Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy 264 (2022) 120189

Contents lists available at ScienceDirect

Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy

journal homepage: www.elsevier.com/locate/saa

A rapid quality grade discrimination method for *Gastrodia elata* powderusing ATR-FTIR and chemometrics

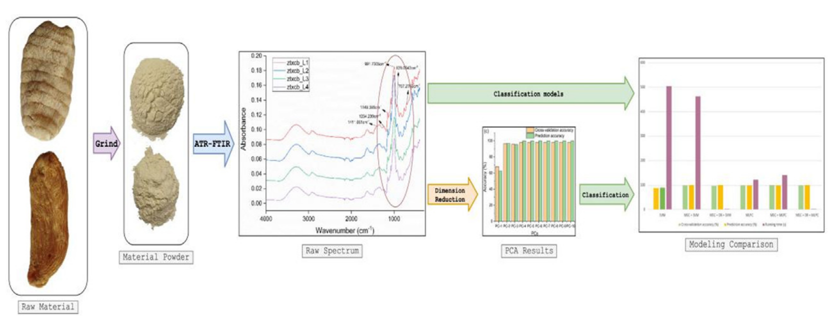
Weixiao Zhan [a], Xiaodong Yang [b], Guoquan Lu [d], Yun Deng [c], Linnan Yang [d,*]

[a] School of Artificial Intelligence, Beijing Normal University, Beijing 100875, China
[b] Department of Automation, Tsinghua University, Beijing 100084, China
[c] Bor Luh Food Safety Center, Department of Food Science & Technology, Shanghai Jiao Tong University, 800 Dongchuan Road, Shanghai 200240, China
[d] School of Big Data, Yunnan Agricultural University, 650201 Kunming, China

HIGHLIGHTS

- A rapid discrimination model of *G. elata* quality was constructed.
- MSC as pretreatment method based on 8-factor modeling effectively classified *G. elata.*
- Extra trees and PCA were effective data manipulators for the IR spectral data.
- MLPC performed better than SVM in distinguishing IR spectral data.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 6 December 2020
Received in revised form 8 July 2021
Accepted 12 July 2021
Available online 16 July 2021

Keywords:
ATR-FTIR spectroscopy
Rapid discrimination
Gastrodia elata
SVM
MLPC

ABSTRACT

Gastrodia elata is an obligate fungal symbiont used in traditional Chinese medicine. There are currently 4 grades of the plant based on the "Commodity Specification Standard of 76 Kinds of Medicinal Materials". The traditional discrimination methods for determining the medicinal grade of *G. elata* powders are complex and time-consuming which are not suitable for rapid analysis. We developed a rapid analysis method for this plant using attenuated total reflection and Fourier-transform infrared spectroscopy (ATR-FTIR) together with machine learning algorithms. The original spectroscopic data was first pretreated using the multiplicative scatter correction (MSC) method and 4 principal components were extracted using extremely randomized trees (Extra-trees) and principal component analysis (PCA) algorithms, and different kinds of classification models were established. We found that multilayer perceptron classifier (MLPC) modeling was superior to support vector machine (SVM) and resulted in validation and prediction accuracies of 99.17% and 100%, respectively and a modeling time of 2.48 s. The methods established from the current study can rapidly and effectively distinguish the 4 different types of *G. elata powders* and thus provides a platform for rapid quality inspection.

1. Introduction

Gastrodia elata is a saprophytic perennial herb in the family Orchidaceae that grows in symbiosis with the fungus*Armillaria mellea*on rotting wood. *G. elata* has been used for more than 2000 years as a Chinese medicinal herb in the mid- and south-western areas of China that include Yunnan, Guizhou, Sichuan, Hunan, and Anhui [1]. *G. elata* extracts have robust anti-convulsive properties and have been used to treat neuralgia, headaches, dizziness, hypertension, epilepsy and tetany [2,3]. This obligate mycoheterotroph is listed in the Food Safety Law of the

* Corresponding author.
E-mail address: lny5400@sina.com (L. Yang).

https://doi.org/10.1016/j.saa.2021.120189
1386-1425/

Table 1
Description of experimental samples.

Sample name	Grade	Standard (ZHI kg^{-1})	Number of training sets (ZHI kg^{-1})	Number of testing sets (ZHI kg^{-1})	Total (ZHI kg^{-1}))
ztxcb_L1	1	<26	37	13	50
ztxcb_L2	2	≥26 to<46	37	13	50
ztxcb_L3	3	≥46 to<90	39	13	52
ztxcb_L4	4	＞90	38	13	51

ZHI kg^{-1}: the number of *G. elata* per kilogram.

People's Republic of China due to its use in both medicine and in the food industry. The active ingredients of *G, elata* include gastrodin (4-β-D-glucopyranosyloxy benzyl alcohol), gastrodigenin, glucoside of 4-hydroxybenzyl alcohol, phenolic aldehyde vanillin and other phenolic compounds [4]. Zhao et al. [5]used a 25 mM borate buffer (pH 10.0) with 10% (v / v) acetonitrile to identify gastrodin (GA), 4-hydroxybenzyl alcohol (HA), vanillyl alcohol (VA), 4-hydroxybenzaldehyde (HD) and vanillin (VL) from *G. elata* extracts. Moreover, Yan et al. [6] used the HPLC method to determine the contents of active ingredients in different grades of *G. elata*, and found that the contents of the active ingredients could not be used as an index for grading. The chemical analysis methods can be used to accurately identify *G. elata* powders but are time- and labor-intensive.

