

Article

Discrimination and Prediction of *Lonicerae japonicae Flos* and *Lonicerae Flos* and Their Related Prescriptions by Attenuated Total Reflectance Fourier Transform Infrared Spectroscopy Combined with Multivariate Statistical Analysis

Yang-Qiannan Tang , Li Li, Tian-Feng Lin, Li-Mei Lin, Ya-Mei Li and Bo-Hou Xia * 

Key Laboratory for Quality Evaluation of Bulk Herbs of Hunan Province, Hunan University of Chinese Medicine, Changsha 410208, China; tangyangqiannan@163.com (Y.-Q.T.); licorece@hotmail.com (L.L.); m13460513070@163.com (T.-F.L.); lizasmile@163.com (L.-M.L.); yameili@hnu.cm.edu.cn (Y.-M.L.)

* Correspondence: xiabohou@163.com

Abstract: LJF and LF are commonly used in Chinese patent drugs. In the *Chinese Pharmacopoeia*, LJF and LF once belonged to the same source. However, since 2005, the two species have been listed separately. Therefore, they are often misused, and medicinal materials are indiscriminately put in their related prescriptions in China. In this work, firstly, we established a model for discriminating LJF and LF using ATR-FTIR combined with multivariate statistical analysis. The spectra data were further preprocessed and combined with spectral filter transformations and normalization methods. These pretreated data were used to establish pattern recognition models with PLS-DA, RF, and SVM. Results demonstrated that the RF model was the optimal model, and the overall classification accuracy for LJF and LF samples reached 98.86%. Then, the established model was applied in the discrimination of their related prescriptions. Interestingly, the results show good accuracy and applicability. The RF model for discriminating the related prescriptions containing LJF or LF had an accuracy of 100%. Our results suggest that this method is a rapid and effective tool for the successful discrimination of LJF and LF and their related prescriptions.

Keywords: *Lonicerae japonicae Flos*; *Lonicerae Flos*; ATR-FTIR; multivariate statistical analysis



Citation: Tang, Y.-Q.; Li, L.; Lin, T.-F.; Lin, L.-M.; Li, Y.-M.; Xia, B.-H. Discrimination and Prediction of *Lonicerae japonicae Flos* and *Lonicerae Flos* and Their Related Prescriptions by Attenuated Total Reflectance Fourier Transform Infrared Spectroscopy Combined with Multivariate Statistical Analysis. *Molecules* **2022**, *27*, 4640. <https://doi.org/10.3390/molecules27144640>

Academic Editor: Lukasz Komsta

Received: 15 June 2022

Accepted: 14 July 2022

Published: 20 July 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

LJF, widely used in common Chinese medicine, is the flower bud of *Lonicera japonica* Thunb. mainly produced in Shandong, Henan, and Hebei Provinces in China [1]. It is commonly applied in the treatments of sores, furuncles, carbuncles, swelling, and infections caused by exopathogenic wind-heat or epidemic febrile diseases [1]. At the same time, there is another Chinese medicine named LF, which is defined by the *Chinese Pharmacopoeia* as the dried flower bud of *Lonicera macranthoides* Hand.-Mazz., *Lonicera fulvotomentosa* Hsu et S. C. Cheng, *Lonicera hypoglauca* Miq., and *Lonicera confusa* DC. These four LF species are mainly grown in the south of the Yangtze River, including the Hunan, Guizhou, Guangxi, and Guangdong Provinces in China [1]. LF possesses a significant antipyretic effect, improves liver functions, and exhibits an antibacterial effect [2]. Although LJF and LF are widely used herbs with similar phenotypes derived from the plants of the same genus, LJF and LF are listed as independent items in the *Chinese Pharmacopoeia* (2020 Edition) for medical safety, especially regarding drug injection. *Chinese Pharmacopoeia* (2005 Edition) lists LJF and LF as independent items because LF and LJF are significantly different in medicinal history, plant morphology, medicinal properties, and chemical constituent, and the only plant source of LJF is again limited to *Lonicera japonica* Thunb. In contrast, LF has three plant sources, including *Lonicera macranthoides* Hand.-Mazz., *Lonicera hypoglauca* Miq., and *Lonicera confusa* DC. Moreover, *Chinese Pharmacopoeia* (2010 Edition) adds another plant

source named *Lonicera fulvotomentosa* Hsu et S. C. Cheng to LF following the 2005 Edition, obtaining a total of four plants of the same genus under the legal species of LF.

According to traditional records, LJF and LF can clear heat and detoxify. Thus, they are often applied for the treatment of sores, furuncles, carbuncles, and issues caused by exopathogenic wind-heat or epidemic febrile diseases. Based on modern pharmacological studies on LJF and LF, there are some different medicinal effects between LJF and LF. Compared with other commonly seen antibacterial drugs, LJF manifests more powerfully in antibacterial activity and inhibition of drug-resistant bacteria [2]. LJF has glucose-lowering, anti-ultraviolet radiation, anti-endotoxin, anti-ulcer, anti-early pregnancy, anti-platelet aggregation, anti-fertility, and neuroprotective activities that are not reported in LF, while LF has effects on balancing intestinal flora and anti-atherosclerotic effects that are not reported in LJF [3]. In recent years, the commercial value of LJF in herbal medicine trading markets has increased by over 400%, and more than 30% of current traditional Chinese medicine prescriptions contain LJF [4]. As a result, LF is often misused as LJF by some pharmaceutical factories to obtain higher interest. Furthermore, due to the difference in pharmacological activity between LJF and LF, the related prescriptions of both have different pharmacological effects. The incorrect addition of LJF and LF to their related prescriptions could cause great damage to human health and industrial development. Therefore, it is essential to develop a rapid identification method to discriminate between LJF, LF, and their related prescriptions to ensure the safety of patients and the healthy development of *Lonicera*-based trade markets.

