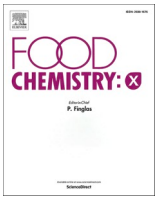
Food Chemistry: X 24 (2024) 101881

Contents lists available at ScienceDirect

Food Chemistry: X

journal homepage: www.sciencedirect.com/journal/food-chemistry-x



Rapid discrimination and quantification of chemotypes in Perillae folium using FT-NIR spectroscopy and GC–MS combined with chemometrics

Dai-xin Yu [a,b], Cheng Qu [a,b,*], Jia-yi Xu [a,b], Jia-yu Lu [a,b], Di-di Wu [a,b], Qi-nan Wu [a,b,c,*]

[a] Jiangsu Collaborative Innovation Center of Chinese Medicinal Resources Industrialization, Nanjing University of Chinese Medicine, Nanjing 210023, China
[b] School of Pharmacy, Nanjing University of Chinese Medicine, Nanjing 210023, China
[c] State Key Laboratory on Technologies for Chinese Medicine Pharmaceutical Process Control and Intelligent Manufacture, Nanjing University of Chinese Medicine, Nanjing 210023, China

ARTICLE INFO

Keywords:
Perillae folium
Chemotype
NIR spectroscopy
Chemometrics
Quantitative analysis

ABSTRACT

Perillae Folium (PF) is a well-known food and herb containing different chemotypes, which affect its quality. Herein, a method was proposed to classify and quantify PF chemotypes using gas chromatography–mass spectrometry (GC–MS) and Fourier transform-near infrared spectroscopy (FT-NIR). GC–MS results revealed that PF contains several chemotypes, including perilla ketone (PK) type, α-asarone (PP-as) type, and dillapiole (PP-dm) type, with the PK type being the predominant chemotype. Based on FT-NIR data, different chemotypes were accurately classified. The random forest algorithm achieved >90 % accuracy in chemotype classification. Furthermore, the main components of perilla ketone and isoegomaketone in PF were successfully quantified using partial least squares regression models, with prediction to deviation values of 3.76 and 2.59, respectively. This method provides valuable insights and references for the quality supervision of PF and other foods.

1. Introduction

Aromatic foods and herbs are widely used in spices, flavoring agents, and herbal remedies owing to their distinctive colors, aromas, and functions (Carvalho Costa et al., 2015). Perillae Folium (PF), derived from the annual herbaceous plant *Perilla frutescens* (L.) Britt. of the Lamiaceae family, is widely distributed in Asia, particularly in China, Japan, Korea, India and Nepal (Yu et al., 2017). As a popular aromatic food, PF is commonly used as a culinary spice and flavoring agent in soups, roasts, vegetable salads, and food colorants because of its distinctive aroma, basil-like taste, and pleasant green or purple hues (Fan et al., 2022). In addition to its edible value, PF is an important herbal medicine used to treat ailments such as food poisoning from fish or crab, food allergies, and common cold (Wu et al., 2023). Modern studies have found that volatile components, amino acids, fatty acids, phenolic acids, and flavonoids are the key nutritional and medicinal constituents of PF. These functional components exert various biological effects, including antioxidant, antimicrobial, antiallergenic, and antidepressant activities (Hou et al., 2022; Yu et al., 2017).

The quality of aromatic foods and herbs plays an important role in determining their flavor and commercial value, and many factors affect their quality, such as geographical origin (Cui et al., 2023), processing method (Zhao et al., 2023), and harvest time (Lu et al., 2023). As the most abundant constituents of PF, volatile compounds are commonly found in essential oils and are recognized as active and flavoring substances. The major volatile compounds identified in PF include perilla ketone, isoegomaketone, perillaldehyde, perillene, piperitenone, asarone, dillapiole, and caryophyllene (Ghimire et al., 2017; Jin et al., 2023). Owing to the various types and contents of volatile components, researchers have recently classified PF samples into distinct chemotypes (Ahmed & Tavaszi-Sarosi, 2019; Ghimire et al., 2017). The classification of chemotypes typically depends on the volatile compound content of the different samples (Rodríguez-Solana et al., 2014). High levels of these components are often considered key factors in determining the quality of PF and other aromatic foods and herbs.

PF has typically been subdivided into several chemotypes, including the PK type (primarily containing perilla ketone and isoegomaketone), PA type (primarily containing perillaldehyde), PL type (primarily

* Corresponding author at: Jiangsu Collaborative Innovation Center of Chinese Medicinal Resources Industrialization, Nanjing University of Chinese Medicine, Nanjing 210023, China
E-mail addresses: qucheng@njucm.edu.cn (C. Qu), wuqn@njucm.edu.cn (Q.-n. Wu).

https://doi.org/10.1016/j.fochx.2024.101881
Received 15 August 2024; Received in revised form 30 September 2024; Accepted 4 October 2024
Available online 5 October 2024

containing perillene), PT type (primarily containing piperitenone), PP type (primarily containing phenylpropanoid), and other composite chemotypes comprising two or three major volatile constituents (Ahmed & Tavaszi-Sarosi, 2019). Because of variations among these chemotypes, the flavor and quality of PF can differ considerably. Among all PF chemotypes, the PK type is commonly distributed in China and other countries (Zhou et al., 2023), and the main components of perilla ketone and isoegomaketone have been found to exhibit promising antifungal, antiobesity, and anti-inflammatory activities (Wang et al., 2022). Given the significant differences in aroma and quality among PF chemotypes, qualitative and quantitative analyses based on chemotypes may offer a new approach for PF quality assessment.

Traditional methods for assessing PF chemotype include gas chromatography (GC) and GC–mass spectrometry (GC–MS) (Sa et al., 2023). Although these methods provide high accuracy and reproducibility for chemotype identification, long analysis time and tedious sample preparation remain challenges. Near infrared spectroscopy (NIR) is a common vibrational spectroscopic technique that acquires spectral information by measuring the overtones and vibrational combinations related to hydrogen-containing groups (O—H, N—H, C—H) in the NIR wavelengths ranging from 750 to 2500 nm, indirectly reflecting chemical composition (Nagy et al., 2022). As a rapid, nondestructive, and environmentally friendly method, NIR coupled with chemometric analysis has been successfully applied for both qualitative and quantitative analyses of foods and herbs (He et al., 2023; Peng et al., 2024). Studies have shown that NIR can establish quantitative models based on volatile components for the rapid quality evaluation of *Florists chrysanthemum* (Fan et al., 2023) and *Monechma ciliatum* (Elrasheid Tahir et al., 2023). However, no reports have yet utilized these techniques to evaluate the quality of different PF chemotypes or other aromatic foods and herbs.

Hence, to develop a rapid and effective method for the chemotype discrimination of PF, this study investigated the feasibility of using NIR combined with chemometrics to distinguish different PF chemotypes and predict the contents of major volatile components. The specific aims were as follows: 1) detect the main volatile components using the GC–MS system and determine the chemotypes of PF, 2) obtain the NIR data of PF to construct discrimination models based on chemometric analysis, and 3) develop and optimize quantitative models using the partial least squares regression (PLSR) algorithm with NIR and GC–MS data to predict the contents of the main components in PF. The results are expected to facilitate rapid, environmentally friendly, and comprehensive quality evaluation of chemotypes in PF or other aromatic foods and herbs.

2. Materials and methods

2.1. Samples and reagents

In this study, sixty-three batches of commercial PF samples were collected from different markets in China (Table 1S). All samples were taxonomically identified as Perillae Folium by Prof. Qinan Wu (Nanjing University of Chinese medicine), and voucher specimens were deposited in the School of Pharmacy, Nanjing University of Chinese Medicine. The PF samples were initially pulverized using a high-speed grinder (FW-80, Tianjin Taisite Instrument Co., China) and then filtered through a 50-mesh sieve (355 μm ± 13 μm) to achieve a fine particle size powder. The prepared samples were stored in centrifuge tubes and kept in a desiccator until analysis. Reference standards of perilla ketone, isoegomaketone, L-limonene, α-asarone, and apiole (purity ≥95 %) were purchased from Chengdu Desite Bio-Technology Co., Ltd. (Chengdu, China). Reference standard of perillaldehyde with a purity ≥98 % was obtained from Chengdu MUST Bio-Technology Co., Ltd. (Chengdu, China). The chromatographic grade n-hexane and n-hexadecane (internal standard, purity ≥98 %) were purchased from Aladdin Chemical Reagents Co., Ltd. (Shanghai, China). The n-alkanes (C_7 – C_{40}) were obtained from Shanghai Yuanye Bio-Technology Co., Ltd. (Shanghai, China). The ultrapure water was supplied by a Mill-Q system (Millipore, United States).

2.2. Volatile component analysis via GC–MS

2.2.1. Extraction of volatile components by simultaneous distillation extraction

The volatile compounds in the PF samples were extracted using the simultaneous distillation extraction (SDE) method, as described in our previous study, with a slight modification (Zhou et al., 2021). Specifically, 5.0 g of the PF sample was weighed and mixed with 200 mL of ultrapure water in a round-bottom flask. Another flask containing 10 mL of n-hexane was used to collect the volatile components. Both flasks were connected to the SDE device and heated to a light boil using an electric heating mantle. The condensate system was turned on, and the process was performed for 1 h after the solvent in the flasks started to boil. Finally, when no liquid drop remained in the condensation tube, the extracted solutions were collected in brown vials and treated with anhydrous sodium sulfate to remove the water. The samples were stored in a refrigerator at −20 °C for 24 h for subsequent analysis.

