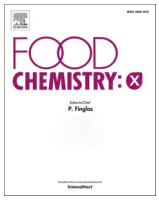
Food Chemistry: X 20 (2023) 100948

Contents lists available at ScienceDirect

Food Chemistry: X

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



Authentication of apples from the Loess Plateau in China based on interannual element fingerprints and multidimensional modelling

Jianyi Zhang [a], Youming Shen [a,*], Ning Ma [b], Guofeng Xu [a]

[a] Quality Inspection and Test Center for Fruit and Nursery Stocks, Ministry of Agriculture and Rural Affairs (Xingcheng), Research Institute of Pomology Chinese Academy of Agricultural Sciences, Xingcheng, Liaoning Province 125100, PR China
[b] College of Veterinary Medicine, Agricultural University of Hebei, Baoding, Hebei 071000, PR China

ARTICLE INFO

Keywords:
Apple
Multielement
Geographical verification
Linear and nonlinear discriminant analysis
Random forest

ABSTRACT

Apple is an important fruit, and fruit authentication is significant for quality and safety control. The Loess Plateau (LP) in China is an important apple-producing region. However, the geographic authentication of LP apples has not been well studied. In this study, we discriminated LP apples based on multielement analysis. We analysed the differences in 29 elements of 522 samples collected from LP and others in 2018–2020 and constructed discriminant models for LP apple authentication. Linear discriminant analysis, partial least square-discriminant analysis, back-propagation artificial neural networks, and random forest (RF) showed different rates in training and validation accuracy. RF showed better tolerance to the removal of the less-important elements in model optimization. The final RF was optimized on 11 elements, which obtained 95.30% training accuracy for the 2018–2019 samples and 97.29% validation accuracy for the 2020 samples. The multielement-based authentication of LP apples could aid further studies of geographical origins.

1. Introduction

Fruit is an essential part of human diets and provide indispensable vitamins, antioxidants, minerals, and many phytonutrients for human health. Apple (*Malus domestica* L.) is a favourite fruit with a high content of nutrients, such as ascorbic acid, K and Mg, and the flavonoids chlorogenic acid, catechin, and quercetin (Acquavia et al., 2021). Apples are suitable for various climatic conditions and are mainly produced in many temperate regions worldwide. China is the leading country in apple production, accounting for approximately half of the total production (https://www.fao.org/faostat/en/#data/QC). Apple production in China is widely distributed in different regions, which mainly include the surrounding Bohai Bay, the Loess Plateau (LP), the ancient Yellow River original and north Qinling Mountains, and the southwest cool highland (Zhang, Zhou, Li, Wei, & Han, 2018; Zhang et al., 2019). Apples from different geographical regions of China are believed to have different internal and external qualities (Kuang, Nie, Li, Cheng, & Shen, 2020). Specifically, apples from Xinjiang, Gansu and Shaanxi Provinces are sweeter in taste and more fragrant. Apples from Yunnan and Sichuan Provinces have good taste and a deeper red color of the fruit peel. Apples from Yantai in Shandong Province have a juicy taste with a suitable acid and sweet flavour. As a result, apples from special areas often demand higher prices. However, dishonest traders often mix and replace these special local products with cheaper or inferior ones to deceive high profits (Zhang et al., 2019). Therefore, the authenticity of apples is essential for ensuring quality and protecting consumer interests.

The Loess Plateau (LP) is one of the most important apple-producing regions in China (Zhang et al., 2018). LP is located in northwestern China, estimated from latitude N 31° to N 42° and longitude E 92° to E 116°, which is close to the central Asia apple origin center (Juniper, Watkins, & Harris, 1998). As a eugenic area for apple growing, the LP planted approximately one million ha of apple trees and output half of the total apple production in China (Zhang et al., 2022). Compared with other regions, the LP has favorable environmental advantages such as relatively arid climates, high altitudes (ranging from 750 to 1700 m), thick soil, and intense and long duration of sunshine, which are suitable for apples growing with higher soluble solid content, and more intensive flavour and extractive appearance (Kuang et al., 2020). However, these characteristics are hardly reliable for distinguishing LP apples from others. Although apples from LP such as Jingning and Tianshui in Gansu Province and Luochuan in Shaanxi Province are specially protected by geographical indications, adulterate behaviours are becoming more serious under the current rapid cargo flows. Poor fruit authentication and quality control frustrate the development of a competitive brand for

* Corresponding author.
E-mail address: shenyouming@caas.cn (Y. Shen).

https://doi.org/10.1016/j.fochx.2023.100948
Received 16 January 2023; Received in revised form 18 September 2023; Accepted 17 October 2023
Available online 20 October 2023

LP apples and infringe on the rights of customers.