Previous studies on the identification of LJF and LF have mainly focused on morphology and anatomical characters [5], chemical analysis [4], and DNA molecular marker techniques [6]. According to research, liquid chromatography and mass spectrometry were used to identify *Lonicera* species flower buds [7,8]. Although these methods have some advantages in the identification of *Lonicera* origins, they are not suitable for online quality control and are time-consuming. The sample pretreatment and detection processes of LC-MS and NMR are complicated, and the absorption bands of FT-NIRs are relatively broad, and the absorption bands are seriously overlapped, which means they cannot be directly performed in qualitative analysis. Recently, the potential of the hyperspectral imaging method is applied for the rapid identification of true and false honeysuckle tea leaves [9]. They are mainly concerned with the study of LJF and LF as tea leaves. Tea belongs to the field of food. However, the LJF- and LF-related pharmaceutical industry is a big business in China. The adulteration problem in the pharmaceutical field is far more important than that in the food field, and the harm is great. Therefore, it is urgent to supplement and modify some research gaps in this field.

ATR-FTIR technology has been developed based on FTIR technology, which can be equipped with ATR accessories for crystal materials [10], and is a fast, convenient, non-destructive, and high-throughput analytical tool that is widely used in the rapid identification and quality control of herbal medicines [11]. However, the obtained FT-IR spectra can be affected by the characteristics of samples in thickness and particle size [12,13], so the acquired raw data should be normalized. Compared with other kinds of infrared spectroscopy, ATR-FTIR is more convenient, simple, and effective. The use of this high-performance ATR accessory does not need time-consuming sample preparation and specialized sample holders such as the KBr pellet method [7,14]. In addition, the detected samples stay in their natural state, whose spectrum could reflect the original chemical information of various metabolites. ATR-FTIR also has excellent sample-to-sample reproducibility and minimal operator-induced variations [8]. In recent years, ATR-FTIR spectra combined with multivariate statistical analysis have been widely applied for quality assessment and authentication of herbal medicines [15–17]. However, the original ATR-FTIR spectra contained some overlapped absorbance and extra absorbance caused by different particle sizes and thickness of sample powders, so it was necessary for them to be preprocessed by spectral filter transformations and normalization methods.

In this work, a comprehensive study of the differences between LJF, LF, and their related prescriptions was proposed based on ATR-FTIR. The spectral process methods (spectral filter transformations and normalization methods), different pattern recognition methods, model parameters, and variable selection methods were optimized to form the optimal distinguishing prediction model. Spectral filter transformations include MSC, SNV, Savitzky–Golay, Row Center, and EWMA. Three normalization methods are applied to ATR-FTIR spectral data: area normalization, min-max normalization, and vector normalization. To construct a more precise discrimination model, we selected an optimal pattern recognition method from RF, SVM, and PLS-DA.

2. Materials and Methods

2.1. Materials and Sample Preparation

LJF and LF dried bud samples recorded in the *Chinese Pharmacopeia* (2020 Edition) were collected from the main *Lonicera*-producing areas in Hunan (*Lonicera macranthoides* Hand.-Mazz.) and Shandong in China, with all samples identified by Prof. Li-Min Gong from the Hunan University of Chinese medicine. Chinese patent medicines containing LJF or LF were purchased from the authentication-passed pharmacies. The ATR-FTIR, LJF, and LF dried buds samples were dried at 60 °C [18] to constant weight (the weight variation was less than 0.1%) in an electric thermostatic drying oven to ensure that moisture was not an interfering factor. Samples were then ground into powders; each sample was finely powdered by agate mortar and screened with a 200-mesh stainless steel sieve. The Chinese patent medicines samples were also ground into powders and then screened with a 200-mesh stainless steel sieve. All samples were stored in a relatively dry environment.

2.2. Spectral Acquisition and Data Preprocessing

Each powder was subjected to an FTIR spectrometer (Nicolet iS5, Thermo Scientific, Waltham, MA, USA) equipped with an ATR accessory for recording the FTIR spectrum. The OMNIC program (version 8.2.0.387, Thermo Scientific, Waltham, MA, USA) was used to obtain all ATR-FTIR spectra. In total, 64 scans were recorded to obtain average analytical results and improve the signal-to-noise ratio. The transmittance of each spectrum was collected between 4000 and 600 cm^{-1} with a spectral resolution of 4 cm^{-1} .

ATR-FTIR spectral filter transformations, including MSC, SNV, Savitzky–Golay, Row Center, and EWMA normalization, were applied to minimize the baseline noise and maximize the differences found in spectra. The normalization methods were area normalization, minimum-maximum normalization, and vector normalization, which could reduce the effect of the physical characteristics of samples (particle size and thickness). In vector normalization, all spectra were converted from transmittance to absorbance, and the ATR-FTIR absorbance spectra were converted into the first and second derivatives with Savitzky–Golay derivative and nine smoothing points in OMNIC software. For vector normalization, the Euclidean norm was performed to calculate the absorbance values to acquire the normalization values of the spectra. In area and min-max normalizations, all spectra were converted from transmittance to absorbance, and then ATR correction was conducted by using OMNIC. For area normalization, each absorbance value at a specific wave number was divided by the total (integrated) absorbance area of the spectrum. For min-max normalization, each absorbance value was divided by the difference between the highest and lowest absorbance values. In addition, for the baseline correction, the airPLS algorithm was used to remove the baseline for all the spectra. The airPLS was proved to be a mature and effective algorithm [19].

2.3. Chemometrics Methods

2.3.1. Random Forest (RF)

RF was developed by Breiman in 2001 and has been widely used to resolve classification problems or regression issues as a non-parametric algorithm based on a learning strategy (called ensemble learning) [20]. The RF model consists of hundreds of trees, and

each tree was grown through a bootstrap sample of the original data. In addition, each node was selected by each tree and corresponded to a random variable subset. Each tree could provide a classification result to decide the final accurate category. The operational process of the RF model could be separated into the following steps.

Firstly, a spectra dataset was divided into the calibration set (bootstrap samples) and the validation set (OOB samples) with the help of the KS algorithm by MATLAB 2017a (MathWorks, Natick, MA, USA). Then, 4/5 of all LJF and LF samples were included in the calibration set, which was applied to obtain the optimal classification trees. The validation set was used to evaluate the ability of the RF model at last. Secondly, the values of n_{tree} and the square root of the number of all variables m_{try} based on the calibration set were both selected to acquire the optimal n_{tree} and m_{try} . The OOB error was the rate of misclassification over all out-of-bag samples. The best n_{tree} was acquired based on the lowest OOB error, which was beneficial for further prediction. Thirdly, a new importance variable matrix was formed after exerting the optimal n_{tree} and m_{try} . Fourthly, the new matrix was inserted in the producer, the original spectra variables were rearranged according to variable importance, and the most important variables were selected by a lower 5-fold cross-validation error rate. Finally, the validation set was inserted in the random forest model established based on the calibration set for the final model prediction. The establishment of the final RF discrimination model was performed by using the optimized n_{tree} and m_{try} parameters. The first two steps were rerun to calculate the final classification accuracy. Steps two to five were running in MATLAB.