2.2.2. GC–MS conditions

An Agilent 7890B gas chromatograph system coupled to an Agilent 7000C triple quadrupole mass spectrometer (Agilent Technologies, Santa Clara, CA, USA), equipped with an HP-5MS fused silica capillary column (30 m × 0.25 mm, 0.25 μm), was used to detect and separate the volatile compounds in PF. For the GC conditions, the temperature of the injection port was set to 220 °C, injection volume was 1 μL, split ratio was 10:1, and carrier gas was high-purity helium (99.999 %) at a flow rate of 1.0 mL/min. The oven temperature program started from 50 °C (held for 3 mins), then increased to 100 °C at a rate of 10 °C/min (held for 3 mins), subsequently heated to 200 °C at a rate of 5 °C/min (held for 3 mins), and finally ramped to 220 °C, remaining for 3 mins. For MS conditions, an electron impact (EI) ionization chamber, operating in the full-scan mode, were used. The EI energy was 70 eV, and the ion source temperature was 230 °C. The transfer line temperature was 280 °C. The full-scan mode was operated at a mass range of 50.0–500.0 amu with a quadrupole temperature of 150 °C. The scanning time was 150 ms, and the solvent delay time was 3 min.

2.2.3. Qualitative and quantitative analyses

The metrics of retention index (RI) values and R match values were used to identify the volatile components in PF. Generally, an experimental RI value that does not differ from the reported RI value by >30 is considered a key indicator for compound identification. At the same time, a higher R match value, based on mass spectral comparison, is recognized as an indicator of identification (Biancolillo et al., 2022). Herein, the potential volatile compounds in PF were initially identified based on R match values of >750. An n-alkane (C_7 – C_{40}) solution was then analyzed *via* GC–MS under the same conditions to calculate the RI values of each peak in PF. The corresponding compounds were identified by comparing the measured RI values with the theoretical RI values from the National Institute of Standards and Technology (NIST, version 14.0) library, while also referencing the highest R match scores. Finally, some volatile standards were used to verify the accuracy of the identified results. The specific calculation formula is shown in Eq. (1):

$$RI = 100 \times n + 100 \times \frac{tx - tn}{tn + 1 - tn}, \quad (1)$$

where RI is the retention index, n is the number of carbon atoms in the n-alkane, t_x is the retention time of the compound to be identified, t_n is the retention time of the n-alkane corresponding to n carbon atoms, and t_{n+1} is the retention time of the n-alkane corresponding to n + 1 carbon atoms.

Relative quantification was performed according to the ratio of the ion flow peak area of each component in the total ion flow chromatogram to the internal standard (n-hexadecane) peak area according to the following formula (2):

$$Ci = \frac{Ai}{A0} \times C0, \quad (2)$$

where C_i represents the relative content of the identified compounds, C_0 represents the content of the internal standard, A_i represents the peak area of the identified volatile compounds, and A_0 refers to the peak area of the internal standard.

2.3. FT-NIR analysis

2.3.1. Acquisition of NIR spectra

The NIR spectra of the PF samples were acquired using an Antaris™ II FT-NIR spectrophotometer (Thermo Fisher Scientific Co., USA) equipped with a rotating sample-cup spinner, an extended InGaAs detector, and a tungsten halogen lamp as the light source. The FT-NIR parameters were set as follows: diffuse reflectance mode, 32 scans within the range of 10,000–4000 cm^{-1} and resolution, 8 cm^{-1}. The Result software (Antaris™ II System, Thermo Fisher Scientific Co., USA) was used to acquire NIR data. The entire NIR procedure was conducted at a room temperature of 20 °C – 24 °C, with relative humidity ranging from 30 % to 35 %. A specific desiccant was used to keep the NIR device dry and ensure the reliability of the acquired spectra. Each NIR spectrum was recorded in triplicate, and the average spectrum was used for further analysis.

2.3.2. Preprocessing of NIR spectra

Generally, some irrelevant interference factors, including background noise and instrumental noise, can arise during NIR acquisition. These undesirable effects may be attributed to light scattering, optical range differences, and sample particle size (Biswas & Chaudhari, 2024). It is essential to apply various preprocessing methods to obtain effective NIR data. Herein, eight pretreatments were employed, including multiplicative scatter correction (MSC), standard normal variate transformation (SNV), first and second derivative (1st Der and 2nd Der) methods, as well as their combinations: MSC + 1st Der, MSC + 2nd Der, SNV + 1st Der, and SNV + 2nd Der. The MSC and SNV methods were employed to eliminate the interference caused by light scattering and particle size. The 1st Der and 2nd Der methods were applied to correct baseline drifts and separate the overlapping spectral bands. All pretreatments were performed using the Savitzky–Golay (SG) algorithm to reduce experimentally generated noise. Specifically, the SG algorithm was configured with 17 smoothing points and a polynomial order of 3 for qualitative models and 13 smoothing points and a polynomial order of 3 for quantitative models. These preprocessing methods were combined to filter the most suitable combinations for further modeling analyses.

2.4. Model construction

2.4.1. Discriminative models

Herein, traditional pattern recognition methods and modern machine learning algorithms were used to classify the chemotypes of the PF samples. Initially, principal component analysis (PCA) was performed for the categorical exploration of the generated data. As an unsupervised model, PCA reduces the dimensionality of complex data by projecting the variables into the first few components, providing intuitive group clustering visualization (Zhou et al., 2020). Subsequently, two common supervised models, partial least squares-discriminant analysis (PLS-DA) and orthogonal partial least squares-discriminant analysis (OPLS-DA), were employed to separate the different groups. Based on partial least squares (PLS) theory, PLS-DA and OPLS-DA can address the issues of intricate variables and covariance in classification (Yu et al., 2022). Finally, three machine learning algorithms, including K-nearest neighbor (KNN), decision tree (DT), and random forest (RF), were developed to extract the most effective information and improve classification accuracy. Briefly, the KNN algorithm selects k objects closest to the test samples in a given space based on Euclidean distance and then fits the optimal discriminative model (Zhang & Wang, 2023). The DT algorithm divides the training dataset into groups with the highest variance between the dependent and independent variables, which can be used to construct decision trees for classification. RF is an ensemble algorithm that generates a forest of decision trees, thereby reducing the impact of outliers and model overfitting and improving the modeling accuracy (Sun, Li, et al., 2021).

2.4.2. Quantitative models

The PLSR algorithm was used to establish quantitative models for predicting the content of the main components in the PF samples. As the most widely utilized spectral multivariate correction analysis, PLSR can fully leverage spectral information to obtain the best correction model. This method decomposes the spectral matrix and the content matrix, removes the irrelevant noise signals from both matrices, and then obtain the relevance between the eigenvectors and indicators (Metz et al., 2021). In the present study, the NIR spectral data were represented by matrix (*X*) and the contents of volatile components based on GC–MS were represented by matrix (*Y*), from which regression models were established.

2.5. Statistical analysis and model evaluation

The clustering heat map, box diagrams, and PCA model were generated using Origin 2021b software (Northampton, MA, USA). The PLS-DA and OPLS-DA models were completed using the SIMCA-P software (Version 14.1, Umetrics, Sweden). The fitting and predictive efficiency of the PLS-DA and OPLS-DA models were assessed using coefficients of determination (R^2X and R^2Y) and predictive capability (Q^2) parameters. R^2X and R^2Y represent the variance scores of the *X* and *Y* matrices, respectively. The *X* matrix refers to GC–MS or FT-NIR data, whereas the *Y* matrix corresponds to the associated categories. Q^2 represents the prediction accuracy of the model. R^2X, R^2Y, and Q^2 values equal to 1.0, suggesting highly effective models (Fu et al., 2021; Li et al., 2024). Machine learning algorithms, including KNN, DT, and RF, were applied to establish classification models based on FT-NIR data using MATLAB R 2022b software (MathWorks Inc., Natick, MA, USA). The area under the curve (AUC), accuracy, precision, recall, and F1-scores were used to evaluate the performance of KNN, DT, and RF models; ideally, these indices are close to 1.00 (Yu et al., 2022).

The PLSR model was implemented in TQ Analyst 9.0 software (Thermo Fisher Scientific Co., USA). The parameters, including the coefficient of determination for calibration (R_c^2), coefficient of determination for prediction (R_p^2), root mean square error of estimation (RMSEC), root mean square error of prediction (RMSEP), and values of prediction to deviation (RPD), were obtained to evaluate the performance of PLSR models. Generally, a good calibration model should have high values of R_c^2 and R_p^2 and low values for RMSEC and RMSEP. Additionally, RPD scores reflect the overall predictive ability of the PLSR model, with RPD of >2.5 indicating excellent performance (Zou et al., 2024).