It is well known that infrared spectroscopy as an advanced non-destructive testing technology has been widely used in various fields. Compared to chemical detection, infrared spectroscopy combined with machine learning is a rapid identification method, which is more convenient. A 3D synchronous fluorescence spectroscopy together with principal component analysis (PCA) has been successfully used to identify 6 types of *G. elata* from different locales [7]. Machine learning algorithms have been adapted to solve biochemical problems. For instance, the use of attenuated total reflection and Fourier-transform infrared spectroscopy (ATR-FTIR) combined with machine learning algorithms are currently being adopted to classify Chinese medicinal materials. FTIR and partial least squares method was used to discriminate 4 adulterants from chia (*Salvia hispanica* L.) and sesame (*Sesamum indicum* L.) oils [8]. Notoginseng powders from different localities have also been classified using ATR-FTIR and machine learning algorithms [9]. However, the use machine learning algorithms in the classification of the potency of *G. elata* have not been reported.

Table 2
FTIR peaks location and assignment of 4 different grades of G. elata.

Peak wavenumber (cm^{-1})	Assignments
1411.661	NO_3 anti-symmetric expansion, CH_2 variable angle vibration, COO symmetrical stretching vibration and C—N telescopic shock
1234.239	the aromatic ring (=C—H) inner-plane bending vibrations and N=O telescopic shock
1149.385	aromatic ring (=C—H) inner-plane bending
991.7305	CH_3 rocking vibrations and aromatic ring (=C—H) inner-plane bending
929.0542	C—O—C telescopic shock and O—P—O symmetrical stretching vibration
707.2769	COO variable angle and NH outer-plane vibrations

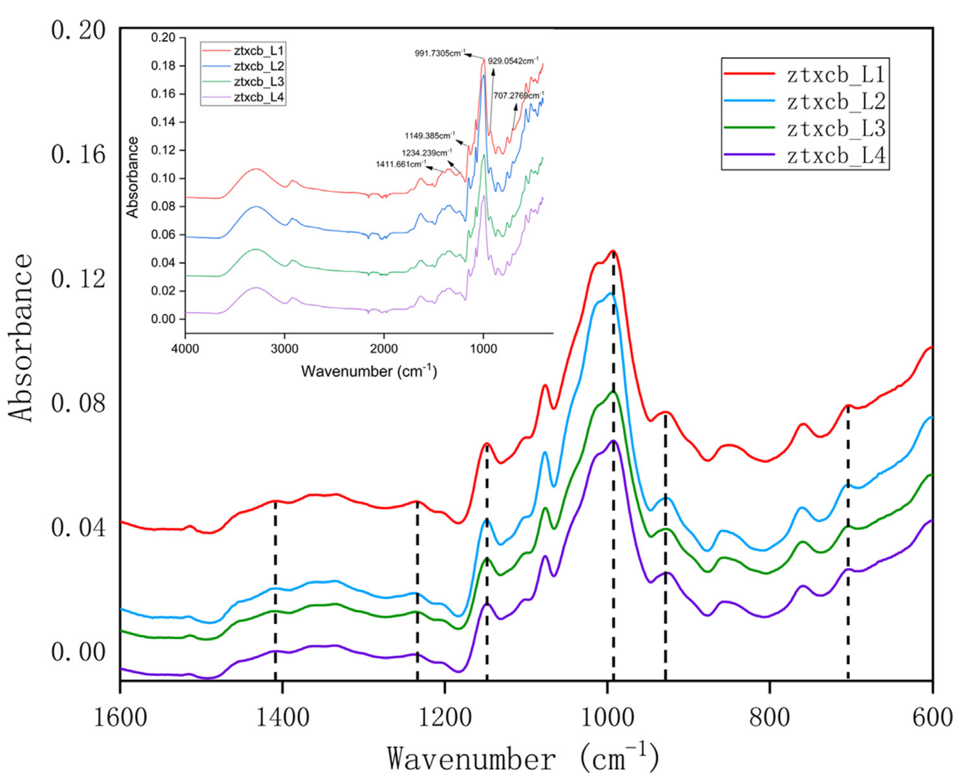


Fig. 1. Primary characteristic peaks from ATR-FTIR spectrum for 4 *G. elata* standard materials. Inset, sample names (see Table 3).