2.3.2. Support Vector Machine (SVM)

SVM is an effective classification method proposed by Vapnik, which is based on SLT and the principle of SRM [21]. SVM, as one of the kernel-based pattern recognition methods, has been successfully applied in the classification of drug and nondrug problems [22].

SVM can obtain nonlinear and global solutions even with the high-dimensional input vector [23]. An optimal classifier generalization is acquired when it minimizes training error along with higher testing accuracy for unknown testing datasets. The training algorithm of SVM maximizes the margin between class boundary and the training data by removing some meaningless data from the training dataset. So, the resulting decision function only depends on the training data called support vectors. Therefore, SVM maximizes the boundary by minimizing the maximum loss and gives good accuracy [24]. Experiments are performed in MATLAB using LIBSVM.

A kernel is a key that determines the performance of the SVM. The radial basis function (RBF) kernel was selected as the kernel function because of the excellent classification performance shown in previous studies [25,26]. The training data consist of the input matrix x_i ($i = 1, 2, \dots, n$) and an output vector y_i ($i = 1, 2, \dots, n$), where +1 and -1 are used to stand for the two classes. The SVM constructed an optimized linear regression by mapping the input vector x to a high-dimensional feature space by a nonlinear mapping using a kernel function $K(x_i, y_j)$. The RBF kernel is expressed as $K(x_i, y_j) = \exp(-\gamma \|x_i - y_j\|^2)$, where γ is the width parameter of the RBF kernel function [27]. There are several methods to the SVM to deal with classification problems, including “One-Versus-Rest (OVR)” [28], “One-Versus-One (OVO)” [29], and DAGSVM [30]. In this paper, we chose the OVO strategy, which can scale well to a large number of classes.

2.3.3. Partial Least Squares Discrimination Analysis (PLS-DA)

PLS-DA, a binary classification algorithm from 0 to 1, is an adaptation of the PLS regression algorithm to the problem of supervised clustering. It has been extensively used in the analysis of multivariate datasets between independent and dependent variables, which are expressed by X and Y , respectively [31]. For PLS-DA, outliers were first eliminated by PCA combined with Mahalanobis distance. PCA A PLS-DA model was created on the training dataset using the preprocessed data. Experiments are performed in MATLAB.

2.4. Data Analysis

RF, SVM, and PLS-DA models can obtain the vote matrices. The three models in MATLAB could calculate the values of TN, TP, FN, and FP, respectively. Sensitivity, SENS (Equation (1)), Specificity, SPEC (Equation (2)), Accuracy, ACC (Equation (3)), and Matthew's correlation coefficient, MCC (Equation (4)) were the four parameters for each class, indicating the identification effects for different samples of the three models. These four parameters with higher values indicate better identification ability for each class. The MCC is a correlation coefficient applied to evaluate the performance of binary classifications. It ranges from ± 1 to 0; +1 indicates a perfect identification while 0 shows the performance of a random classification. Only binary PLS-DA classification models were calculated by the MCC [32].

$$\text{SENS} = \frac{\text{TP}}{(\text{TP} + \text{FN})} \quad (1)$$

$$\text{SPEC} = \frac{\text{TN}}{(\text{TN} + \text{FP})} \quad (2)$$

$$\text{ACC} = \frac{(\text{TN} + \text{TP})}{(\text{TP} + \text{TN} + \text{FP} + \text{FN})} \quad (3)$$

$$\text{MCC} = \frac{(\text{TP} \times \text{TN} - \text{FP} \times \text{FN})}{\sqrt{(\text{TP} + \text{FP})(\text{TP} + \text{FN})(\text{TN} + \text{FP})(\text{TN} + \text{FN})}} \quad (4)$$

3. Results and Discussion

3.1. Band Assignment Comparison between LJF and LF

Spectral data were obtained in the range of 4000–600 cm^{-1} . The representative samples of LJF and LF spectra are shown in Figure 1. The spectra were analyzed, and several noticeable peaks were observed. Assignments of the main absorption peaks are also summarized in Table 1.

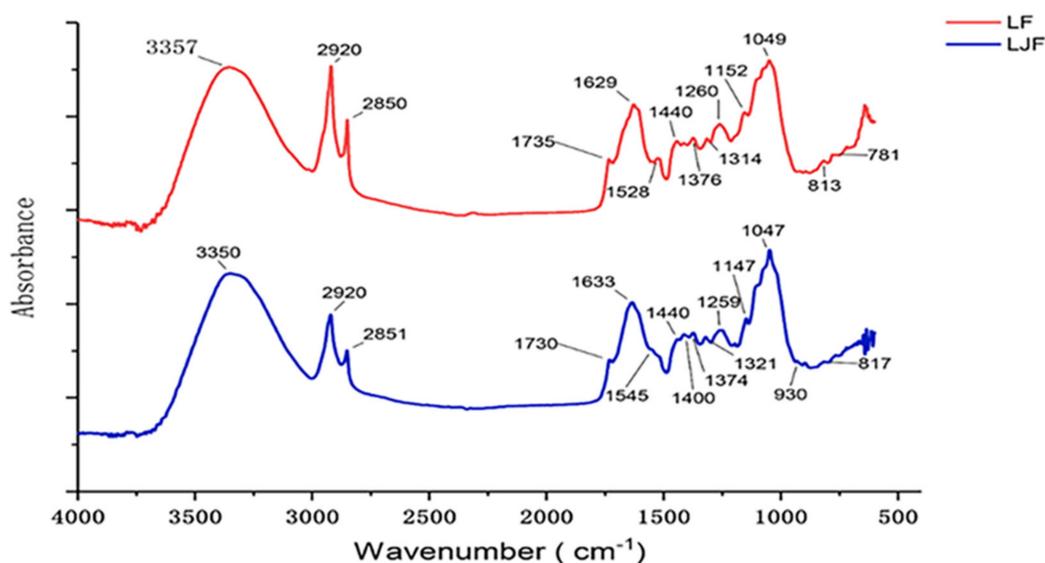


Figure 1. ATR-FTIR spectra of representative LJF and LF.