3. Results and discussion

3.1. Volatile component analysis of the PF

3.1.1. Identification and quantification of volatile compounds via SDE-GC–MS

Herein, the volatile compounds in the PF samples were detected and analyzed using the SDE-GC–MS system. The total ion chromatograms of

all the PF samples are shown in Fig. 1A. Using the NIST 14 database, 20 main volatile components were identified by comparing their mass information (Fig. 1S). The identified compounds include major volatile components of PF previously reported (Jin et al., 2023), such as perilla ketone, perillaldehyde, isoegomaketone, caryophyllene, β-asarone, α-asarone, dillapiole, and apiole (Table 1). Based on the peak area of each identified compound, the relative contents of the PF samples were calculated using the ratio of the peak intensity of the sample to that of the internal standard. As shown in Table 2S, the relative contents of perilla ketone, isoegomaketone, caryophyllene, α-asarone, and dillapiole were significantly higher than those of the other volatile compounds in PF. Owing to the significant variances in the types and contents of the major volatile components, 63 PF samples were classified into three chemotypes. Specifically, 41 PF samples were grouped into PK (isoegomaketone) types, whereas 22 PF samples were categorized into phenylpropanoids (PP) types, respectively. The PP type samples can be

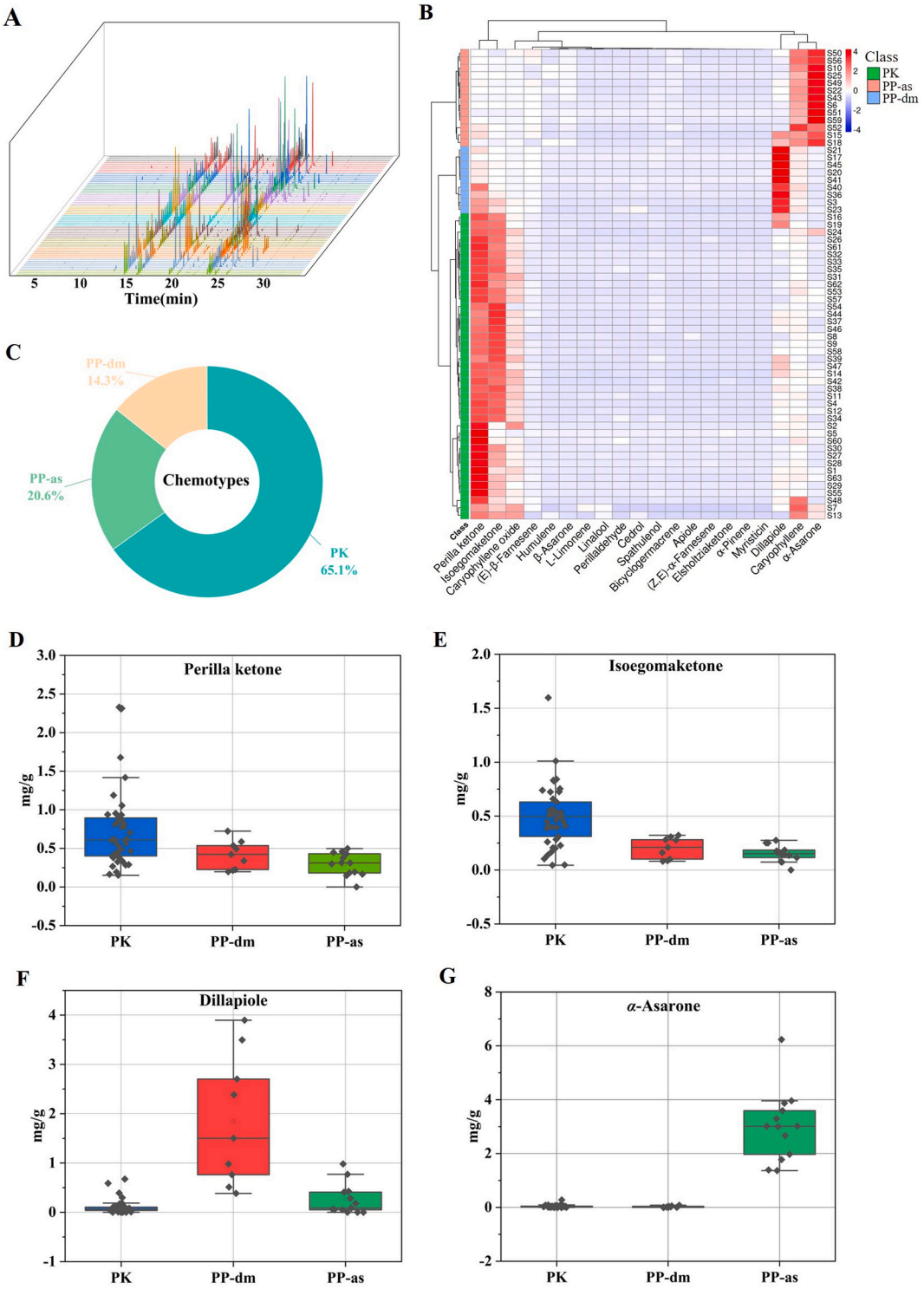


Fig. 1. The total ion chromatogram (TIC) of the PF samples (A); the clustering heat map of PF (B); percentage of different chemotypes in PF (C); the relative contents of perilla ketone (D), isoegomaketone (E), dillapiole (F), and α-asarone (G) in different PF chemotypes.

Table 1
The main volatile components identified in PF by GC–MS.

NO.	RT (min)	Compounds	Formula	CAS	R Match	RI (measured)	RI (reference)
1	6.823	α-Pinene	$C_{10}H_{16}$	7785-70-8	781	936	937
2	8.634	L-Limonene*	$C_{10}H_{16}$	5989-27-5	868	1031	1036
3	10.283	Linalool	$C_{10}H_{18}O$	106–24-1	827	1100	1099
4	13.438	Elsholtziaketone	$C_{10}H_{14}O_2$	488–05-1	862	1204	1177
5	14.849	Perilla ketone*	$C_{10}H_{14}O_2$	553–84-4	878	1251	1253
6	15.621	Perillaldehyde*	$C_{10}H_{14}O$	18,031–40-8	846	1276	1272
7	16.377	Isoegomaketone*	$C_{10}H_{12}O_2$	34,348–59-9	765	1324	1302
8	19.768	Caryophyllene	$C_{15}H_{24}$	87–44-5	882	1425	1419
9	20.655	(*E*)-β-Farnesene	$C_{15}H_{24}$	18,794–84-8	770	1458	1457
10	20.667	Humulene	$C_{15}H_{24}$	6753-98-6	847	1459	1454
11	21.627	(Z, E)-α-Farnesene	$C_{15}H_{24}$	26,560–14-5	813	1496	1491
12	21.763	Bicyclogermacrene	$C_{15}H_{24}$	24,703–35-3	840	1501	1495
13	22.329	Myristicin	$C_{11}H_{12}O_3$	607–91-0	884	1504	1519
14	23.830	Spathulenol	$C_{15}H_{24}O$	6750-60-3	765	1586	1576
15	23.905	Caryophyllene oxide	$C_{15}H_{24}O$	1139-30-6	861	1589	1581
16	24.368	Cedrol	$C_{15}H_{26}O$	77–53-2	799	1609	1598
17	24.692	β-Asarone	$C_{12}H_{16}O_3$	5273-86-9	759	1623	1622
18	24.809	Dillapiole	$C_{12}H_{14}O_4$	484–31-1	758	1628	1628
19	26.044	α-Asarone*	$C_{12}H_{16}O_3$	2883-98-9	841	1682	1680
20	26.117	Apiole*	$C_{12}H_{14}O_4$	523–80-8	771	1686	1682

Note: RT: retention time; CAS: chemical abstracts service registry number; R Match: reverse matching score; RI (measured): retention index calculated by n-alkanes and actual measured value; RI (reference): the theoretical retention index in NIST 14 library; the * indicates that the ingredient was identified by standards.

further classified into PP-as and PP-dm types based on the distinctions between α-asarone and dillapiole. The clustering heat map (Fig. 1B) shows a clear distribution of different chemotypes in PF, with higher contents of perilla ketone and isoegomaketone in PK type samples and higher levels of dillapiole or α-asarone in PP type samples. The PK type samples accounted for >65 % of PF, with the remaining PP-as and PP-dm types accounting for nearly 35 % (Fig. 1C). This indicates that PK type is the predominant chemotype in commercial PF, consistent with the findings of previous reports (Zhou et al., 2023). Regarding the specific contents of different chemotypes, the mean values of perilla ketone and isoegomaketone in PK type samples were 0.7625 and 0.5000 mg/g, respectively, which were significantly higher than those in the PP type samples (Fig. 1D and E). The average content of dillapiole in the PP-dm type was 1.8459 mg/g, whereas the mean value of α-asarone in the PP-as type was 3.0105 mg/g (Fig. 1F and G). Although the PK type samples also contained dillapiole and α-asarone, their levels were lower than those of perilla ketone and isoegomaketone. Overall, the commercial PF samples can be categorized into three common chemotypes. Considerable variations in the contents of major compounds among the different PF chemotypes may be attributed to genetic and environmental factors. Therefore, the classification and assessment of the qualities of various PF chemotypes are crucial.

3.1.2. Chemometric analysis via traditional PCA, PLS-DA, and OPLS-DA methods

To better understand the distribution trends of the different chemotypes in the PF samples, several pattern recognition models, including PCA, PLS-DA, and OPLS-DA, were employed. The PCA model was initially applied for the unsupervised exploration of the generated data. Despite some overlap, the PK, PP-as, and PP-dm types were clustered into three categories (Fig. 2A), indicating acceptable separation. As shown in the loading plot (Fig. 2B), perilla ketone and isoegomaketone significantly contributed to the classification of the PK type samples. α-Asarone and dillapiole played important roles in the categorization of PP-as and PP-dm type samples, respectively. Compared with the PCA model, PLS-DA and OPLS-DA are supervised methods that maximize the separation between observation groups, resulting in better prediction capability. In the PLS-DA model, good separation was achieved for PF samples of different chemotypes, with the main parameters of $R^2X = 0.502$, $R^2Y = 0.849$, and Q^2 (*cum*) = 0.737, indicating strong explanatory and predictive capacity. The three chemotypes were significantly classified compared with the PCA model (Fig. 2C). Permutation tests were conducted 200 times to determine whether the model was overfitting. The results (Fig. 2D) indicated that the intercept of Q^2 (−0.408) was <0.05, which enhanced the robustness and reliability of the model. The OPLS-DA model was conducted using one predictive component and two orthogonal components through seven-fold cross-validation, with $R^2X = 0.416$, $R^2Y = 0.870$, and $Q^2 = 0.793$. Better separation efficiency was achieved, with the PK and PP type samples on the left and right sides, respectively (Fig. 2E). Permutation tests also demonstrated that the OPLS-DA model was not overfitting because the intercept of Q^2 was <0.05 (Fig. 2F). Overall, the PCA, PLS-DA, and OPLS-DA models confirmed their feasibility in discriminating different PF chemotypes using volatile component data combined with multivariate statistics. The above results show that PF samples from different sources can be classified into different chemotypes. The quality of PF may be affected by variations in the major volatile components of different chemotype samples. However, because of undesirable factors such as tedious extraction in SDE and the long analysis time of GC–MS, further studies are necessary for the rapid discrimination of chemotypes in PF.