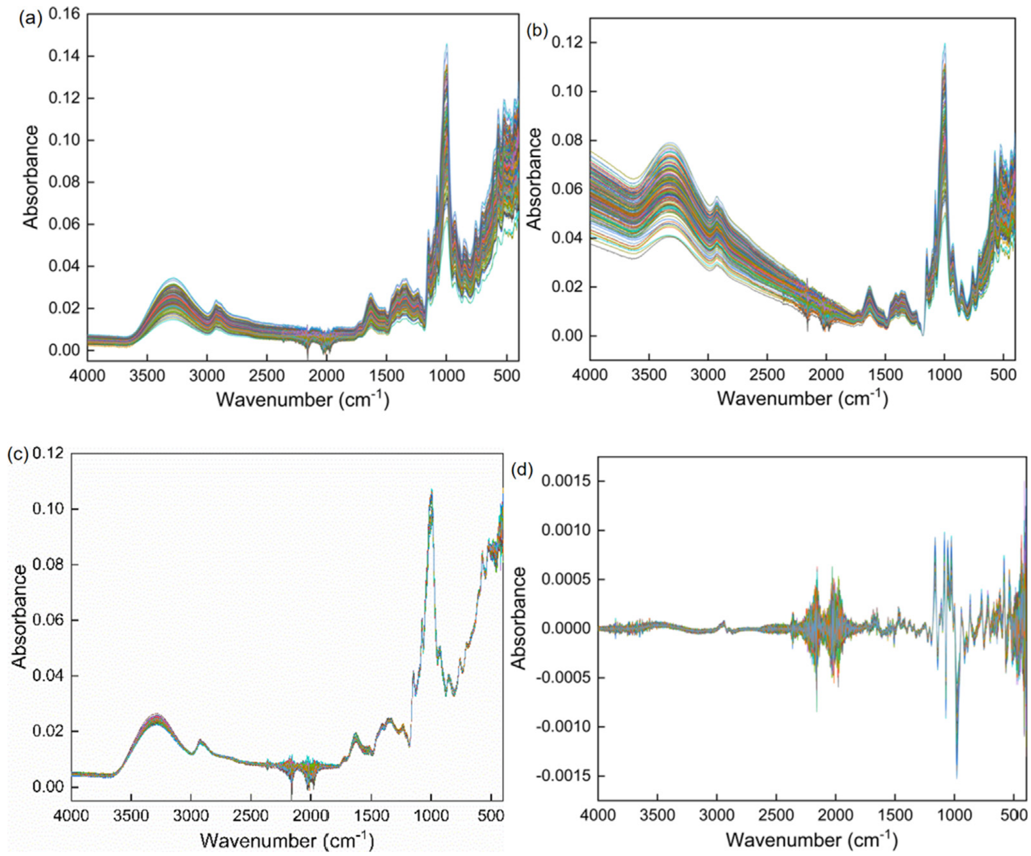


Fig. 2. ATR-FTIR spectra using *G. elata* powdered samples (a) Original unmodified spectrum (b) baseline corrected (c) post-MSC processing (d) post-first derivative processing and 9-point Savitzky-Golay smoothing.

According to the "Commodity Specification Standard of 76 Kinds of Medicinal Materials", the classification standard of *G. elata* is the size of the rhizome, and the grading of the material conformed to the general rules [10]. Therefore, the current study reports the development of a rapid method for the classification of *G. elata* powders using ATR-FTIR combined with a machine learning algorithm. This method is simple in operation and is a direct spectral analysis of *G. elata*. The method guarantees a high accuracy rate as well as efficiency, and can be used for on-site inspections. Pre-treatment methods of one sort or another like baseline, multiplicative scatter correction (MSC), first derivative with 9-points smoothing are tested in order to get a better discrimination result. The data is then processed using extremely randomized trees (Extra-trees) and PCA analyses. The characteristic spectral data following the pre-treatment and primary extraction steps were analyzed to develop a rapid discrimination method suitable for classifying the spectral data generated from *G. elata* tissues.

2. Materials and methods

2.1. Collection of samples

G. elata samples used in the study were collected from Xiaocaoba Town, Yiliang County, Zhaotong city, the world authentic site for *G. elata* in Yunnan province. The geographic locations and climatic conditions are ideally suited for *G. elata* cultivation and have served as sources for the standard material. The grading of the material conformed to the general rules of "Commodity Specification Standard of 76 Kinds of Medicinal Materials" as previously described [10] (Supplementary file 1). The *G. elata* specimens used for this study were thus divided into the 4 different grade types (Table 1).

2.2. Sample preparation

The *G. elata* samples, purchased from local Chinese medicine supplier, were rinsed of surface soil and then dried naturally in

Table 3
Evaluation results for 3 data pretreatment methods.

Preprocessing method	Factors	*RMSECV*	R^2_{cv}	*RMSEP*	R^2_p
Raw spectral data	7	0.301	0.927	0.319	0.923
Baseline	7	0.258	0.946	0.274	0.942
MSC	8	0.194	0.97	0.209	0.967
D1, with 9 smoothing points	5	0.232	0.96	0.235	0.965
D2, with 9 Smoothing points	5	0.273	0.94	0.249	0.964
D1, with 13 Smoothing points	5	2.232	0.96	0.239	0.961
D2, with 13 Smoothing points	5	0.267	0.943	0.247	0.963