In recent years, multielement fingerprints have been proven to be important characteristics for the authentication of agroproducts (Dou et al., 2022; Luykx & Van Ruth, 2008). Elements are the chemical compositions that provide information about the environmental condition of plant-derived products. The multielement fingerprints of agroproducts are closely shaped by the soil and related to the climate in which they grow (Zhao, Tang, & Yang, 2021). Recently, inductively coupled plasma–mass spectrometry (ICP–MS) has been used for the analyses of multielement concentrations in comprehensive biological samples such as grains, vegetables and fruits (Giorgia Potorti et al., 2022; von Wuthenau, Segelke, Müller, Behlok, & Fischer, 2022; Li et al., 2023). Its advantages include being fast, having high throughput, and reducing costs when compared with atomic absorption or atomic emission spectrometry. In addition, statistical analysis methods were used to reorganize variations in data and build discriminate models in mathematical linguistics (Jandric et al., 2021). Specifically, principal component analysis (PCA), linear discriminant analysis (LDA), and partial least square-discriminant analysis (PLS-DA) represent linear discriminant methods, which explain most of the variations in reduced data dimensionality (Luykx & Van Ruth, 2008). The nonlinear discriminate prediction of back-propagation artificial neural networks (BP-ANNs) and random forests (RFs) were equipped with multivariate analysis, and complex iterative calculations were performed to filter out noise from the raw data in effective ways (Maione & Barbosa, 2019). Thus, element fingerprints combined with multidimensional discrimination analysis provide a significant methodology for the authentication of fruit products, such as grapes (Longobardi, Casiello, Centonze, Catucci, & Agostiano, 2017), wine (Cruz et al., 2015; Rapa, Ferrante, Rodushkin, Paulukat, & Conti, 2023), apple (Zhang et al., 2022), apple juice (Bat et al., 2016; Xu, Xu, Wang, Wang, & Liao, 2020), and kiwifruit (Guo, Yuan, Dou, & Yue, 2017).

Our previous studies showed that apples from five regions in China had different element profiles, and the optimized discrimination model of BP-ANN scored approximately 80 % in discriminate accuracy (Zhang et al., 2019). In addition, we found that Se, Pb, Zn, Na, and Mn were significant for the authentication of apples from the southwest cold highlands in China (Zhang et al., 2022). To obtain sufficient accuracy for the discrimination of LP apples, the present study was conducted by performing ICP–MS based element fingerprint detection and multidimensional discriminate modelling. The element concentrations of apples and their variations from different regions and years were obtained. Linear and nonlinear discriminate models were constructed for LP apple authentication. The discrimination rates of different models were compared. The applicable model was also optimized with the important elements, which were obtained with sufficient training and validation accuracy for LP apple discrimination. This study provides a basis for further apple geographical authentication.

2. Materials and methods

2.1. Reagents and solvents

ICP–MS multielement calibration standard solutions with Ag, Al, As, Ba, Be, Ca, Cd, Co, Cr, Cs, Cu, Fe, Ga, K, Li, Mg, Mn, Na, Ni, Pb, Rb, Se, Sr, Tl, U, V, and Zn at 10 μg/mL and internal standard solutions with Bi, Ge, In, Li, Lu, Rh, Sc, and Tb at 100 μg/mL were purchased from Agilent Technologies (California, USA). Standard solutions of elements including B and P at 100 μg/mL and the certified reference material (CRM) of apple (GBW10019, serial number: GSB-10) were purchased from the National Institute of Metrology of China (Beijing, China). Nitric acid (70 %) was purchased from Sinopharm Chemical Reagent Beijing Co., Ltd. (Beijing, China). High-purity argon and helium gases were purchased from Huludao Weiye Gas Reagent Co., Ltd. (Liaoning, China). Ultra-pure water (18 MΩ) was obtained from Wahaha Co., Ltd. (Hangzhou, China) and further purified on a water purification system (Milli-Q-Direct 8, Millipore, California, USA).