Table 1. Peak assignments of the ATR-FTIR spectra of LJF and LF.

LJF (cm ⁻¹)	LF (cm ⁻¹)	Vibration	Suggested Biomolecular Assignment
4000–3500	4000–3500	O-H, ν	water
3350	3357	O-H, ν	saccharides
2920	2920	CH ₂ , CH ₃ ν_{as}	lipids (cutin and waxes), proteins, carbohydrates
2851	2850	CH ₂ , ν_s	lipids (cutin and waxes), proteins, carbohydrates
2442–2208	2442–2208	C-O-C, ν	CO ₂
1730	1735	C=O, ν	lipids (cutin and waxes)
1633	1629	C-O, ν C-N, ν	amide I band
1545		Amide II bands	proteins
	1528	Amide II bands	phenolic acids, flavonoids
14,401,374	14,401,374	O-H, ν	organic acid, flavonoids
		O-H, ν	organic acid, flavonoids
1400		C-H, δ	saccharides
1321	1314	C-O, ν	lipids, flavonoid
1259	1260	C-O, ν	lipids, flavonoid
1147	1152	C-O, ν CO-O-C, ν_{as}	cholesterol ester, oligosaccharides, triacylglycerols
1047	1049	C-O, ν	starch
930		C-O-C, skeletal	saccharides
817	813	C-H, δ_{oop}	
	781	COO-, skeletal	saponins

ν —stretching, ν_s —symmetrical stretching, ν_{as} —asymmetrical stretching, δ —bending, δ_{oop} —bending out of the plane.

The tiny band between 4000 and 3500 cm⁻¹ corresponded to water-vapor O-H stretching, and the band was attributed to the O-C-O stretching of carbon dioxide at 2442–2208 cm⁻¹ [33]. The stretching band of O-H near 3350 cm⁻¹ and the stretching bands of C-O in the region of 1200–950 cm⁻¹ indicated the existence of saccharides in LJF and LF [34]. LJF and LF spectra showed different patterns in this region, which meant the saccharides were different in LJF and LF. The peaks at 2920 cm⁻¹ and 2851 cm⁻¹ are attributed to the asymmetrical and symmetrical stretching bands of CH₂- [35], in conjunction with the stretching band of C=O near 1729 cm⁻¹ [36], indicating the existence of lipids. LJF and LF are covered by cuticles [1,37], which are continuous lipid membranes including cutin, waxes, cutan, and polysaccharides [38,39]. The absorption peaks near 2920, 2851, and 1729 cm⁻¹ were from cutin and waxes [40], which consisted of acids, alcohols, esters, alkanes, etc. The peaks near 1630 cm⁻¹ were due to C-O and C-N protein stretching [41]. This is known as the amide I band and is the main amide band. The peaks at 1440 and 1374 cm⁻¹ were attributed to organic acid OH vibrational modes [42]. The peaks at 1321 and 1259 cm⁻¹ were due to C-O stretching vibrations. The peak near 1150 cm⁻¹ was assigned to CO-O-C asymmetric stretching of cholesterol ester and C-O stretching of oligosaccharides and triacylglycerols [43,44]. One was at 1051 cm⁻¹ due to C-O stretching of starch [45].

Moreover, LJF showed the amide II bands at 1545 cm⁻¹ while LF showed the aromatic skeletal bands near 1528 cm⁻¹. This indicated that LJF contains more proteins, but LF contains more aromatic compounds (phenolic acids, flavonoids, etc.) [46,47]. In addition, LJF showed a peak at 1400 cm⁻¹ and a weak peak at 930 cm⁻¹, which were both absent in LF. The peak at 1400 cm⁻¹ could be assigned to the bending mode of O-C-H, while the peak at 930 cm⁻¹ could correspond to the skeletal mode of saccharides [48]. These two peaks indicated the different saccharides in LJF and LF. Another discriminating peak was near 780 cm⁻¹, which was present in LF but not in LJF. This difference may be caused by the high content of saponins in LF. In summary, LJF and LF can be distinguished by the four peaks near 1545, 1400, 930, and 780 cm⁻¹.

3.2. Classification of LJF and LF

In this study, RF, SVM, and PLS-DA models were applied in the classification and prediction of LJF and LF samples in a MATLAB programming environment. Spectra datasets were divided into the calibration set and the validation set with the help of the KS algorithm. The calibration set was 4/5 of all LJF and LF samples. In the training process of

these three models, 5-fold cross-validation was used to demonstrate the performance of our methods.

As for establishing the RF model, the first step was to select the number of trees for the optimal classification trees model by the training process. Before this step, the initial raw dataset and normalization datasets were calculated with 300 trees to choose which dataset was better to establish the RF model. SVM and PLS-DA models were also established based on the raw dataset and normalization datasets. As shown in Table 2, RF models showed the performance by applying suitable vector normalization after the first differentiation in the calibration set. The sensitivity, specificity, accuracy, and MCC by the RF method reached 0.9706, 1, 0.9844, and 0.9692, respectively, which outperformed the other pretreatment methods. In terms of accuracies, the RF model also outperformed PLS-DA and SVM models in terms of overall accuracies. When compared with classifiers, the highest accuracy of RF was 0.0071 higher than that of PLS-DA and 0.0071 higher than that of SVM, respectively. Moreover, the validation sets for the assessment of the three models' performance are shown in Table 3, so it can be seen that the RF model with the first derivative vector normalization was the optimal prediction model for discriminating LJF and LF, having accuracy of 0.9744.

Table 2. Comparing the performance of RF, PLS-DA, and SVM models according to various normalization and spectral filter transformations between LJF and LF samples.