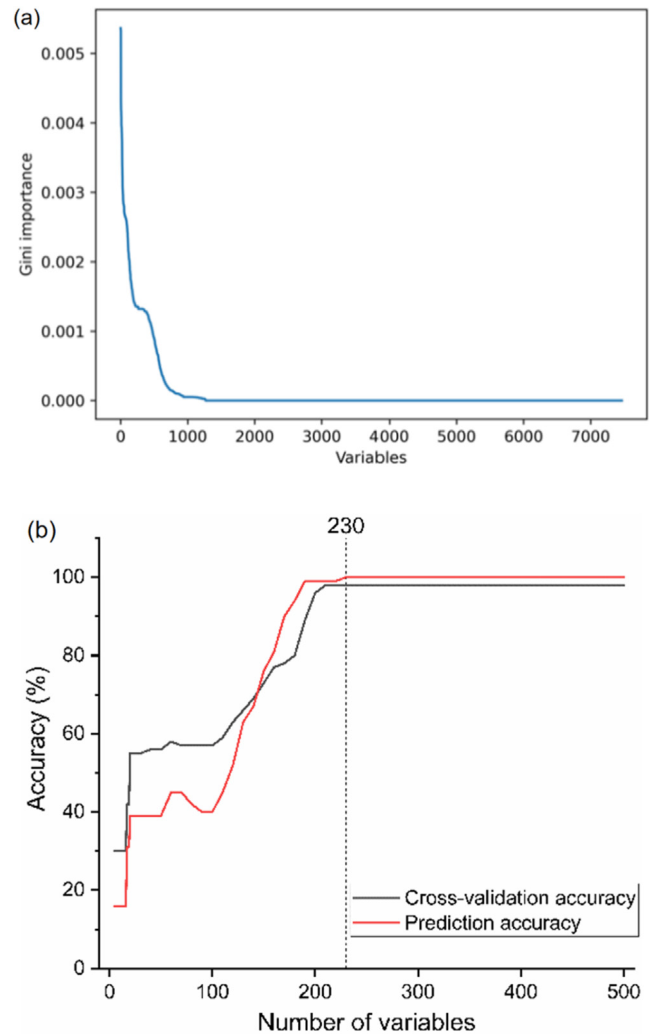


Fig. 3. Selective characteristic variables generated using the Extra-trees algorithm (a) Plots of Gini importance (b) Plots of Cross-validation accuracy and predict accuracy.

the sun. The material was then ground to powder using a grinder and passed through a 200 mesh sieve. The powders were vacuum dried at 50℃ to constant weight. The numbers of powder samples for each grade were 50, 50, 52 and 51 for grades 1–4, respectively.

2.3. Spectrum acquisition and analysis

A Nicolet iS50 FTIR spectrometer (Thermo Fisher, Pittsburg, PA, USA) equipped with an ATR accessory was employed to collect FT-MIR (mid-infrared) spectra in the range 4000–400 cm^{-1} using a wave number resolution of 4 cm^{-1} and an interval of 0.482 cm^{-1}. In addition, 200 mesh sieves were used to control fineness. Briefly, background spectra were recorded to eliminate the impacts of atmospheric H_2O and CO_2 and then recorded every half an hour. A total of 16 scans were taken and an average value was calculated to yield a high signal-to-noise ratio. The spectral analysis was implemented by using Python, Unscrambler version 14, and Microsoft Excel 2016. A total of 230 original spectra were obtained that encompassed 7468 dimensions of raw spectrum profiles. A cutoff at 25% was used for the testing set and 75% for the training set.

2.4. Identification method for G. elata based on ATR-FTIR and chemometrics

2.4.1. Preprocessing of spectral data

The absorption peaks for the characteristic of ATR-FTIR spectra was the key to reveal chemical information from the *G. elata* samples. Background noise was eliminated first using baseline correction followed by MSC[11]and Savitzky-Golaymethod [12]. Partial Least Squares Discriminant Analysis (PLS-DA) was then used to evaluate different pre-treatment methods. In brief, (1) the spectroscopic data along with the original spectroscopic data of pre-treatment input (7468 data points) were combined into 203 × 4 dimensional vectors, (2) each category has an only one category label, if the prediction is correct, the model will output the corresponding label, and (3) an individual evaluation index was calculated using RMSECV, R^2_{CV}, RMSEP and the value for R^2_P, respectively.

2.4.2. Feature selection for spectral data

In order to improve the efficiency for the used modeling, characteristics for spectroscopic data were selected and extracted. Common selecting methods included interval Partial Least Square (iPLS) [13], competitive adaptive reweighted sampling (CARS) [14], successive projection algorithm (SPA) [15] and principal component analysis (PCA) [16]. Usually, the selection procedure involves first selecting the characteristic wavelength and then one or more differential methods is used to select further characteristic variables. To improve efficiency, the iPLS procedure can be omitted to allow manual selection of characteristic bands [17]. Since a characteristic band contains many variables including invalid data and noise data, the Extra-trees algorithm [18] was used to select feature variables and define the order of importance for variables based on Gini coefficients. The characteristic bands were selected by the "elbow" rule and then PCA was used to further select the appropriate principal components for the modeling.

2.4.2.1. Extra-trees. The Extra-trees algorithm consisted of a set of untrimmed decision trees in which training sets were used to generate trees and split nodes in a completely random manner resulting in the characteristics to define the best segmentation. All of these characteristics were evaluated using Gini coefficient analysis.

2.4.2.2. PCA. The original data were projected into a new coordinate system using orthogonal linear alterations and the coordinate axis with the most prominent variable was defined as the first principal component followed by the second principal component and so on, based on the decreasing order of the variables. We used two methods to confirm principal components: a calculation of the cumulative contribution rate for principal components normally greater than 90% and the second was using modeling with different principal components and then observing the cross-validation accuracy until a stable value was achieved [19].