3.2. Qualitative analyses based on FT-NIR spectroscopy

3.2.1. NIR spectral features and optimal preprocessing

The raw NIR spectra of PF were collected in the wavenumber range of 10,000–4000 cm^{-1} (Fig. 3A). Overall, the intensity of the absorption peaks increased, as seen in the weaker absorption peaks between 5000 and 4000 cm^{-1}, and higher absorption values between 10,000 and 5000 cm^{-1}. The average NIR spectra exhibited valuable chemical information about PF from different chemotypes. It can be seen that the characteristic peaks of the NIR spectrum were primarily located at 8396, 6876, 5782, 5168, 4628, 4331, and 4254 cm^{-1} (Fig. 3B). Generally, the absorption peak at 8396 cm^{-1} is from the second overtone of C—H stretching (Zhao et al., 2024). The peak around 6876 cm^{-1} is assigned to O—H or N—H stretching vibrations in the first overtone (Hong et al., 2019). The band at 5782 cm^{-1} is induced by the first overtone of C—H stretching vibrations, and the wavelength of 5168 cm^{-1} corresponds to the combination of O—H and C—O stretching (Zhao et al., 2020). The absorption peak near 4628 cm^{-1} corresponds to the combined vibrations of C—H_2 symmetric stretching and bending. The band around 4331 cm^{-1} is related to the overtone vibrations of C—H bending, whereas the band at 4254 cm^{-1} corresponds to the first overtone of the C—H stretching vibrations (Xia et al., 2024).

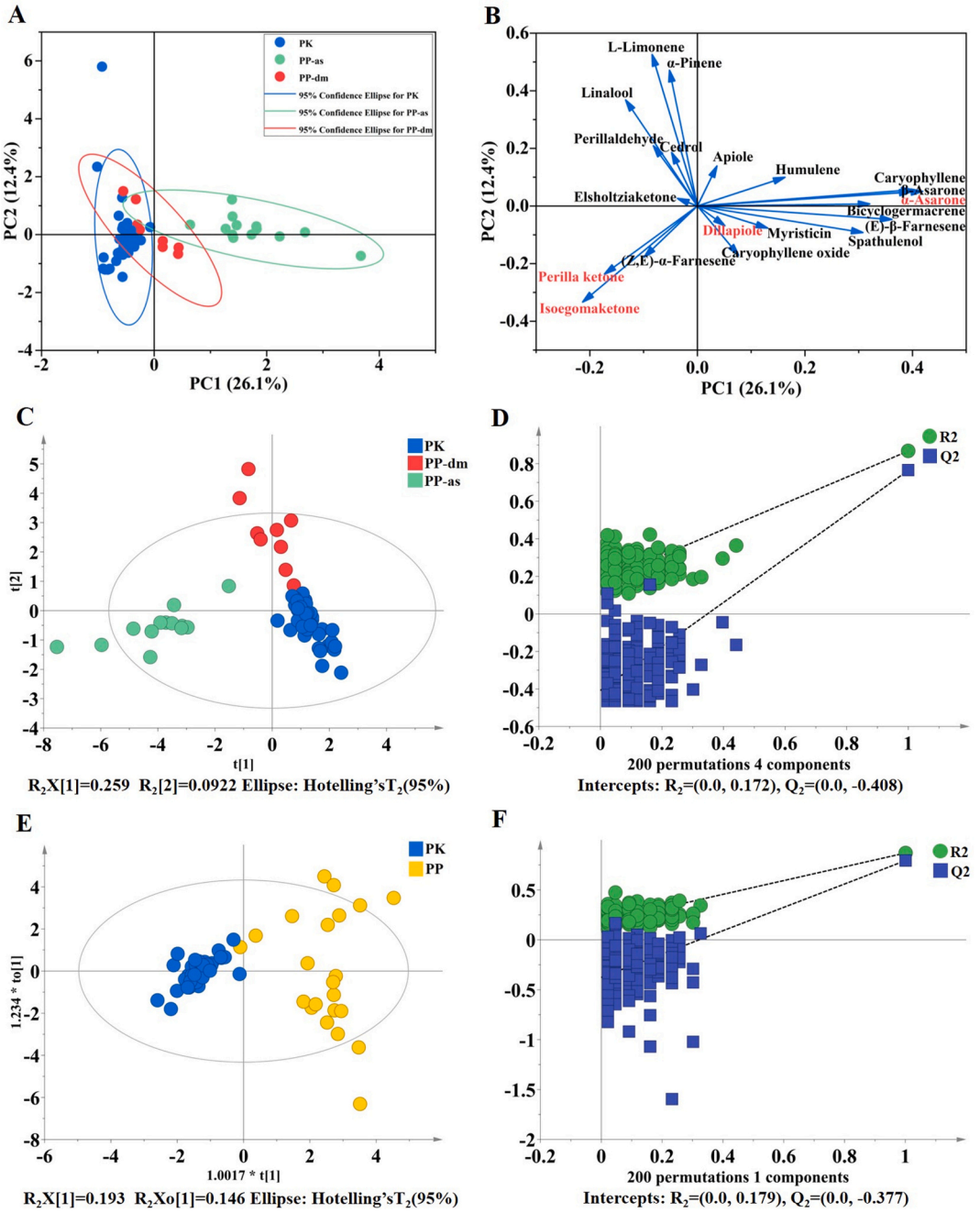


Fig. 2. Score plot (A) and loading analysis (B) of PCA model for three PF chemotypes; score plot of PLS-DA model (C) and OPLS-DA model (E) for different PF chemotypes; cross-validation results with 200 calculations using a permutation test (D and F).

The vibrational differences between the NIR absorption peaks may be important factors for the quality discrimination of different chemotypes in the PF samples. However, several invalid information exists in complex spectra, and it is difficult to interpret such information through visual observation. To solve this problem, eight NIR preprocessing methods were compared using the TQ Analyst software, and the prediction accuracy and performance index were used to assess the reliability of the models. Table 3S shows that except for SNV pretreatment, the preprocessing methods exhibited higher accuracy than the original spectra for discriminating PF chemotypes. Among them, the combination of SNV + 2nd Der was selected as the best method with a high predictive accuracy of 90.48 % and a good performance index of 85.0. The optimized NIR spectra are shown in Fig. 2S.

3.2.2. PF Chemotype discrimination by PLS-DA and OPLS-DA

Based on the optimal preprocessing method, the feasibility of NIR in discriminating PF chemotypes was further explored. The PLS-DA model was fitted using two principal components, with $R^2Y = 0.574$ and $Q^2 = 0.354$, indicating that 57.4 % and 35.4 % of the total variation can be explained and predicted, respectively. As shown in Fig. 3C, the PF samples of PK, PP-as and PP-dm types could be clustered independently with an initial tendency toward separation, and only a few PF samples were incorrectly classified into other categories. The OPLS-DA model was further employed to validate the reliability of the PLS-DA result. The main parameters R^2Y and Q^2 were 0.775 and 0.545, respectively, demonstrating that the OPLS-DA model has good explanatory power and predictive ability in classifying the PK and PP chemotypes of PF. Fig. 3E exhibited a distinctive result in chemotype classification, which was