RF Model					
Pretreatment Methods	SENS	SPEC	ACC	MCC	AUC
No methods	0.9167	0.8250	0.8750	0.7481	0.9190
Vector (first)	0.9706	1	0.9844	0.9692	0.9710
Vector (second)	0.9583	1	0.9773	0.9554	0.9630
Min-max	0.9375	0.9750	0.9545	0.9097	0.9368
Area	0.9500	0.9375	0.9432	0.8859	0.9491
EWMA	0.9167	0.9000	0.9091	0.8167	0.9090
MSC	0.9500	0.9583	0.9545	0.9083	0.9310
RC	0.9500	0.9583	0.9545	0.9083	0.9430
S-G	0.9250	0.9167	0.9205	0.8401	0.9168
SNV	0.9500	0.9375	0.9432	0.8859	0.9291
airPLS	0.9750	0.9792	0.9773	0.9542	0.9690
PLS-DA Model					
Pretreatment methods	SENS	SPEC	ACC	MCC	AUC
No methods	0.9412	1	0.9687	0.9393	0.9229
Vector (first)	0.9750	0.9792	0.9773	0.9542	0.9710
Vector (second)	0.9500	0.9583	0.9545	0.9083	0.9630
Min-max	0.9333	0.9667	0.9687	0.9389	0.9218
Area	0.9118	0.9706	0.9531	0.9104	0.9091
EWMA	0.9412	1	0.9687	0.9393	0.9290
MSC	0.9706	0.9667	0.9687	0.9373	0.9610
RC	0.9750	0.9792	0.9773	0.9542	0.9430
S-G	0.9412	1	0.9687	0.9393	0.9268
SNV	0.9667	0.9706	0.9687	0.9373	0.9491
airPLS	0.9750	0.9792	0.9773	0.9542	0.9690
SVM Model					
Pretreatment methods	SENS	SPEC	ACC	MCC	AUC
No methods	0.9500	0.9792	0.9659	0.9314	0.9390
Vector (first)	0.9792	0.9981	0.9716	0.9724	0.9710
Vector (second)	0.9750	0.9792	0.9773	0.9542	0.9630
Min-max	0.9750	0.9792	0.9773	0.9542	0.9668
Area	0.7250	0.5208	0.6136	0.2490	0.1291
EWMA	0.9500	0.9792	0.9659	0.9314	0.9290
MSC	0.9250	0.9792	0.9545	0.9089	0.9010
RC	0.9500	0.9583	0.9545	0.9083	0.9230
S-G	0.9500	0.9792	0.9659	0.9314	0.9168
SNV	0.9750	0.9792	0.9773	0.9542	0.9491
airPLS	0.9500	0.9792	0.9659	0.9314	0.9390

RF: random forest; PLS-DA: partial least squares-linear discriminant analysis; SVM: support vector machine regression; EWMA: exponentially weighted moving average; MSC: multiplicative scatter correction; RC: row center; S-G: Savitzky–Golay; SNV: standard normal variate; airPLS: adaptive iteratively reweighted penalized least squares.

Table 3. The classification results and evaluation parameters between LJF and LF combined with RF, PLS-DA, and SVM models by vector normalization applied after the first differentiation.

Calibration Set	SENS	SPEC	ACC	MCC	AUC
RF	0.9706	1	0.9844	0.9692	0.9775
PLS-DA	0.9750	0.9792	0.9773	0.9542	0.9546
SVM	0.9792	0.9981	0.9716	0.9724	0.9668
Validation set	SENS	SPEC	ACC	MCC	AUC
RF	0.9706	0.9981	0.9744	0.9592	0.9425
PLS-DA	0.9250	0.9792	0.9545	0.9089	0.9006
SVM	0.9500	0.9792	0.9659	0.9314	0.9218

Furthermore, the number of trees, the wave-number regions, and the variable importance cutoff value were employed to acquire the best prediction RF model. The optimal number of trees was obtained according to the lowest OOB classification error value. As shown in Table 4, these trees with the lowest error were in the numbers 100, 200, 300, 500, 800, and 1000. The number of grown trees with the highest accuracy was 300. Then, 300 trees were chosen for branch nodes selection (the m_{try} value). Using these 300 trees, the result of nodes calculation was acquired in Table 4. In total, 86 branch nodes were selected after inputting the optimal number of trees with the lowest OOB error and the highest accuracy of 0.9886 in the calibration set. The results indicated that 300 trees and 86 branch nodes were used for further training and prediction of the model.

Table 4. The parameter screening in the RF model for variables is ranked by permutation accuracy importance.

n_{tree}	SENS	SPEC	ACC	MCC	AUC
100	0.9750	0.9792	0.9773	0.9554	0.9390
200	0.9286	1	0.9583	0.9188	0.9010
300	0.9706	1	0.9844	0.9692	0.9775
500	0.9583	0.9750	0.9659	0.9316	0.9168
800	0.9583	0.9750	0.9659	0.9316	0.9168
1000	0.9286	1	0.9583	0.9188	0.9018
m_{try}	SENS	SPEC	ACC	MCC	AUC
82	0.9583	1	0.9773	0.9554	0.9390
84	0.9583	0.9750	0.9659	0.9316	0.9168
86	0.9792	1	0.9886	0.9774	0.9875
88	0.9706	1	0.9844	0.9692	0.9775
90	0.9583	0.9750	0.9659	0.9316	0.9168
92	0.9583	1	0.9773	0.9554	0.9390
94	0.9792	0.9750	0.9773	0.9542	0.9390
96	0.9583	0.9750	0.9659	0.9316	0.9168

Since ATR-FTIR spectra may be influenced by environmental factors including water vapor and carbon dioxide, RF models established by utilizing different wave-number ranges were compared to identify the best prediction model. As presented in Table 5, the ATR-FTIR spectral region between 4000 and 600 cm^{-1} was the best prediction model for discriminating LJF and LF since it had the highest accuracy ($=0.9844$). Furthermore, the original spectra variable was rearranged according to cutoff values. As listed in Table 6, the RF model had the higher four parameters data with a VIP cutoff value of 0.01, having the highest prediction accuracy of 0.9886. Thus, we could conclude that the RF method was selected as an appropriate classifier in this study with 300 trees, 86 branch nodes, 4000–600 cm^{-1} wavenumber area, and the VIP cutoff value of 0.01 under the first derivative vector normalization.

Table 5. List of permutation parameters of the random forest model obtained using variables selected by vector normalization applied after the first differentiation.