2.2. Sample collection and prepreparation

A total of 522 apple orchards were selected for sample collection from the main Chinese apple-producing regions from 2018 to 2020, including 198 samples from the LP region and 324 representatives from the non-Loess Plateau (NLP) region (Fig. 1 and Table S1). According to the geographic distribution, the LP region contained Gansu, Shaanxi, Shanxi and northwest Henan Provinces. The NLP region contained Liaoning, Shandong, Hebei, Yunnan and southeast Henan Provinces. Fuji apple was the selected cultivar, the orchard chosen was considered to represent the apple distribution, and the geographic locations were recorded in detail (Table S1). For each sample, 5 kg of apple fruits at harvest maturity were collected from at least 6 trees and transported to the laboratory immediately. Samples were washed with water, rinsed with deionized water, and naturally dried. Then, apples were obtained, and the core was discarded, cut into particles of 1 cm^3, frozen in liquid nitrogen, ground by a SPEX SamplePrep system (Metuchen, NJ, USA), and stored at −20 °C for further use.

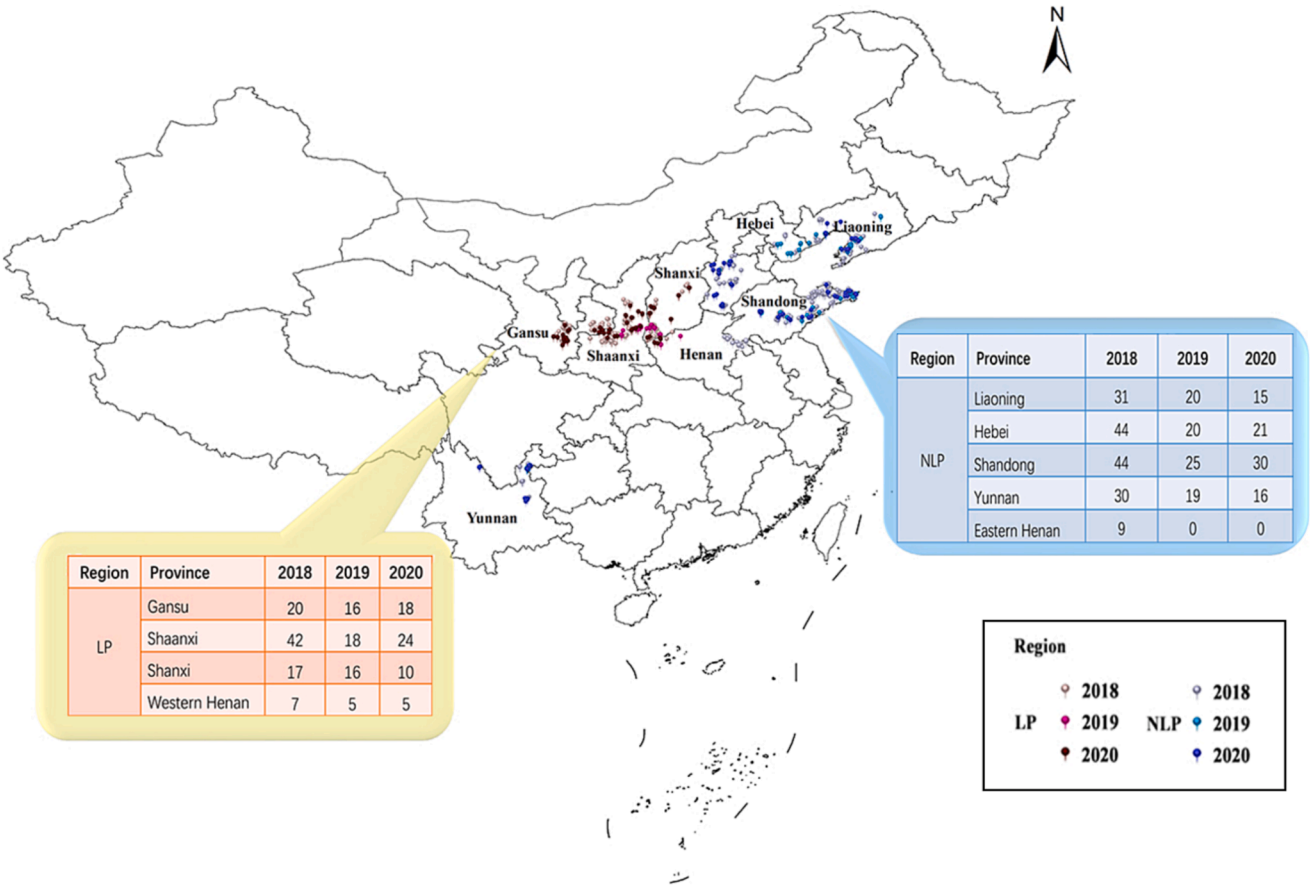
2.3. Sample digestion and instrumental analysis