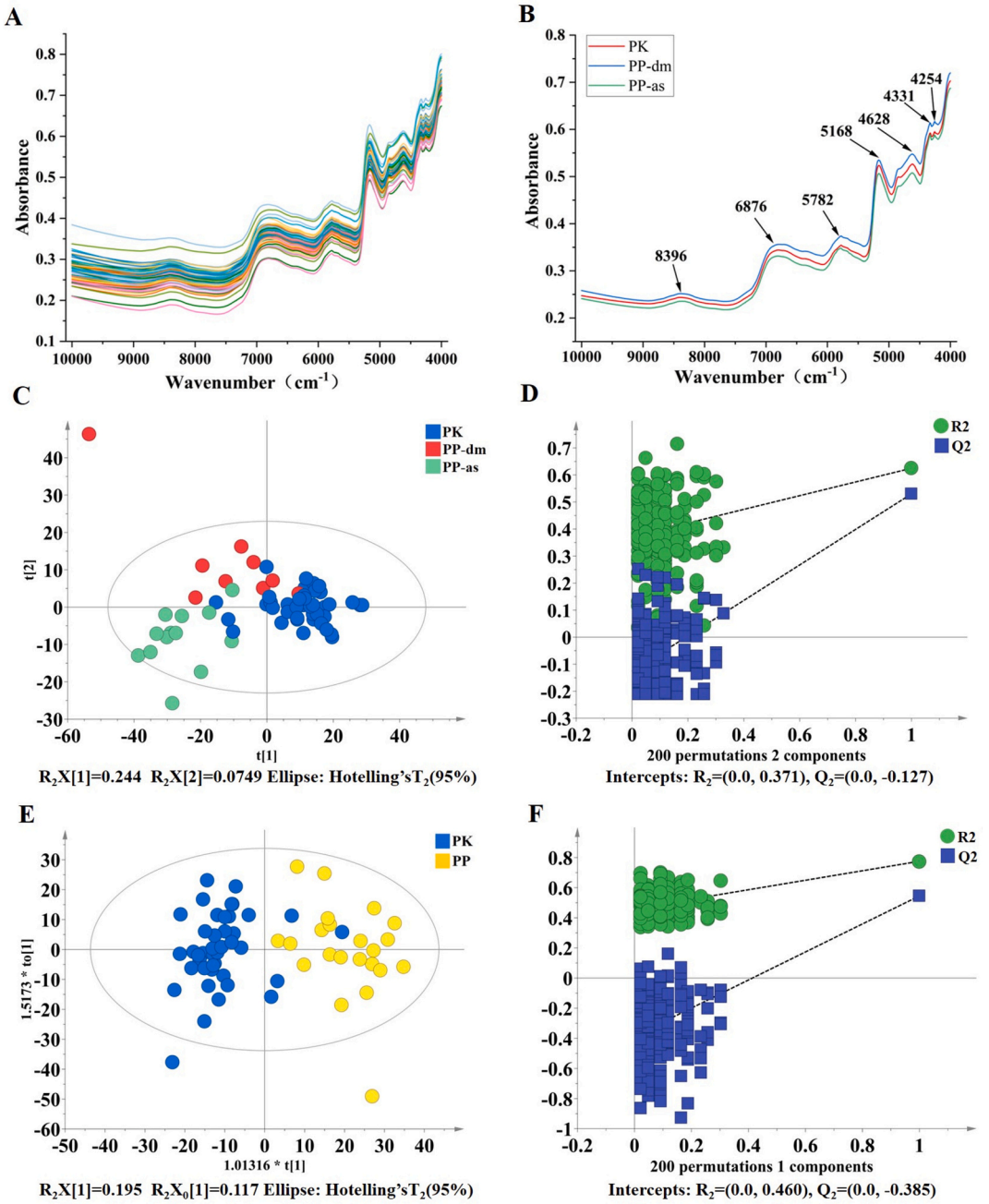


Fig. 3. Average raw NIR spectra (A) and characteristic peaks (B) of the PF samples; score plots of PLS-DA model (C) and OPLS-DA model (E) for different PF chemotypes; cross-validation results with 200 calculations using a permutation test (D and F).

consistent with the modeling from GC–MS. In addition, 200 times of permutation tests were performed on the PLS-DA (Fig. 3D) and OPLS-DA (Fig. 3F) models, and the Q^2 values were − 0.127 and − 0.385, respectively, indicating that the models did not overfit. The above results indicate that NIR spectral information can initially be used to classify PF chemotypes from different commercial sources. However, the Q^2 parameters in PLS-DA and OPLS-DA were close to or below 0.5, indicating that the predictive ability of these models was not adequately strong for chemotype classification. Machine learning algorithms should be developed to construct a classification model to solve this problem.

3.2.3. *PF Chemotypes discrimination based on machine learning*

To verify the accuracy and reliability of the above results and achieve a more efficient classification of PF chemotypes based on NIR data, three machine learning algorithms, namely KNN, DT, and RF, were developed for further analysis. The optimal NIR spectra were processed using the SNV + 2nd Der method to construct data matrices and then imported into MATLAB 2020b software to create KNN, DT, and RF models. A flowchart illustrating the algorithmic process is presented in Fig. 4A, and the fitting results are presented in Fig. 5. Specifically, 60 % and 40 % of the PF samples were selected as the training and testing sets, respectively, to construct the classification models. The receiver operating characteristic (ROC) curves were fitted to ensure the classification performance of the three models (Fig. 4B). The AUC values were calculated from the ROC curves: 0.8987 for KNN, 0.9281 for DT, and 1.000 for RF. These results indicate that the three models achieved optimal performance in the classification task. Accuracy is an important metric to assess the performance of algorithmic models. Herein, the training accuracy of the KNN, DT, and RF classifiers was 100.0 %, 97.3 %, and 100.0 %, respectively, and the total predictive accuracy was 84.6 %,

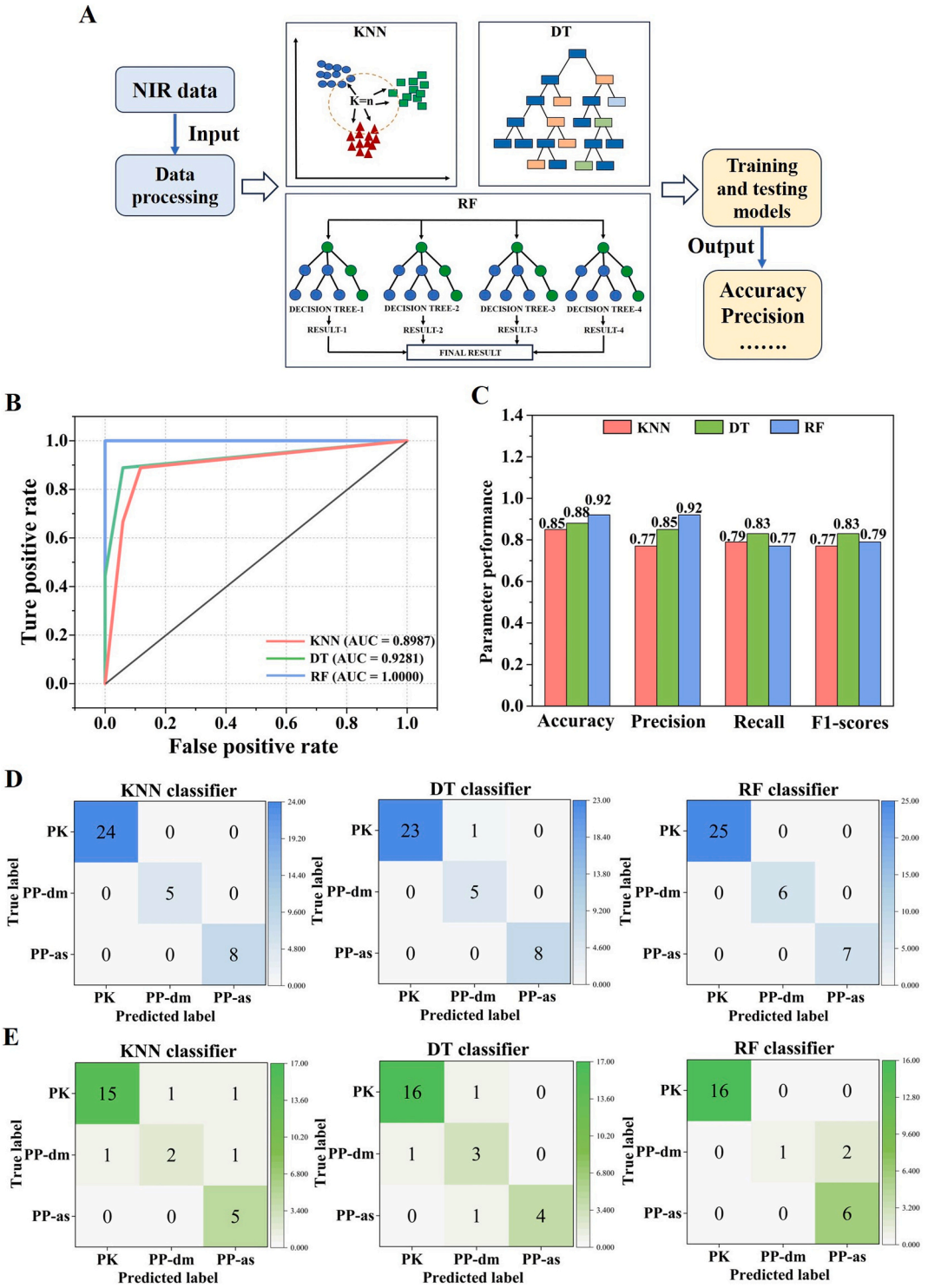


Fig. 4. The basic flowchart of three machine learning algorithms (KNN, DT, and RF) (A); the ROC curves and AUC values in three classifiers (B); the evaluation metrics of three algorithms (C); confusion matrixes of training sets (D) and testing sets (E) for classification of three chemotypes (PK, PP-as, and PP-dm types) based on three models.