Normalization Method	SENS	SPEC	ACC	MCC	AUC
Vector (First)					
4000–600 cm ⁻¹ except for water vapor, carbon dioxide region	0.9750	0.9792	0.9773	0.9542	0.9425
2000–600 cm ⁻¹	0.9792	0.9750	0.9773	0.9542	0.9390
4000–2000 cm ⁻¹	0.9583	1	0.9773	0.9554	0.9390
4000–600 cm ⁻¹	0.9706	1	0.9844	0.9692	0.9775

Table 6. Various VIP cutoff values using 4000–600 cm⁻¹ wavenumber areas for the comparison of LJF and LF.

VIP Cutoff	SENS	SPEC	ACC	MCC	AUC
0.05	0.9412	1	0.9688	0.9393	0.9168
0.01	0.9750	1	0.9886	0.9773	0.9775
0.015	0.9512	0.9867	0.9731	0.9465	0.9425
0.020	0.9000	0.9706	0.9375	0.8758	0.8625

3.3. Classification of LJF- and LF-Related Prescriptions

LJF- and LF-related prescriptions are mainly composed of LJF and LF, respectively. Under the guidance of TCM theory, varied compound prescriptions, containing different herbs, are formulated based on the principle of “Emperor, Minister, Adjuvant, Courier”, a metaphor suggesting that different herbs have different effects and that the TCM practitioner should prescribe them according to the specific status of a patient. Due to the difference in ingredients, compound prescriptions differ in curative effects and side effects. Therefore, we can apply the methods of classification and prediction of LJF and LF in identifying LJF- and LF-related prescriptions. As mentioned above, we believe that the optimal prediction RF model of discriminating LJF and LF can show the appropriate performance in the identification of LJF- and LF-related prescriptions. After calculating LJF and LF traditional medicine ATR-FTIR spectra datasets by the RF model, four evaluation parameters (SENS, SPEC, ACC, and MCC) were all 1, which suggested that the RF model was an effective classifier for prescriptions. Thus, the present results indicate that the established RF model is the optimal predictable one to discriminate LJF and LF and their related prescriptions.

4. Conclusions

In this study, LJF and LF and their related prescriptions were analyzed by ATR-FTIR spectra combined with multivariate classification methods. The chemical information differences between LJF and LF were revealed in the light of band assignments based on the ATR-FTIR spectra. This is the first study that established classification models for the identification of LJF and LF using various factors, including spectral filter transformations, normalization methods, VIP cutoff, and wave-number region. The feasible results indicated that ATR-FTIR combined with RF is a rapid, efficient, reliable, and stable online classification and prediction method for LJF and LF and their related prescriptions, which has wide adaptability and can be applied in the application of Chinese medicine.

Author Contributions: Y.-Q.T. and B.-H.X. designed the experiment. Y.-Q.T. and T.-F.L. conducted the experiment and analyzed the data. L.L., T.-F.L., Y.-M.L. and B.-H.X. interpreted the results. Y.-Q.T., L.-M.L. and B.-H.X. drafted and edited the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This study was funded by The Science and Technology Innovation Program of Hunan Province (2021RC4064), Science and Technology Project of Changsha (kq2004054), Natural Science Foundation of Hunan Province (2021JJ40551, 2022JJ50008), and First-class Discipline Project on Chinese Pharmacology of Hunan University of Chinese Medicine (201803).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: We thank Jianhua Huang for his support in data analysis.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

LJF	<i>Lonicerae japonicae</i> Flos
LF	<i>Lonicerae</i> Flos
ATR-FTIR	attenuated total reflectance Fourier transform spectroscopy
PLS-DA	partial least squares-linear discriminant analysis
RF	random forest
SVM	support vector machine
FT-IR	Fourier-transform infrared
LC-MS	liquid chromatography-mass spectrometer
NMR	nuclear magnetic resonance spectroscopy
FT-NIRs	Fourier-transform near infrared spectroscopy
EWMA	exponentially weighted moving average
min-max	minimum-maximum
ATR	attenuated total reflection
airPLS	adaptive iteratively reweighted penalized least squares
MSC	multiplicative scatter correction
SNV	standard normal variate
OOB	out-of-bag
KS	Kennard–Stone
SLT	statistical learning theory
SRM	structural risk minimization
PCA	principal component analysis
TN	true negative
TP	true positive
FN	false negative
FP	false positive
TCM	traditional Chinese medicine

References

1. Chinese Pharmacopoeia Commission. *Pharmacopoeia of the People's Republic of China*; China Medical Science Press: Beijing, China, 2020; Volume 1, pp. 31–32, 229–230.
2. Lin, H.W.; Lee, Y.J.; Yang, D.J.; Hsieh, M.C.; Chen, C.C.; Hsu, W.L.; Chang, Y.Y.; Liu, C.W. Anti-inflammatory effects of *Flos Lonicerae Japonicae* Water Extract are regulated by the STAT/NF- κ B pathway and HO-1 expression in Virus-infected RAW264.7 cells. *Int. J. Med. Sci.* **2021**, *18*, 2285. [[CrossRef](#)] [[PubMed](#)]
3. Li, Y.; Li, W.; Fu, C.; Li, Y.; Li, W.; Fu, C.; Song, Y.; Fu, Q. *Lonicerae japonicae* flos and *Lonicerae* flos: A systematic review of ethnopharmacology, phytochemistry and pharmacology. *Phytochem. Rev.* **2020**, *19*, 1–61. [[CrossRef](#)] [[PubMed](#)]
4. Tang, Y.; Zeng, T.; Zafar, S.; Yuan, H.; Li, B.; Peng, C.; Wang, S.; Jian, Y.; Qin, Y.; Choudhary, M.I.; et al. *Lonicerae* Flos: A review of chemical constituents and biological activities. *Digit. Chin. Med.* **2018**, *1*, 173–188. [[CrossRef](#)]
5. Liu, S.; Zhou, L.; Huang, J.; Zeng, H.; Qiao, Z.; Li, Y.; Zhang, G. Comparative Analysis of the Complete Chloroplast Genome Sequences of Four Origin Plants of *Lonicerae* Flos (*Lonicera*; *Caprifoliaceae*). *Phyton-Int. J. Exp. Bot.* **2022**, *91*, 1503–1516. [[CrossRef](#)]
6. Yu, J.; Wu, X.; Liu, C.; Newmaster, S.; Ragupathy, S.; Kress, W.J. Progress in the use of DNA barcodes in the identification and classification of medicinal plants. *Ecotoxicol. Environ. Saf.* **2021**, *208*, 111691. [[CrossRef](#)]
7. Zhang, F.; Shi, P.; Liu, H.; Zhang, Y.; Yu, X.; Li, J.; Pu, G. A simple, rapid, and practical method for distinguishing *Lonicerae japonicae* Flos from *Lonicerae* Flos. *Molecules* **2019**, *24*, 3455. [[CrossRef](#)]
8. Baravkar, A.A.; Kale, R.N.; Sawant, S.D. FT-IR spectroscopy: Principle, technique and mathematics. *Int. J. Pharma Bio Sci.* **2011**, *2*, 513–519.
9. Feng, J.; Liu, Y.; Shi, X.; Wang, Q. Potential of hyperspectral imaging for rapid identification of true and false honeysuckle tea leaves. *J. Food Meas. Charact.* **2018**, *12*, 2184–2192. [[CrossRef](#)]