Three replicates of each sample of 5.0 g were introduced into digestion tubes, and 8.0 mL of nitric acid (70 %) was added and incubated at room temperature overnight (Zhang et al., 2022). The sample digestion was conducted on a MARS6 system (CEM, Matthews, North Carolina, USA) in a 1600 kW microwave, and run with the following temperature programs: maintaining 40 °C for 5 min; heating at 10 °C/min for 10 min to 140 °C; maintaining 140 °C for 5 min; heating at 5 °C/min for 8 min to 180 °C; and maintaining 180 °C for 10 min. Then, the samples were transferred to a vacuum acid-catching system to evaporate the nitric acid until the residual aqueous layer was concentrated to 2 to 3 mL. After cooling to room temperature, the samples were diluted to 50 mL with purified water and filtered through a 0.22 μm nylon membrane from Tianjin Jinteng Experimental Equipment Co., Ltd. (Tianjing, China). The supernatant was subjected to ICP–MS analysis.

The analyses of 29 elements were conducted by an ICP–MS system (Agilent ICP–MS 7850, California, USA). High purity argon was used as the carrier gas, with a total pressure of 600 kPa. Helium was used as the MS shield gas, with a pressure of 100 kPa. The tuning of the mass spectrum was used to confirm the instrument's performance. The error of the mass axis was within 0.1 Da, and the 10 % peak width was higher than 0.65, which is considered a good performance of mass spectrometry. Calibration curves with Ag, Al, As, Ba, Be, Cd, Co, Cr, Cs, Cu, Fe, Ga, Li, Mn, Na, Ni, Pb, Rb, Se, Sr, Tl, U, and V were prepared with concentrations ranging from 0 to 200 μg/L within 10 gradients. Macroelements of Ca, K, Mg, and P were prepared at a concentration of 2000 μg/L for the upper limit of the standard curves. Samples were injected with a flow rate of 0.8 mL/min. The internal standard solution was mixed with each sample to monitor the stability of the system and evaluate the reliability of the experimental data. In each branch analysis, the reference standard of apple was simultaneously prepared and analysed for the analytical quality assessment. The elemental concentrations of the reference material were compared with the certified values. Only if the elements of the reference material obtained with acceptable recoveries and variations, were the elemental concentrations of the samples recorded.

2.4. Data statistics

The concentrations of the elements were calculated and arranged in Microsoft Excel (version 2013). Student's *t* test was performed to identify the elements with different concentrations between LP and NLP samples by SPSS software (version 24.0). PCA was used to construct several uncorrelated principal components (PCs) to observe the



Region	Province	2018	2019	2020
LP	Gansu	20	16	18
	Shaanxi	42	18	24
	Shanxi	17	16	10
	Western Henan	7	5	5

Region	Province	2018	2019	2020
NLP	Liaoning	31	20	15
	Hebei	44	20	21
	Shandong	44	25	30
	Yunnan	30	19	16
	Eastern Henan	9	0	0

Fig. 1. The picture shows a map of China with detailed origins of the apple samples collected. Provinces of Gansu, Shaanxi, Shanxi and northwest Henan represent the Loess Plateau (LP) apple-producing region. Provinces of Liaoning, Shandong, Hebei, Yunnan and southeast Henan represent the non-Loess Plateau (NLP) region.

intergroup differences in reduced data dimensionalities (Jandric et al., 2021). Sample classification models were performed by R (version 3.5.1) and MATLAB (version 2017b), including the linear models LDA (Jandric et al., 2021) and PLS-DA (Luykx & Van Ruth, 2008) and the nonlinear models BP-ANN (Maione & Barbosa, 2019) and RF (Maxwell, Warner, & Fang, 2018). The validation and training accuracies of the models were obtained by calculating the proportions of effective sample discrimination. Models were further optimized according to the values of variable importance in projection (VIP) of elements. Receiver operating characteristic (ROC) analyses were performed to evaluate the predictive performance of these discriminate models.