2.4.3. Classifiers suitable for IR spectral data

2.4.3.1. SVM. The purpose of SVM [20] was to generate the best classification hyperplane among different kinds of samples. A nearest sample point for the hyperplane was regarded as the support vector, i.e., a hyperplane was confirmed as long as samples passed through the support vector (much fewer than the sample numbers). This was used to classify samples from different grades. Furthermore, SVM allowed better tolerance for the classification hyperplane to interference information from training samples and thus ensured a model that could be better generalized in handling unknown data sets. Normally, classification efforts can be achieved using linear support vector machines. We also try to use some nonlinear kernel functions, including polynomials, Gaus-

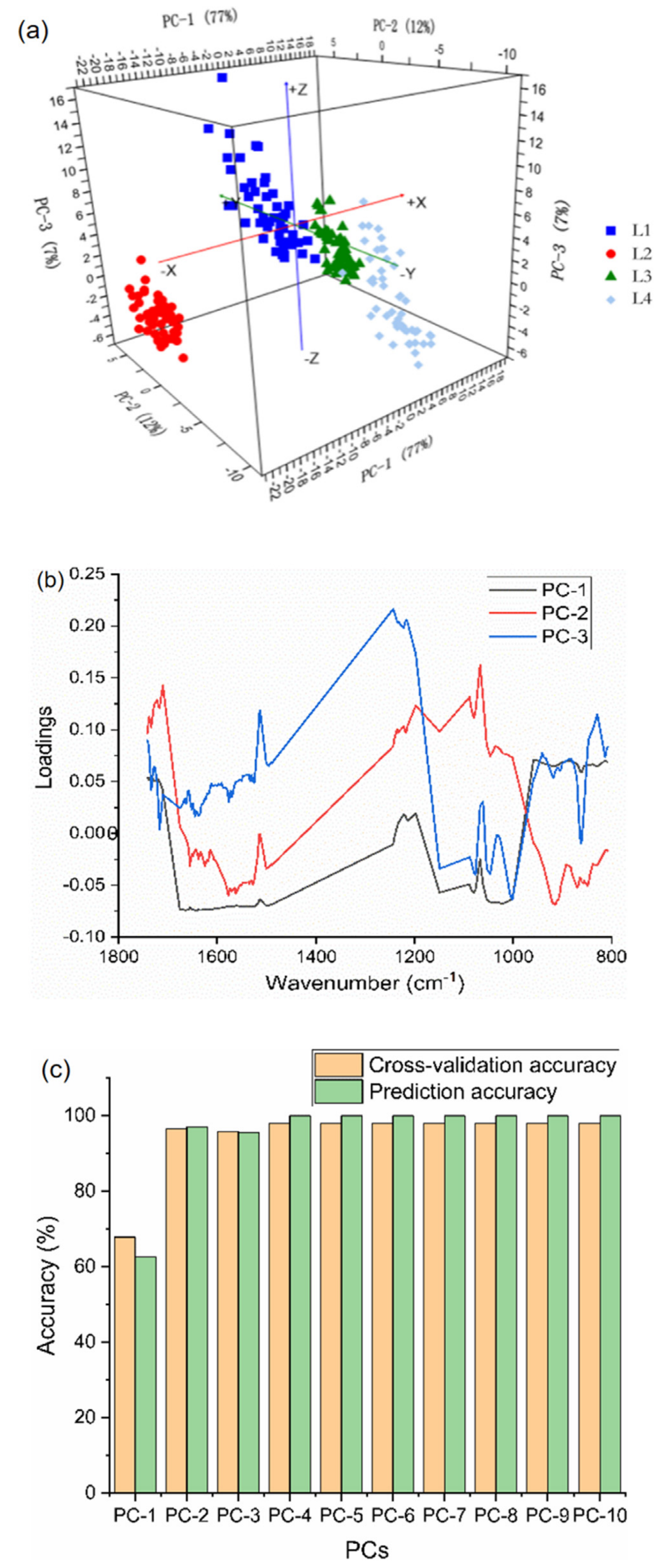


Fig. 4. PCA dimensional reductions (a) Plots of principal component scores (b) Plots of principal component loadings (c) The accuracy of calibration set.

sian, sigmoid, and radial basis kernels. The original spectral data (203 × 7468 dimension) acted as the input that was derived from the MSC processed spectroscopic data (203 × 7468 dimension) of pre-treatments as well as the spectroscopic data (203 × 4 dimension) that were underwent dimension reduction processing by Extra-trees and PCA. Additionally, outputs were the classification numbers of 0, 1, 2, 3. The kernel function was the radial basis kernel function that was used to calculate cross-validation accuracy and prediction accuracy.

2.4.3.2. MLPC. Multilayer perceptron classifier (MLPC) [21] used was a classifier based on forward feeding artificial neural network and was composed of many network layers. Each network layer completely connected to the next layer. This manuscript describes a method using original spectral data (203 × 7468 dimensions), the MSC processed spectroscopic data (203 × 7468 dimensions) of pre-treatments and the spectroscopic data (203 × 4 dimensions) underwent dimension reduction processing by Extra-trees and PCA method as inputs while the classification numbering 0, 1, 2, and 3 were used as output. In conclusion, cross-validation accuracy, prediction accuracy and operating time were optimized.