10. Zhang, Y.C.; Deng, J.; Lin, X.L.; Li, Y.M.; Sheng, H.X.; Xia, B.H.; Lin, L.M. Use of ATR-FTIR Spectroscopy and Chemometrics for the Variation of Active Components in Different Harvesting Periods of *Lonicera japonica*. *Int. J. Anal. Chem.* **2022**, *2022*, 8850914. [CrossRef]
11. Sun, S.; Zhou, Q.; Chen, J. *Infrared Spectroscopy for Complex Mixtures*; Chemical Industry: Beijing, China, 2011; pp. 24–66.
12. Filik, J.; Frogley, M.D.; Pijanka, J.K.; Wehbe, K.; Cinque, G. Electric field standing wave artefacts in FTIR micro-spectroscopy of biological materials. *Analyst* **2012**, *137*, 853–861. [CrossRef]
13. Zhao, X.; Zhu, H.; Chen, J.; Ao, Q. FTIR, XRD and SEM Analysis of Ginger Powders with Different Size. *J. Food Process. Preserv.* **2015**, *39*, 2017–2026. [CrossRef]
14. Guiliano, M.; Asia, L.; Onoratini, G.; Mille, G. Applications of diamond crystal ATR FTIR spectroscopy to the characterization of ambers. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2007**, *67*, 1407–1411. [CrossRef]
15. Escamilla, M.N.; Sanz, F.R.; Li, H.; Schönbichler, S.; Yang, B.; Bonn, G.; Huck, C. Rapid determination of baicalin and total baicalein content in *Scutellariae radix* by ATR-IR and NIR spectroscopy. *Talanta* **2013**, *114*, 304–310. [CrossRef]
16. Nikzad-Langerodi, R.; Ortmann, S.; Pferschy-Wenzig, E.M.; Bochkov, V.; Zhao, Y.M.; Miao, J.H.; Atanasov, A.G. Assessment of anti-inflammatory properties of extracts from Honeysuckle (*Lonicera* sp. L., Caprifoliaceae) by ATR-FTIR spectroscopy. *Talanta* **2017**, *175*, 264–272. [CrossRef]
17. Schönbichler, S.; Bittner, L.; Pallua, J.; Popp, M.; Abel, G.; Bonn, G.; Huck, C. Simultaneous quantification of verbenaol and verbascoside in *Verbena officinalis* by ATR-IR and NIR spectroscopy. *J. Pharm. Biomed. Anal.* **2013**, *84*, 97–102. [CrossRef]
18. Li, W.; Cheng, Z.; Wang, Y.; Qu, H. Quality control of *Lonicerae Japonicae Flos* using near infrared spectroscopy and chemometrics. *J. Pharm. Biomed. Anal.* **2013**, *72*, 33–39. [CrossRef]
19. Baek, S.J.; Park, A.; Ahn, Y.-J.; Choo, J. Baseline correction using asymmetrically reweighted penalized least squares smoothing. *Analyst* **2015**, *140*, 250–257. [CrossRef]
20. Strobl, C.; Boulesteix, A.-L.; Zeileis, A.; Hothorn, T. Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinform.* **2007**, *8*, 25. [CrossRef]
21. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [CrossRef]
22. Byvatov, E.; Fechner, U.; Sadowski, J.; Schneider, G. Comparison of support vector machine and artificial neural network systems for drug/non-drug classification. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1882–1889. [CrossRef]
23. Chauchard, F.; Cogdill, R.; Roussel, S.; Roger, J.M.; Bellon-Maurel, V. Application of LS-SVM to non-linear phenomena in NIR spectroscopy: Development of a robust and portable sensor for acidity prediction in grapes. *Chemom. Intell. Lab. Syst.* **2004**, *71*, 141–150. [CrossRef]
24. Boser, B.E.; Guyon, I.M.; Vapnik, V.N. A training algorithm for optimal margin classifiers. In Proceedings of the Fifth Annual Workshop on Computational Learning Theory, Pittsburgh, PA, USA, 27–29 July 1992; ACM: New York, NY, USA, 1992; pp. 144–152.
25. Huang, J.; Liu, J.; Wang, K.; Yang, Z.; Liu, X. Classification and identification of molecules through factor analysis method based on terahertz spectroscopy. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* **2018**, *198*, 198–203. [CrossRef]
26. Maguire, A.; Vega-Carrascal, I.; Bryant, J.; White, L.; Howe, O.; Lyng, F.M.; Meade, A.D. Competitive evaluation of data mining algorithms for use in classification of leukocyte subtypes with Raman microspectroscopy. *Analyst* **2015**, *140*, 2473–2481. [CrossRef]
27. Liang, R.P.; Huang, S.Y.; Shi, S.P.; Sun, X.Y.; Suo, S.B.; Qiu, J.D. A novel algorithm combining support vector machine with the discrete wavelet transform for the prediction of protein subcellular localization. *Comput. Biol. Med.* **2012**, *42*, 180–187. [CrossRef]
28. Meyer, D.; Wien, F.T. Support Vector Machines. The Interface to Libsvm in Package e1071. 2015. Available online: <https://mran.revolutionanalytics.com/snapshot/2016-03-14/web/packages/e1071/vignettes/svmdoc.pdf> (accessed on 14 June 2022).
29. Platt, J.C.; Cristianini, N.; Shawe-Taylor, J. Large margin DAGs for multiclass classification. Advances in Neural Information Processing Systems. In Proceedings of the 12th International Conference on Neural Information Processing Systems, Online, 29 November 1999; pp. 547–553.
30. Joachims, T. Making large-scale SVM learning practical. In *Advances in Kernel Methods: Support Vector Learning*; Scholkopf, B., Burges, C., Smola, A., Eds.; MIT Press: Cambridge, MA, USA, 1999.
31. Stähle, L.; Wold, S. Partial least squares analysis with cross-validation for the two-class problem: A Monte Carlo study. *J. Chemom.* **1987**, *1*, 185–196. [CrossRef]
32. Sandt, C.; Féraud, O.; Oudrhiri, N.; Bonnet, M.L.; Meunier, M.C.; Valogne, Y.; Bertrand, A.; Raphaël, M.; Griscelli, F.; Turhan, A.G. Identification of spectral modifications occurring during reprogramming of somatic cells. *PLoS ONE* **2012**, *7*, e30743.
33. Bruun, S.W.; Kohler, A.; Adt, I.; Sockalingum, G.D.; Manfait, M.; Martens, H. Correcting attenuated total reflection–Fourier transform infrared spectra for water vapor and carbon dioxide. *Appl. Spectrosc.* **2006**, *60*, 1029–1039. [CrossRef]
34. Schulz, H.; Baranska, M. Identification and quantification of valuable plant substances by IR and Raman spectroscopy. *Vib. Spectrosc.* **2007**, *43*, 13–25. [CrossRef]
35. Wang, L.; Yang, Q.; Zhao, H. Sub-regional identification of peanuts from Shandong Province of China based on Fourier transform infrared (FT-IR) spectroscopy. *Food Control* **2021**, *124*, 107879. [CrossRef]
36. Liu, J.; Mukherjee, J.; Hawkes, J.J.; Wilkinson, S.J. Optimization of lipid production for algal biodiesel in nitrogen stressed cells of *Dunaliella salina* using FTIR analysis. *J. Chem. Technol. Biotechnol.* **2013**, *88*, 1807–1814. [CrossRef]
37. Heredia-Guerrero, J.A.; Benítez, J.J.; Domínguez, E.; Bayer, I.S.; Cingolani, R.; Athanassiou, A.; Heredia, A. Infrared and Raman spectroscopic features of plant cuticles: A review. *Front. Plant Sci.* **2014**, *5*, 305. [CrossRef] [PubMed]
38. Yeats, T.H.; Rose, J.K. The formation and function of plant cuticles. *Plant Physiol.* **2013**, *163*, 5–20. [CrossRef] [PubMed]