88.5 %, and 92.0 %. This result indicates that the three classifiers successfully distinguished different PF chemotypes and exhibited strong explanatory ability. The confusion matrix diagrams are shown in Fig. 4D and E. It is clear that the RF model exhibits the highest discriminative accuracy, with only two PP-dm type samples misclassified into the PP-as type group. The PF samples from the PP-dm type were misjudged in all three classifiers. In addition to accuracy, other metrics (precision, recall, and F1-scores) were calculated to evaluate the reliability of the algorithms (Fig. 4C), and these metrics were satisfactory for each model, further enhancing the stability and robustness of the classification. Furthermore, we explored the discriminant accuracy of classifying the PK and PP types based on the three algorithms. The modeling results and evaluation indicators are shown in Fig. 3S and 4S. When grouping the PP-dm and PP-as type samples, the accuracy of the model improved, as RF misclassified only one sample. The results demonstrated that NIR spectral information can rapidly achieve the classification of

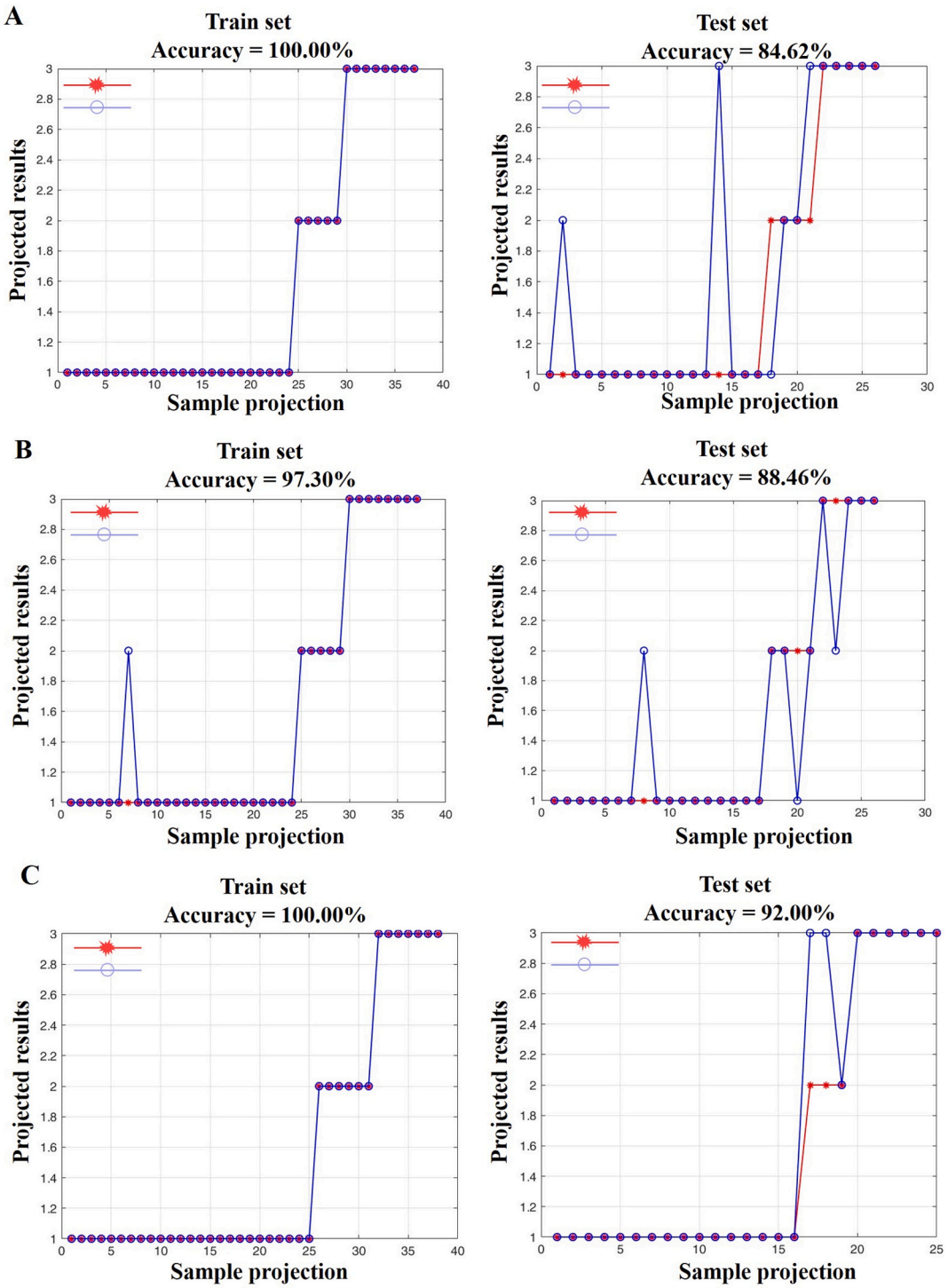


Fig. 5. The modeling training and testing process of KNN (A), DT (B), and RF (C) algorithms in three chemotypes (PK, PP-as, and PP-dm types).

chemotypes in PF, and the RF classifiers showed the best discriminatory performance in both two or three chemotypes. Overall, the FT-NIR technique combined with machine learning algorithms can be recognized as an effective tool for discriminating PF samples from different chemotypes.

3.3. Quantitative analysis based on FT-NIR spectroscopy

3.3.1. Selection of the spectral pretreatment method

As key constituents, perilla ketone and isoegomaketone play crucial roles in determining the flavor quality of PF. Previous studies have revealed that perilla ketone and isoegomaketone are the primary indicators of the PK chemotype in PF and they can also be detected and quantified in other PF chemotypes (Zhou et al., 2023). Given that NIR spectra contain rich chemical information, the feasibility of the quantitative prediction of perilla ketone and isoegomaketone in PF was investigated. PLSR was employed to establish quantitative models, which were analyzed using multivariate calibration through a linear correlation between the independent variable X (actual concentration of volatile compounds based on GC–MS analysis) and the dependent variable Y (spectral data from FT-NIR analysis). Eight different pretreatments (MSC, SNV, 1st Der, 2nd Der, MSC + 1st Der, MSC + 2nd Der,

SNV + 1st Der, and SNV + 2nd Der.) were compared and the main metrics of the different PLSR models are listed in Fig. 6A and Table 4S. Regarding perilla ketone, an optimal quantification model was developed after preprocessing MSC + 1st Der, with an RPD value of 3.76, R_c^2 and R_p^2 values of 0.9065 and 0.8965, and RMSEC and RMSEP values of 0.2420 and 0.1320, respectively. When comparing the eight pretreatments, only 1st Der, MSC + 1st Der, and SNV + 1st Der showed better prediction in PLSR than that using the raw spectra, indicating that 1st Der played an important role in correlating perilla ketone with spectral information of PF. For isoegomaketone, all pretreatments resulted in better predictions than those based on raw spectra. Interestingly, the combination of MSC + 1st Der also exhibited the best prediction performance, with an RPD value of 2.59, R_c^2 and R_p^2 values of 0.9389 and 0.8361, and RMSEC and RMSEP values of 0.1090 and 0.1100, respectively. Generally, the 1st Der method can remove baseline drifts, separate overlapping NIR bands, and reduce the effects of noise, thereby improving the accuracy of prediction results (Nagy et al., 2022). After combining MSC with the 1st Der, the PLSR models achieved the best results, indicating that MSC further enhanced the utilization of spectral information by eliminating the interference from spectral light scatter and particle size (Zheng et al., 2023). The RPD value is a key parameter for evaluating the PLSR model, reflecting the overall predictive ability of the model. In this study, the RPD values in the two optimal PLSR models were > 2.5, indicating that the selected MSC + 1st Der method exhibited excellent content prediction performance (Zou et al., 2024). In addition, the R_c^2 and R_p^2 values were close to each other in both optimal PLSR models, demonstrating that the established models were neither underfitted nor overfitted (Chen et al., 2024). Overall, the spectral pretreatment using the MSC + 1st Der combination improved the performance of the PLSR model for the quantitative prediction of perilla ketone and isoegomaketone in PF.

3.3.2. *Regression curves of the main components obtained using the PLSR model*

The calibration models for the PF compounds were constructed using the selected optimal PLSR models. The quantitative regression curves in Fig. 6B (for perilla ketone) and Fig. 6C (for isoegomaketone) demonstrate that the NIR spectra of the PF samples have a linear relationship with the content values of perilla ketone and isoegomaketone. Generally, the regression line represents the most desirable result in a quantitative model, and scatter points close to this line indicate that the model is excellent (Sun, Liu, et al., 2021). All the data points in the calibration and prediction sets are tightly clustered around the diagonal lines, indicating the excellent predictive ability of the models for the studied quality components, allowing for better practical applications (Tang et al., 2023). Currently, some reports exist on the rapid prediction of volatile compounds in various foods using FT-NIR and GC–MS techniques, providing methodological references for the quality control of volatile components (Fan et al., 2023; Tahir et al., 2021). However, related studies on PF are lacking. When using FT-NIR combined with PLSR to rapidly predict volatile compounds in various foods, our research observed a similar trend (Elrasheid Tahir et al., 2023). The current findings confirmed that the PK chemotype can be clearly separated, and the two main volatile compounds can be accurately and simultaneously predicted using FT-NIR spectroscopy.