Samples were grouped into training and validation sets according to a fixed pattern, which corresponded to each discriminate model (Zhang et al., 2019; Zhang et al., 2022). First, samples from 2018 were selected as the training set (244 samples), and samples from 2019 (139 samples) and 2020 (139 samples) were selected as two individual validation sets. Then, samples from 2019 were selected as the training set, and samples from 2020 were selected as the validation set. Samples from 2018 and 2019 (383 samples) were selected as the training set, and samples from 2020 were selected as the validation set. Finally, all the samples were selected for constructing discriminate models, which were randomly and evenly subdivided into two disjoint subsets at a 3 to 1 ratio: a training set containing 392 samples and a test set containing 130 samples.

3. Results

3.1. Analytical method assurance

By performing ICP–MS analysis, we established a method for the detection of 29 elements in apples. The calibration curves for element detection were calculated by the predetection of standard solutions (Table S2). Elements were all obtained with correlation coefficient values higher than 0.99. The CRM of apple was simultaneously detected in each independent sample test, and from all the tests, the experimental values of 27 elements were detected with acceptable ranges, and their recoveries ranged from 95 % to 102 % (Table S3). Two elements, Ag and Ga, were without the recommended concentrations in the reference apple, which were obtained with relative standard deviations (RSDs) within ± 15 %. These results proved that the apple CRM (GBW10019) was an effective reference material for analytical quality control. Therefore, the current methods were sufficient to obtain the apple element compositions with high accuracy and precision.

3.2. Multielements, intergroup differences and PCA results of apples

Table 1 and Fig. S1 show the multielement compositions of apples from the LP and NLP regions from 2018 to 2020. Apples were detected with relatively higher concentrations of K, P, Mg, and Ca, with average concentrations of 1138.00, 118.72, 41.00, and 48.66 mg/kg, respectively. LP apples generally contained a higher concentration of Ca but a lower concentration of Mg than NLP apples. Apples were detected with moderate concentrations of B, Na, Fe, Al, and Rb, with average concentrations of 2.47, 4.78, 2.09, 1.16, and 0.88 mg/kg, respectively. LP apples were generally detected with lower concentrations of B, Na, and Fe than the NLP apples (Student's *t* test $p < 0.05$, Fig. S1). The other 16 elements were trace in apples and were detected at concentrations below 1.00 mg/kg. Among them, Mn, Sr, Pb, As, Co, Se, Cd, Tl, and Be were detected with significantly different concentrations between LP and NLP apples ($p < 0.05$, Fig. S1).

Table 1
Multielement concentrations (mg/kg, marked a × 10^{-3}) in apple samples from the Loess Plateau (LP) and non-LP (NLP).