3. Results and discussion

3.1. Spectral analysis following ATR-FTIR

The primary and characteristic ATR-FTIR peaks spectrum were initially used to distinguish the different substances [22,23]. As for *G. elata*, a per attribution analysis for the characteristic peaks revealed that 1411.661 cm^{-1} represented NO_3 anti-symmetric expansion, CH_2 variable angle vibration, COO symmetrical stretching vibration and C—N telescopic shock [24]. The 1234.239 cm^{-1} peak represented the aromatic ring (=C—H) inner-plane bending and N=O telescopic shock, while 1149.385 cm^{-1} represented aromatic ring (=C—H) inner-plane bending, and 991.7305 cm^{-1} was mainly assigned to CH3 rocking vibrations and aromatic ring (=C—H) inner-plane bending. The 929.0542 cm^{-1} peak represented C—O—C telescopic shock and O—P—O symmetrical stretching vibration while 707.2769 cm^{-1} represented COO variable angle and NH outer-plane vibrations (Fig. 1 and Table 2).

3.2. Fine pre-treatment method

3.2.1. Comparisons following data pre-treatment

The raw data from pre-treatments were analyzed using baseline correction [25], MSC and first derivative with 9-points smoothing. The original spectrum (Fig. 2a) was initially baseline corrected (Fig. 2b) and processed using MSC (Fig. 2c). These steps were further modified using 9-point smoothing to reveal a characteristic spectrum (Fig. 2d). These data sets were further refined using PLS-DA to establish an evaluation model used to compare the results based on analysis of original spectral data using RMSECV, R^2_{CV}, RMSEP, R^2_P as evaluation standards. The adopted MSC generated the best result using PLS-DA based on 8-factor modeling and RMSECV, R^2_{CV}, RMSEP, R^2_P values were 0.194, 0.97, 0.209 and 0.967, respectively (Table 3).

3.2.2. Selection based on Extra-trees

The results were then analyzed using the Extra-trees algorithm to select variables and the significance was compared based on the Gini coefficient and a final model was developed using SVM. The ranking results of Gini significance indicated that when the variation number was >1000, the significance for the variation of characteristic tended to be 0 (Fig. 3a). Elbow was then used to choose characteristic variations from the top 500 with high Gini significance to establish an SVM model. The cross-validation accuracy and the prediction accuracy indicated that when more than 230 characteristic variations were used to establish the SVM model, cross-validation and prediction accuracies for the modeling were high and stable at 98% and 100%, respectively. Therefore, the final

Table 4
IR spectral classification results for 4 different grades of *G. elata* based on SVM modeling.

Number	Datasets	Feature selection methods	Factors	SVM		
				Cross-validation accuracy (%)	Prediction accuracy (%)	Running time (s)
A1	Raw data	\	7468	88.45	89.55	504.54
A2	MSC	\	7468	99.29	100	462.75
A3	MSC	Extra-Trees, PCA	4	97.86	100	2.47

characteristic variations based on the Extra-trees algorithm was 230 (Fig. 3b).

3.2.3. Selection based on PCA algorithms

In order to further reduce the characteristic variations for modeling, the 230 characteristic variations were applied to PCA to extract the principal components. The accumulative contribution rate for the top 3 principal components was 86% and the 4 *G. elata* sample types could be distinguished. In particular, grade 2 samples were the most characteristic compared with the other 3 grades (Fig. 4a). The loading diagram for the top 3 primary PCA elements revealed great fluctuations for characteristic variables that had a higher contribution rate (Fig. 4b). The modeling results using 1 to 10 principal components based on the SVM model demonstrated that when 4 principal components were used for the modeling, the 20-fold cross-validation and prediction accuracies were 97.86% and 100%, respectively. An increase in the number of principal components used did not result in increased accuracy (Fig. 4c). Therefore, a final classification model was established using PCA to select the top 4 principal components in this study.

3.3. ATR-FTIR spectral classification methods

3.3.1. SVM modeling-based

SVM modeling was used to classify IR spectral data for the 4 different grades of *G. elata* powder. The method A1 classified the original IR data using SVM and resulted in cross-validation and prediction accuracies of 88.45% and 89.55%, respectively, with a modeling time of 504.54 s. This result was not satisfactory. However, method A2 that used pre-treatment of the original IR data with MSC resulted in cross-validation and prediction accuracies that were higher than the method using original spectral data and increased by 10.84% and 10.45%, respectively. The method A3 utilized Extra-trees and PCA to select characteristic variations from IR data that were pre-treated using MSC. The variation numbers in the modeling decreased from 7468 to 4 and the cross-validation and prediction accuracies were 97.86% and 100%. The method A3 did not significantly improve the results over method A2 but the modeling time increased by 460.28 s (Table 4)