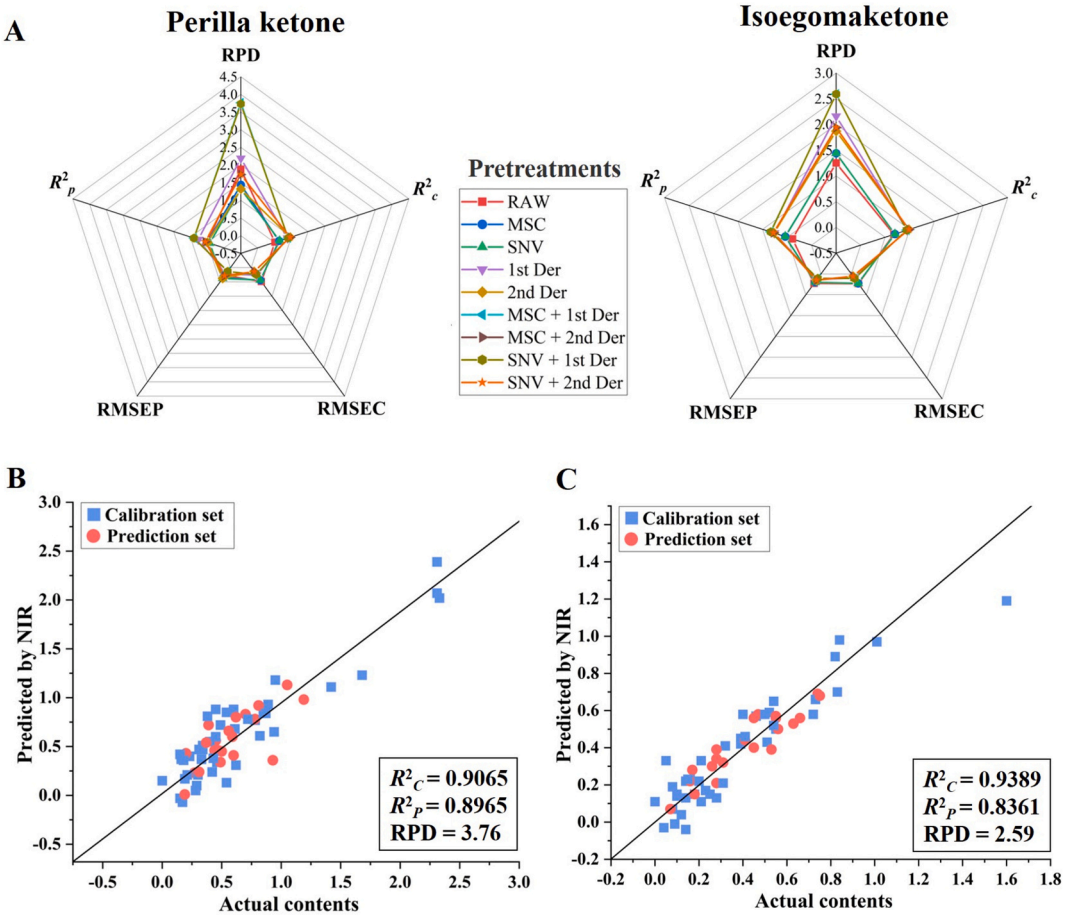


Fig. 6. The main parameters of PLSR models for determination of perilla ketone and isoegomaketone based on different preprocessing methods (A); quantitative regression curves of perilla ketone (B) and isoegomaketone (C) in PK type by PLSR model.

3.4. Summary

To the best of our knowledge, this is the first study to use FT-NIR for the rapid classification of different PF chemotypes. As a fast and nondestructive method, FT-NIR offers unique advantages for the characterization of chemical information. Machine learning algorithms exhibited high chemotype discrimination performance. The quantitative regression models successfully predicted the contents of major PF components. This study achieved the initial discrimination of PF chemotypes, providing a rapid quality evaluation method for PF and other aromatic foods or herbs from a novel perspective based on chemotype variations. However, because of the limited sample size, the current strategy focused only on a few chemotypes with high PF percentages, such as the PK and PP types. According to other reports, PF also contains chemotypes such as PA type (perillaldehyde), C type (citral), EK type (elsholtziaketone), PL type (perilla perillene), PT type (piperitenone), and SF type (shisofuran) (Ahmed & Tavaszi-Sarosi, 2019). In the future, it will be essential to collect more PF samples with different chemotypes for further analysis using our established method and to develop quantitative models for other major components. In addition, this strategy can be explored for chemotype classification and quality evaluation in other aromatic foods and herbs, such as mint, anise, and pepper.

4. Conclusion

Herein, an accurate and rapid strategy was used for the first time to classify and quantify different PF chemotypes using FT-NIR and GC–MS techniques combined with machine learning methods. The GC–MS results showed that the PF samples can be divided into PK, PP-as, and PP-dm chemotypes based on the main volatile compounds. The average perilla ketone and isoegomaketone contents in the PK type samples were 0.7625 and 0.5000 mg/g, respectively. Chemometric analyses using PCA, PLS-DA, and OPLS-DA demonstrated that different PF chemotypes can be accurately distinguished using GC–MS data. The FT-NIR technique was introduced to obtain spectral information from different sources and enhance the efficiency of PF chemotype classification. The PLS-DA and OPLS-DA models successfully facilitated the initial division of different PF chemotypes based on spectral fingerprints. The machine learning algorithms, including KNN, DT, and RF, achieved better discrimination than traditional statistical analyses, with the RF classifier attaining the highest accuracies of 92.0 % and 96.0 % for three types and two types of PF chemotypes, respectively. Subsequently, the two main compounds, perilla ketone and isoegomaketone, in the PK type were successfully quantified using the PLSR model based on FT-NIR and GC–MS data, with RPD values of 3.76 and 2.59, respectively. The quantitative regression curves exhibit high predictive performance. In conclusion, the present study provided a novel, rapid, and accurate method for chemotype discrimination and quantitative analysis of PF based on NIR spectral information combined with chemometrics. In addition, this approach, which differed from other evaluation perspectives, such as geographical origin and processing method, yielded satisfactory results based on the differences in PF chemotypes.

CRediT authorship contribution statement

Dai-xin Yu: Writing – review & editing, Writing – original draft, Software, Methodology, Conceptualization. **Cheng Qu:** Writing – review & editing, Supervision, Project administration, Funding acquisition. **Jia-yu Lu:** Investigation. **Di-di Wu:** Investigation. **Qi-nan Wu:** Writing – review & editing, Supervision, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that has been used is confidential.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (No. 82204597), the Research of Assurance Ability Improvement of Chinese Medicinal Resources (2023), and the Postgraduate Research & Practice Innovation Program of Jiangsu Province (KYCX24_2278).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.fochx.2024.101881.

References

Ahmed, H. M., & Tavaszi-Sarosi, S. (2019). Identification and quantification of essential oil content and composition, total polyphenols and antioxidant capacity of *Perilla frutescens* (L.) Britt. *Food Chemistry, 275*, 730–738. https://doi.org/10.1016/j.foodchem.2018.09.155

Biancolillo, A., Aloia, R., Rossi, L., & D'Archivio, A. A. (2022). Organosulfur volatile profiles in Italian red garlic (*allium Sativum* L.) varieties investigated by HS-SPME/GC-MS and chemometrics. *Food Control, 131, Article 108477*. https://doi.org/10.1016/j.foodcont.2021.108477

Biswas, A., & Chaudhari, S. R. (2024). Exploring the role of NIR spectroscopy in quantifying and verifying honey authenticity: A review. *Food Chemistry, 445*, 138712. https://doi.org/10.1016/j.foodchem.2024.138712

Carvalho Costa, D., Costa, H. S., Goncalves Albuquerque, T., Ramos, F., Castilho, M. C., & Sanches-Silva, A. (2015). Advances in phenolic compounds analysis of aromatic plants and their potential applications. *Trends in Food Science & Technology, 45*(2), 336–354. https://doi.org/10.1016/j.tifs.2015.06.009

Chen, R., Li, S., Cao, H., Xu, T., Bai, Y., Li, Z., Leng, X., & Huang, Y. (2024). Rapid quality evaluation and geographical origin recognition of ginger powder by portable NIRS in tandem with chemometrics. *Food Chemistry, 438, Article 137931*. https://doi.org/10.1016/j.foodchem.2023.137931

Cui, C., Xia, M., Wei, Z., Chen, J., Peng, C., Cai, H., Jin, L., & Hou, R. (2023). 1H NMR-based metabolomic approach combined with machine learning algorithm to distinguish the geographic origin of huajiao (*Zanthoxylum bungeanum* maxim.). *Food Control, 145, Article 109476*. https://doi.org/10.1016/j.foodcont.2022.109476

Elrasheid Tahir, H., Adam Mariod, A., Hashim, S. B. H., Arslan, M., Komla Mahunu, G., Xiaowei, H., … Xiaobo, Z. (2023). Classification of black Mahlab seeds (*Monechma ciliatum*) using GC–MS and FT-NIR and simultaneous prediction of their major volatile compounds using chemometrics. *Food Chemistry, 408, Article 134948*. https://doi.org/10.1016/j.foodchem.2022.134948

Fan, Y., Cao, X., Zhang, M., Wei, S., Zhu, Y., Ouyang, H., & He, J. (2022). Quantitative comparison and chemical profile analysis of different medicinal parts of *Perilla frutescens* (L.) Britt. From different varieties and harvest periods. *Journal of Agricultural and Food Chemistry, 70*(28), 8838–8853. https://doi.org/10.1021/acs.jafc.2c03104

Fan, Y., Bai, X., Chen, H., Yang, X., Yang, J., She, Y., & Fu, H. (2023). A novel simultaneous quantitative method for differential volatile components in herbs based on combined near-infrared and mid-infrared spectroscopy. *Food Chemistry, 407, Article 135096*. https://doi.org/10.1016/j.foodchem.2022.135096

Fu, H., Wei, L., Chen, H., Yang, X., Kang, L., Hao, Q., Zhou, L., Zhan, Z., Liu, Z., Yang, J., & Guo, L. (2021). Combining stable C, N, O, H, Sr isotope and multi-element with chemometrics for identifying the geographical origins and farming patterns of Huangjing herb. *Journal of Food Composition and Analysis, 102*. https://doi.org/10.1016/j.jfca.2021.103972. Article 103972.

Ghimire, B., Yoo, J., Yu, C., & Chung, I. (2017). GC-MS analysis of volatile compounds of *Perilla frutescens* Britton var. *Japonica* accessions: Morphological and seasonal variability. *Asian Pacific Journal of Tropical Medicine, 10*(7), 705–714. https://doi.org/10.1016/j.apjtm.2017.07.004

He, G., Yang, S., & Wang, Y. (2023). An integrated chemical characterization based on FT-NIR, and GC–MS for the comparative metabolite profiling of 3 species of the genus Amomum. *Analytica Chimica Acta*. https://doi.org/10.1016/j.aca.2023.341869. *1280*, article 341869.