39. Mazurek, S.; Mucciolo, A.; Humbel, B.M.; Nawrath, C. Transmission Fourier transform infrared microspectroscopy allows simultaneous assessment of cutin and cell-wall polysaccharides of Arabidopsis petals. *Plant J.* **2013**, *74*, 880–891. [[CrossRef](#)] [[PubMed](#)]
40. Kong, J.; Yu, S. Fourier Transform Infrared Spectroscopic Analysis of Protein Secondary Structures. *Acta Biochim. Biophys. Sin.* **2007**, *39*, 549–559. [[CrossRef](#)]
41. Li, Y.; Kong, D.; Wu, H. Comprehensive chemical analysis of the flower buds of five Lonicera species by ATR-FTIR, HPLC-DAD, and chemometric methods. *Rev. Bras. Farmacogn.* **2018**, *28*, 533–541. [[CrossRef](#)]
42. Ordoudi, S.A.; de los Mozos Pascual, M.; Tsimidou, M.Z. On the quality control of traded saffron by means of transmission Fourier-transform mid-infrared (FT-MIR) spectroscopy and chemometrics. *Food Chem.* **2014**, *150*, 414–421. [[CrossRef](#)]
43. Staniszewska-Slezak, E.; Fedorowicz, A.; Kramkowski, K.; Leszczynska, A.; Chlopicki, S.; Baranska, M.; Malek, K. Plasma biomarkers of pulmonary hypertension identified by Fourier transform infrared spectroscopy and principal component analysis. *Analyst* **2015**, *140*, 2273–2279. [[CrossRef](#)]
44. Lammers, K.; Arbuckle-Keil, G.; Dighton, J. FT-IR study of the changes in carbohydrate chemistry of three New Jersey pine barrens leaf litters during simulated control burning. *Soil Biol. Biochem.* **2009**, *41*, 340–347. [[CrossRef](#)]
45. Chun-Yun, C.; Lian-Wen, Q.; Hui-Jun, L.; Ping, L.; Ling, Y.; Hong-Liang, M.; Dan, T. Simultaneous determination of iridoids, phenolic acids, flavonoids, and saponins in *Flos Lonicerae* and *Flos Lonicerae Japonicae* by HPLC-DAD-ELSD coupled with principal component analysis. *J. Sep. Sci.* **2007**, *30*, 3181–3192. [[CrossRef](#)]
46. Ren, M.T.; Chen, J.; Song, Y.; Sheng, L.S.; Li, P.; Qi, L.W. Identification and quantification of 32 bioactive compounds in Lonicera species by high performance liquid chromatography coupled with time-of-flight mass spectrometry. *J. Pharm. Biomed. Anal.* **2008**, *48*, 1351–1360. [[CrossRef](#)]
47. Chen, J.; Guo, B.; Yan, R.; Sun, S.; Zhou, Q. Rapid and automatic chemical identification of the medicinal flower buds of Lonicera plants by the benchtop and hand-held Fourier transform infrared spectroscopy. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2017**, *182*, 81–86. [[CrossRef](#)]
48. Rui, Y.; Chen, J.B.; Sun, S.Q.; Guo, B.L. Rapid identification of *Lonicerae japonicae Flos* and *Lonicerae Flos* by Fourier transform infrared (FT-IR) spectroscopy and two-dimensional correlation analysis. *J. Mol. Struct.* **2016**, *1124*, 110–116.