Sample set	2018		2019		2020	
	LP (86)	NLP (158)	LP (55)	NLP (84)	LP (57)	NLP (82)
K	1079.95 ± 199.49	1227.81 ± 231.65	1024.23 ± 249.17	1192.23 ± 243.80	1101.79 ± 263.80	1077.76 ± 235.26
P	118.66 ± 39.52	132.76 ± 82.20	95.77 ± 37.45	123.39 ± 31.64	97.52 ± 30.05	119.30 ± 79.01
Ca	41.40 ± 5.64	34.95 ± 8.57	47.17 ± 7.61	42.22 ± 11.77	43.95 ± 7.68	43.64 ± 9.85
Mg	48.63 ± 6.74	52.29 ± 9.19	43.07 ± 6.55	51.10 ± 9.84	43.52 ± 6.18	46.96 ± 8.75
B	2.56 ± 0.92	2.60 ± 0.84	2.08 ± 0.91	2.69 ± 1.11	2.03 ± 0.73	2.29 ± 0.86
Na	4.98 ± 3.71	6.47 ± 2.74	2.58 ± 1.78	4.76 ± 2.54	2.58 ± 2.03	4.59 ± 1.91
Fe	1.48 ± 0.57	2.49 ± 3.85	1.42 ± 0.56	2.02 ± 0.89	1.54 ± 0.79	1.91 ± 0.62
Al	1.31 ± 0.74	1.25 ± 0.80	0.93 ± 0.50	1.19 ± 0.73	1.01 ± 1.00	1.10 ± 0.71
Rb	0.83 ± 0.50	0.96 ± 0.58	0.72 ± 0.35	0.84 ± 0.49	1.03 ± 0.72	0.83 ± 0.52
Cu	0.58 ± 0.21	0.57 ± 0.21	0.42 ± 0.21	0.53 ± 0.21	0.46 ± 0.42	0.54 ± 0.20
Mn	0.33 ± 0.10	0.47 ± 0.15	0.32 ± 0.08	0.47 ± 0.19	0.36 ± 0.20	0.44 ± 0.16
Zn	0.25 ± 0.07	0.29 ± 0.08	0.42 ± 0.53	0.36 ± 0.18	0.34 ± 0.18	0.31 ± 0.19
Sr	0.39 ± 0.34	0.19 ± 0.16	0.40 ± 0.43	0.18 ± 0.10	0.30 ± 0.35	0.19 ± 0.10
Ba	0.15 ± 0.05	0.18 ± 0.15	0.16 ± 0.09	0.16 ± 0.12	0.11 ± 0.10	0.07 ± 0.08
Pb[a]	26.37 ± 22.80	14.04 ± 12.46	26.33 ± 27.69	12.05 ± 7.45	18.03 ± 21.47	13.21 ± 7.26
Cr[a]	14.08 ± 13.01	34.94 ± 52.90	21.90 ± 17.59	27.38 ± 36.54	18.20 ± 17.60	18.25 ± 15.19
Ni[a]	55.11 ± 188.04	22.32 ± 47.20	13.02 ± 35.07	12.85 ± 14.83	8.52 ± 7.76	11.00 ± 6.51
As[a]	2.46 ± 1.26	4.73 ± 8.16	2.38 ± 1.58	5.48 ± 13.96	1.38 ± 1.47	7.73 ± 16.40
Co[a]	2.39 ± 0.88	5.66 ± 3.47	2.20 ± 0.86	5.09 ± 3.37	3.59 ± 6.40	3.40 ± 2.66
Cs[a]	2.82 ± 1.53	3.15 ± 2.45	2.55 ± 1.57	2.76 ± 2.25	3.87 ± 3.92	3.31 ± 2.60
Li[a]	8.78 ± 16.76	4.29 ± 10.23	7.37 ± 9.12	5.77 ± 25.91	5.05 ± 7.86	5.73 ± 26.12
Se[a]	3.12 ± 1.16	3.34 ± 1.45	4.38 ± 1.71	3.99 ± 1.54	3.61 ± 1.18	3.77 ± 3.16
V[a]	1.31 ± 0.51	1.26 ± 0.46	1.62 ± 0.70	1.69 ± 0.81	1.63 ± 0.78	1.61 ± 0.73
Ag[a]	0.52 ± 0.74	0.42 ± 0.28	0.79 ± 1.03	0.83 ± 1.37	0.73 ± 1.03	0.77 ± 1.97
Cd[a]	0.23 ± 0.32	0.53 ± 0.80	0.38 ± 0.30	0.51 ± 0.64	1.23 ± 3.33	0.50 ± 0.65
Ga[a]	0.35 ± 0.14	0.37 ± 0.18	0.38 ± 0.16	0.42 ± 0.21	0.36 ± 0.20	0.42 ± 0.21
Tl[a]	0.22 ± 0.17	0.38 ± 0.38	0.20 ± 0.10	0.34 ± 0.28	0.19 ± 0.10	0.31 ± 0.19
U[a]	0.17 ± 0.23	0.13 ± 0.21	0.13 ± 0.07	0.17 ± 0.18	0.12 ± 0.09	0.43 ± 0.31
Be[a]	0.08 ± 0.06	0.27 ± 0.21	0.04 ± 0.07	0.25 ± 0.27	0.08 ± 0.13	0.17 ± 0.19