3.3.2. MLPC Modeling-based

The data were then used to train a multilayer perceptron model that contained one input layer, a hiddenlayer, and an output layer. There were 20 "neurons" in the hide layer and the ReLU function was used to activate the "neurons" in the hide layer and output layer. Limited-memory Broyden-Fletcher-Goldfarb-Shanno (LBFGS) was then used to optimize the model weights and the learning rate for the model was thus set at 0.01. To ensure the convergence of the model, the maximum number of iterations for the model was set at 2000. We were then able to classify the IR spectral data for each of the different grades of *G. elata* based on the MLPC model (Table 5). Method B1 using MLPC directly generated cross-validation and prediction accuracies 99% and 99.25%, respectively, with a modeling time of 122.69 s. Method B2 using MSC pretreatment did not improve prediction accuracy but the cross-validation accuracy reaches 100%. Therefore, method B3 that utilized Extra-trees and PCA resulted in a decrease in the variations in the model from 7468 to 4 although there was no change in cross-validation and prediction accuracies. There was however, a significant decrease in modeling time to 2.48 s (Table 5). These results indicated that both methods A3 and B3 can be used for the rapid identification of the different grades of *G. elata* powders. However, when the difference in the modeling times was close, the cross-accuracy rate of B3 increased 1.31% over A3. In addition, the processing of original IR spectra using MLPC was better due to its strong feedback connections and can accommodate a large amount of characteristic variation data directly to establish its classification model. The process of pre-treatment of IR spectral data could improve the effects of using the SVM classifier while this had no impact on MLPC because the IR spectral noise can also influence the SVM classifier to obtain support vector. However, the neurons of MLPC may help the modeling learn data in a noisy background and therefore has a practical usefulness.

4. Conclusion

This study established a rapid identification model for different grades of *G. elata* based on ATR-FTIR spectrum technology and machine learning-related algorithms. MSC as a pre-treatment method based on 8-factor modeling was more effective and suitable to pre-treat the IR spectroscopic data used in the study. Extra-trees and PCA algorithms could reduce the dimensions for the original data resulting in a decrease in the operation period. This resulted in cross validation and prediction accuracies of 97.86% and 100%, respectively. Following pre-treatment and dimensional reduction, both the SVM classifier and MLPC could be used for rapid identification of different grades of the *G. elata* samples although the performance by MLPC was superior. The quality of *G. elata* potency is greatly affected by the planting environment so that our models can be further improved with the input of additional sample data.

Table 5
IR spectral classification results for *G. elata* grades based on MLPC.

Number	Datasets	Feature selection methods	Factors	MLPC		
				Cross-validation accuracy (%)	Prediction accuracy (%)	Running time (s)
B1	Raw data	\	7468	99	99.25	122.69
B2	MSC	\	7468	100	98.95	140.71
B3	MSC	Extra-Trees, PCA	4	99.17	100	2.48

CRediT authorship contribution statement

Weixiao Zhan: Investigation, Validation, Data curation, Methodology, Formal analysis, Writing – original draft. **Xiaodong Yang:** Formal analysis, Software. **Guoquan Lu:** Resources, Validation, Methodology. **Yun Deng:** Writing - review & editing. **Linnan Yang:** Conceptualization, Methodology, Visualization, Supervision, Project administration.

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.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.saa.2021.120189.

References

[1] S. Sun, Y. Li, L. Zhu, H. Ma, L. Li, Y. Liu, Accurate discrimination of Gastrodia elata from different geographical origins using high-performance liquid chromatography fingerprint combined with boosting partial least-squares discriminant analysis, J. Sep. Sci. 42 (17) (2019) 2875–2882, https://doi.org/10.1002/jssc.v42.1710.1002/jssc.201900073.

[2] Y.S. Ch'ng, Y.C. Loh, C.S. Tan, M. Ahmad, M.Z. Asmawi, W.M. Wan Omar, M.F. Yam, Vasorelaxant properties of vernonia amygdalina ethanol extract and its possible mechanism, Pharm. Biol. 55 (1) (2017) 2083–2094, https://doi.org/10.1080/13880209.2017.1357735.

[3] Y. Zuo, X. Deng, Q. Wu, Discrimination of gastrodia elata from different geographical origin for quality evaluation using newly-build near infrared spectrum coupled with multivariate analysis, Molecules 23 (2018) 1–18, https://doi.org/10.3390/molecules23051088.

[4] L.i. Wang, H. Xiao, X. Liang, L. Wei, Identification of phenolics and nucleoside derivatives in Gastrodia elata by HPLC-UV-MS, J. Sep. Sci. 30 (10) (2007) 1488–1495, https://doi.org/10.1002/jssc.200600469.

[5] Y. Zhao, Q.-e. Cao, Y. Xiang, Z. Hu, Identification and determination of active components in Gastrodia elata Bl. by capillary electrophoresis, J. Chromatogr. A 849 (1) (1999) 277–283, https://doi.org/10.1016/S0021-9673(99)00534-8.

[6] G. Yan, Y. Cai, Y. Yang, G. Wu, K. Che, Z. Yin, G. Jie, Influencing factors analysis on the content of gastrodin and p-hydroxybenzyl alcohol in fresh Gastrodia elata, Sci. Technol. Food Indust. 42 (2021) 6 (in Chinese).