Hong, X., Fu, X., Wang, Z., Zhang, L., Yu, X., & Ye, Z. (2019). Tracing geographical origins of teas based on FT-NIR spectroscopy: Introduction of model updating and imbalanced data handling approaches. *Journal of Analytical Methods in Chemistry, 2019, Article.* , Article 1537568. https://doi.org/10.1155/2019/1537568

Hou, T., Netala, V. R., Zhang, H., Xing, Y., Li, H., & Zhang, Z. (2022). *Perilla frutescens*: A rich source of pharmacological active compounds. *Molecules, 27(11), Article 3578*. https://doi.org/10.3390/molecules27113578

Jin, G., Zhu, Z., Wu, Z., Wang, F., Li, J., Raghavan, V., Li, B., & Song, C. (2023). Characterization of volatile components of microwave dried perilla leaves using

GC–MS and E-nose. *Food. Bioscience, 56, Article 103083*. https://doi.org/10.1016/j.fbio.2023.103083

Li, Y., Logan, N., Quinn, B., Hong, Y., Birse, N., Zhu, H., … Wu, D. (2024). Fingerprinting black tea: When spectroscopy meets machine learning a novel workflow for geographical origin identification. *Food Chemistry, 438, Article 138029*. https://doi.org/10.1016/j.foodchem.2023.138029

Lu, H., Tian, M., Han, X., Shi, N., Li, H., Cheng, C., Chen, W., Li, S., He, F., Duan, C., & Wang, J. (2023). Vineyard soil heterogeneity and harvest date affect volatolomics and sensory attributes of cabernet sauvignon wines on a *meso-terroir* scale. *Food Research International, 174, Article 113508*. https://doi.org/10.1016/j.foodres.2023.113508

Metz, M., Abdelghafour, F., Roger, J.-M., & Lesnoff, M. (2021). A novel robust PLS regression method inspired from boosting principles: RoBoost-PLSR. *Analytica Chimica Acta*. https://doi.org/10.1016/j.aca.2021.338823. *1179*, article 338823.

Nagy, M., Wang, S., & Farag, M. (2022). Quality analysis and authentication of nutraceuticals using near IR (NIR) spectroscopy: A comprehensive review of novel trends and applications. *Trends in Food Science and Technology, 123*, 290–309. https://doi.org/10.1016/j.tifs.2022.03.005

Peng, L., He, M., Wang, X., Guo, S., Zhang, Y., & Li, W. (2024). Fast discrimination and quantification analysis of Atractylodis rhizoma using NIR spectroscopy coupled with chemometrics tools. *Journal of Agricultural and Food Chemistry, 72*(14), 7707–7715. https://doi.org/10.1021/acs.jafc.3c08812

Rodríguez-Solana, R., Daferera, D. J., Mitsi, C., Trigas, P., Polissiou, M., & Tarantilis, P. A. (2014). Comparative chemotype determination of lamiaceae plants by means of GC-MS, FT-IR, and dispersive-Raman spectroscopic techniques and GC-FID quantification. *Industrial Crops and Products, 62*, 22–33. https://doi.org/10.1016/j.indcrop.2014.08.003

Sa, K., Jang, S., Lee, S., Park, H., Cho, J., Sung, J., & Lee, J. (2023). Characterization of volatile compounds of *Perilla* crop (*Perilla frutescens* L.) in South Korea. *Applied. Biological Chemistry, 66*(1), Article 41. https://doi.org/10.1186/s13765-023-00801-6

Sun, X., Li, H., Yi, Y., Hua, H., Guan, Y., & Chen, C. (2021). Rapid detection and quantification of adulteration in Chinese hawthorn fruits powder by near-infrared spectroscopy combined with chemometrics. *Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy, 250, Article 119346*. https://doi.org/10.1016/j.saa.2020.119346

Sun, Y., Liu, N., Kang, X., Zhao, Y., Cao, R., Ning, J., Ding, H., Sheng, X., & Zhou, D. (2021). Rapid identification of geographical origin of sea cucumbers *Apostichopus japonicus* using FT-NIR coupled with light gradient boosting machine. *Food Control, 124, Article 107883*. https://doi.org/10.1016/j.foodcont.2021.107883

Tahir, H., Mahunu, G., Arslan, M., Zhihua, L., Wen, Z., Xiaobo, Z., Mariod, A., & Jiyong, S. (2021). Feasibility study for the use of colorimetric sensor arrays, NIR and FT-IR spectroscopy in the quantitative analysis of volatile components in honey. *Microchemical Journal, 160, Article 105730*. https://doi.org/10.1016/j.microc.2020.105730

Tang, C., Jiang, B., Ejaz, I., Ameen, A., Zhang, R., Mo, X., & Wang, Z. (2023). High-throughput phenotyping of nutritional quality components in sweet potato roots by near-infrared spectroscopy and chemometrics methods. *Food Chemistry, X, 20, Article 100916*. https://doi.org/10.1016/j.fochx.2023.100916

Wang, R., Zhang, Q., Feng, C., Zhang, J., Qin, Y., & Meng, L. (2022). Advances in the pharmacological activities and effects of perilla ketone and isoegomaketone. *Evidence-Based Complementary and Alternative Medicine, 2022, Article*, 8809792. https://doi.org/10.1155/2022/8809792

Wu, X., Dong, S., Chen, H., Guo, M., Sun, Z., & Luo, H. (2023). Perilla frutescens: A traditional medicine and food homologous plant. *Chinese Herbal Medicines, 15*(3), 369–375. https://doi.org/10.1016/j.chmed.2023.03.002

Xia, H., Chen, W., Hu, D., Miao, A., Qiao, X., Qiu, G., Liang, J., Guo, W., & Ma, C. (2024). Rapid discrimination of quality grade of black tea based on near-infrared spectroscopy (NIRS), electronic nose (E-nose) and data fusion. *Food Chemistry, 440, Article 138242*. https://doi.org/10.1016/j.foodchem.2023.138242

Yu, D., Zhang, X., Guo, S., Yan, H., Wang, J., Zhou, J., Yang, J., & Duan, J. (2022). Headspace GC/MS and fast GC e-nose combined with chemometric analysis to identify the varieties and geographical origins of ginger (*Zingiber officinale* roscoe). *Food Chemistry, 396, Article 133672*. https://doi.org/10.1016/j.foodchem.2022.133672

Yu, H., Qiu, J., Ma, L., Hu, Y., Li, P., & Wan, J. (2017). Phytochemical and phytopharmacological review of *Perilla frutescens* L. (Labiatae), a traditional edible-medicinal herb in China. *Food and Chemical Toxicology, 108*, 375–391. https://doi.org/10.1016/j.fct.2016.11.023

Zhang, Y., & Wang, Y. (2023). Machine learning applications for multi-source data of edible crops: A review of current trends and future prospects. *Food Chemistry, X, 19, Article 100860*. https://doi.org/10.1016/j.fochx.2023.100860

Zhao, C., Zhang, F., Chen, S., Hu, W., Dong, L., Zhao, Y., Han, M., & Li, Z. (2023). Effects of drying methods on the quality of Hanyuan Zanthoxylum bungeanum based on physicochemical and functional metabolite analysis. *LWT-food. Science and Technology, 180*. https://doi.org/10.1016/j.lwt.2023.114674. Article 114674.

Zhao, J., Cui, P., Liu, H., Wang, C., Liu, M., Li, G., Pan, M., Li, Z., & Suo, T. (2020). Rapid screening and quantitative analysis of adulterant Lonicerae Flos in Lonicerae Japonicae Flos by Fourier-transform near infrared spectroscopy. *Infrared Physics & Technology, 104, Article 103139*. https://doi.org/10.1016/j.infrared.2019.103139

Zhao, L., Zhao, W., Zhao, Z., Xian, R., Jia, M., Jiang, Y., Li, Z., Pan, X., Lan, Z., & Li, M. (2024). Rapid discrimination of Alismatis Rhizoma and quantitative analysis of triterpenoids based on near-infrared spectroscopy. *Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy, 321, Article 124618*. https://doi.org/10.1016/j.saa.2024.124618

Zheng, C., Li, J., Liu, H., & Wang, Y. (2023). Data fusion of FT-NIR and ATR-FTIR spectra for accurate authentication of geographical indications for Gastrodia elata Blume. *Food. Bioscience, 56, Article 103308*. https://doi.org/10.1016/j.fbio.2023.103308

Zhou, G., Yuan, Y., Yin, Y., Tang, Y., Xu, R., Liu, Y., Chen, P., Yin, L., & Duan, J. (2020). Hydrophilic interaction chromatography combined with ultrasound-assisted ionic liquid dispersive liquid-liquid microextraction for determination of underivatized neurotransmitters in dementia patients' urine samples. *Analytica Chimica Acta, 1107*, 74–84. https://doi.org/10.1016/j.aca.2020.02.027

Zhou, P., Shao, Y., Jiang, Z., Dang, J., Qu, C., & Wu, Q. (2023). The revealing of a novel double bond reductase related to perilla ketone biosynthesis in Perilla frutescens. *BMC Plant Biology, 23*(1), 1–12. https://doi.org/10.1186/s12870-023-04345-1

Zhou, P., Yin, M., Dai, S., Bao, K., Song, C., Liu, C., & Wu, Q. (2021). Multi-omics analysis of the bioactive constituents biosynthesis of glandular trichome in Perilla frutescens. *BMC Plant Biology, 21*(1), 1–15. https://doi.org/10.1186/s12870-021-03069-4

Zou, Z., Guo, B., Guo, Y., Ma, X., Luo, S., Feng, L., Pan, Z., Deng, L., Pan, S., Wei, J., & Su, Z. (2024). A comprehensive "quality-quantity-activity" approach based on portable near-infrared spectrometer and membership function analysis to systematically evaluate spice quality: Cinnamomum cassia as an example. *Food Chemistry, 439, Article 138142*. https://doi.org/10.1016/j.foodchem.2023.138142