidate: inhibition potency and mechanistic analysis. *Phytomedicine*, 102, 154053. http://dx.doi.org/10.1016/j.phymed.2022.154053. PMid:35567993.

Wang, Q., Jiang, Y., Mao, X., Yu, W., Lu, J., & Wang, L. (2022b). Integration of morphological, physiological, cytological, metabolome and transcriptome analyses reveal age inhibited accumulation of flavonoid biosynthesis in Ginkgo biloba leaves. *Industrial Crops and Products*, 187, 115405. http://dx.doi.org/10.1016/j.indcrop.2022.115405.

Wang, Y., He, H., Jiang, S., & Ma, H. (2022c). Nondestructive determination of IMP content in chilled chicken based on hyperspectral data combined with chemometrics. *International Journal of Agricultural and Biological Engineering*, 15(1), 277-284. http://dx.doi.org/10.25165/j.ijabe.20221501.6612.

Wang, X., Xing, X., Zhao, M., & Yang, J. (2020). Comparison of multispectral modeling of physiochemical attributes of greengage: Brix and pH values. *Food Science and Technology (Campinas)*, 41(Suppl 2), 611-618. http://dx.doi.org/10.1590/fst.21320.

Wu, D. D., Qu, C., Liu, X. G., Li, P., Gao, W., & Yang, H. (2021). A simple high-performance liquid chromatography method for the assay of flavonoids in Ginkgo biloba Leaves. *World Journal of Traditional Chinese Medicine*, 7(1), 47. http://dx.doi.org/10.4103/wjtcm.wjtcm_9_21.

Xin, Z., Jun, S., Yan, T., Quansheng, C., Xiaohong, W., & Yingying, H. (2020). A deep learning based regression method on hyperspectral data for rapid prediction of cadmium residue in lettuce leaves. *Chemometrics and Intelligent Laboratory Systems*, 200, 103996. http://dx.doi.org/10.1016/j.chemolab.2020.103996.

Yang, J. H., Meng, F. Z., Beseler, C. L., Li, H., Liu, X. M., Guo, Y. P., Qin, L. L., Zuo, R. Y., & Ren, S. P. (2021a). Application of ginkgo biloba l. extract at the early stage of tumor development helps cyclophosphamide inhibit the growth of tumor cells. *Biomedical and Environmental Sciences*, 34(12), 1010-1014. PMid:34981725.

Yang, X., Zhou, T., Su, X., Wang, G., Zhang, X., Guo, Q., & Cao, F. (2021b). Structural characterization and comparative analysis of the chloroplast genome of Ginkgo biloba and other gymnosperms. *Journal of Forestry Research*, 32(2), 765-778. http://dx.doi.org/10.1007/s11676-019-01088-4.

Ye, J., Yang, K., Li, Y., Xu, F., Cheng, S., Zhang, W., Liao, Y., Yang, X., Wang, L., & Wang, Q. (2022). Genome-wide transcriptome analysis reveals the regulatory network governing terpene trilactones biosynthesis in ginkgo biloba. *Tree Physiology*, 42(10), 2068-2085. http://dx.doi.org/10.1093/treephys/tpac051. PMid:35532090.

Zhang, D., Ji, H. W., Luo, G. X., Chen, H., Liu, S. C., & Mao, W. J. (2021). Insight into aroma attributes change during the hot-air-drying process of white shrimp using GC-MS, E-Nose and sensory analysis. *Food Science and Technology (Campinas)*, 42, e70820. http://dx.doi.org/10.1590/fst.70820.

Zhang, F., Li, D. X., Lu, D. Y., Lu, Y. F., Zhang, R., Zhao, L. L., Ji, S., Guo, M. Z., Du, Y., & Tang, D. Q. (2022a). Analysis of plasma free amino acids in diabetic rat and the intervention of Ginkgo biloba leaves extract using hydrophilic interaction liquid chromatography coupled with tandem mass-spectrometry. *Journal of Chromatography B, Analytical Technologies in the Biomedical and Life Sciences*, 1196, 123230. http://dx.doi.org/10.1016/j.jchromb.2022.123230. PMid:35349934.

Zhang, S., Gong, X., & Qu, H. (2022b). Near-infrared spectroscopy and HPLC combined with chemometrics for comprehensive evaluation of six organic acids in Ginkgo biloba leaf extract. *The Journal of Pharmacy and Pharmacology*, 74(7), 1040-1050. http://dx.doi.org/10.1093/jpp/rgab177. PMid:35294552.

Zhao, L. J., Liu, W., Xiong, S., Tang, J., Lou, Z., Xie, M., Xia, B. H., Lin, L. M., & Liao, D. F. (2018). Determination of Total Flavonoids Contents and Antioxidant Activity of Ginkgo biloba Leaf by Near-Infrared Reflectance Method. *International Journal of Analytical Chemistry*, 2018, 8195784. http://dx.doi.org/10.1155/2018/8195784. PMid:30154851.

Zou, K., Liu, X., Zhang, D., Yang, Q., Fu, S., Meng, D., Chang, W., Li, R., Yin, H., & Liang, Y. (2019). Flavonoid biosynthesis is likely more susceptible to elevation and tree age than other branch pathways involved in phenylpropanoid biosynthesis in ginkgo leaves. *Frontiers in Plant Science*, 10, 983. http://dx.doi.org/10.3389/fpls.2019.00983. PMid:31417595.

Zou, Z., Chen, J., Zhou, M., Zhao, Y., Long, T., Wu, Q., & Xu, L. (2022a). Prediction of peanut seed vigor based on hyperspectral images. *Food Science and Technology (Campinas)*, 42, e32822. http://dx.doi.org/10.1590/fst.32822.

Zou, Z., Wu, Q., Chen, J., Long, T., Wang, J., Zhou, M., Zhao, Y., Yu, T., Wang, Y., & Xu, L. (2022b). Rapid determination of water content in potato tubers based on hyperspectral images and machine learning algorithms. *Food Science and Technology (Campinas)*, 42, e46522. http://dx.doi.org/10.1590/fst.46522.