[7] Q. Fan, C. Chen, Z. Huang, C. Zhang, P. Liang, S. Zhao, Discrimination of Rhizoma Gastrodiae (Tianma) using 3D synchronous fluorescence spectroscopy coupled with principal component analysis, Spectrochim. Acta – Part A Mol. Biomol. Spectrosc. 136 (2015) 1621–1625, https://doi.org/10.1016/j.saa.2014.10.056.

[8] S.D. Rodríguez, M. Gagneten, A.E. Farroni, N.M. Percibaldi, M.P. Buera, FT-IR and untargeted chemometric analysis for adulterant detection in chia and sesame oils, Food Control. 105 (2019) 78–85, https://doi.org/10.1016/j.foodcont.2019.05.025.

[9] X. Yang, J. Song, X. Wu, L. Xie, X. Liu, G. Li, Identification of unhealthy Panax notoginseng from different geographical origins by means of multi-label classification, Spectrochim. Acta – Part A Mol. Biomol. Spectrosc. 222 (2019) 117243, https://doi.org/10.1016/j.saa.2019.117243.

[10] State Pharmaceutical Administrationof the People's Republic of China, Commodity specification standard of 76 kinds of medicinal materials, Ministry of Health of the People's Republic of China, 1984..

[11] M.R. Maleki, A.M. Mouazen, H. Ramon, J. De Baerdemaeker, Multiplicative Scatter Correction during On-line Measurement with Near Infrared Spectroscopy, Biosyst. Eng. 96 (3) (2007) 427–433, https://doi.org/10.1016/j.biosystemseng.2006.11.014.

[12] H. Li, Y. Liang, Q. Xu, D. Cao, Key wavelengths screening using competitive adaptive reweighted sampling method for multivariate calibration, Anal. Chim. Acta 648 (1) (2009) 77–84, https://doi.org/10.1016/j.aca.2009.06.046.

[13] M.C.U. Araújo, T.C.B. Saldanha, R.K.H. Galvão, T. Yoneyama, H.C. Chame, V. Visani, The successive projections algorithm for variable selection in spectroscopic multicomponent analysis, Chemom. Intell. Lab. Syst. 57 (2) (2001) 65–73, https://doi.org/10.1016/S0169-7439(01)00119-8.

[14] S. Valle, W. Li, S.J. Qin, Selection of the number of principal components: The variance of the reconstruction error criterion with a comparison to other methods, Ind. Eng. Chem. Res. 38 (1999) 4389–4401, https://doi.org/10.1021/ie990110i.

[15] X. Yang, G. Li, J. Song, M. Gao, S. Zhou, Rapid discrimination of Notoginseng powder adulteration of different grades using FT-MIR spectroscopy combined with chemometrics, Spectrochim. Acta – Part A Mol. Biomol. Spectrosc. 205 (2018) 457–464, https://doi.org/10.1016/j.saa.2018.07.056.

[16] P. Geurts, D. Ernst, L. Wehenkel, Extremely randomized trees, Mach. Learn. 63 (1) (2006) 3–42, https://doi.org/10.1007/s10994-006-6226-1.

[17] P.R. Peres-Neto, D.A. Jackson, K.M. Somers, How many principal components? stopping rules for determining the number of non-trivial axes revisited, Comput. Stat. Data Anal. 49 (4) (2005) 974–997, https://doi.org/10.1016/j.csda.2004.06.015.

[18] O. Chapelle, Choosing multiple parameters for support, Mach. Learn. 46 (2002) 131–159, https://doi.org/10.1023/A:1012450327387.

[19] M.W. Gardner, S.R. Dorling, Artificial neural networks (the multilayer perceptron) - a review of applications in the atmospheric sciences, Atmos. Environ. 32 (14-15) (1998) 2627–2636, https://doi.org/10.1016/S1352-2310(97)00447-0.

[20] R. Karoui, G. Downey, C. Blecker, Mid-infrared spectroscopy coupled with chemometrics: A tool for the analysis of intact food systems and the exploration of their molecular structure-quality relationships-A review, Chem. Rev. 110 (10) (2010) 6144–6168, https://doi.org/10.1021/cr100090k.

[21] B. Sadhana, Department of Information Science and Engineering, Retinal image diabetic detection using multilayer perceptron classifier, in: Lecture Notes on Data Engineering and Communications Technologies, 2021, pp. 591–601.

[22] Z. Yang, H. Yang, H. Yang, Effects of sucrose addition on the rheology and microstructure of k-carrageenan gel, Food Hydrocolloids 75 (2018) 164–173, https://doi.org/10.1016/j.foodhyd.2017.08.032.

[23] D. Yang, S. Gao, H. Yang, Effffects of sucrose addition on the rheology and structure of iota-carrageenan, Food Hydrocolloids 99 (2020), https://doi.org/10.1016/j.foodhyd.2019.105317 105317.

[24] Z. Furong, Study on the mechanism and method of spectral diagnosis of Gastrodia elata quality in Guizhou.Guizhou University, 2009. (in Chinese).

[25] L.C. Sow, H. Yang, Effects of salt and sugar addition on the physicochemical properties and nanostructure of fish gelatin, Food Hydrocolloids 45 (2015) 72–82, https://doi.org/10.1016/j.foodhyd.2014.10.021.