Based on the 29 elements in the apples, PCA was performed to construct several uncorrelated PCs to reduce the data dimensionality. From samples collected in 2018, 2019, and 2020, the first five PCs obtained 64.76 %, 62.21 %, and 80.21 % of the total variance, respectively. Based on the score plots of the first two PCs, the PCA diagrams showed overlaps between LP and NLP samples for three consecutive years (Fig. S2). From the combined samples of 2018–2020, 2018–2019, and 2019–2020, the first five PCs scored 74.65 %, 55.79 % and 57.39 % of the total variance, respectively. As shown in Fig. 2A and B, NLP samples were mainly obtained with relatively higher PC1 scores than LP samples, suggesting considerable differences in element profiles between LP and NLP samples. However, major PCA diagrams showed overlaps between LP and NLP samples, reflecting that PCA showed weak function in sample discrimination. The contribution values of 29 elements to different PCA models are shown in Table S4. Elements V, Mn, Ba, Ca, Co, Rb, Ga, K, and Tl have higher contribution rates to multiple models, while Al, U, B, P, Li, and Se have lower contributions to these models. Therefore, PCA initially showed the differences in multielement profiles between LP and NLP apples.

3.3. Linear models for discrimination of apples from LP and NLP regions

PLS-DA was conducted to discriminate the LP and NLP apples by linear classification methodology. The PLS-DA model was first established based on samples from 2018, and its training accuracy of LP and NLP samples reached 69.76 % and 74.05 %, respectively (Table 2). The predictive abilities of the model were tested by the validation samples of 2019 and 2020, which obtained validation accuracies that reached 75.54 % and 75.53 %, respectively. Based on the training samples of 2019, the established PLS-DA model obtained a training accuracy of 78.42 %. The predictive abilities of this model for the validation samples of 2020 reached 74.82 %. PLS-DA of 2018–2019 samples obtained 75.72 % training accuracy, and its validation accuracy for 2020 samples was 74.82 %. PLS-DA two-dimensional plots showed the separation between LP and NLP samples from 2018 and 2019, and the separation trends were related to the different scores in function 1 (Fig. 2C). PLS-DA of all 2018–2020 samples was obtained with a training accuracy of 77.97 %. ROC analysis showed that the area under the curve (AUC) averaged 0.95, indicating that the model shows consistency with reality (Table S5). The values of variable importance in projection (VIP) for 29 elements were calculated in PLS-DA models (Table S6). Be, Sr, Co, Pb, As, Cr, Mn, and Na gave significant VIP scores, suggesting that these elements have an important contribution to the model (Fig. 3A).

LDA was used to construct a linear discriminate model for classifying the LP and NLP apples. Based on the samples from 2018, LDA obtained a training accuracy of 91.80 % (Table 2). The validation accuracy for the 2018 and 2019 samples reached 97.84 % and 94.96 %, respectively. The LDA model of 2019 samples had a training accuracy of 90.64 %, and its validation accuracy for 2020 samples reached 91.37 %. LDA of 2018–2019 samples obtained 92.69 % training accuracy, and its validation accuracy for 2020 samples was 94.24 %. LDA two-dimensional plots showed the separation between LP and NLP samples from 2018 and 2019 (Fig. 2D). LDA of all 2018–2020 samples was obtained with a training accuracy of 91.57 %. ROC analysis showed that the AUC averaged 0.96, indicating that the model showed good performance (Table S5). The values of variable importance in projection (VIP) for 29 elements were calculated in LDA models (Table S6). Co, Be, Mn, Sr, Pb, Na, K, Mg, Ca, and Tl gave significant VIP scores (Fig. 3A), suggesting that these elements have an important contribution to the model.

From the results of the linear models, the LDA models generally obtained higher training and validation accuracies than those of the PLS-DA models. Therefore, LDA was selected for linear model optimization. Based on the elemental VIP scores, the significant elements were selected for the linear model reconstruction, and the model accuracies were obtained accordingly. The final model was optimized based on the training set of samples from 2018 and 2019 and the validation set of samples from 2020 (Fig. 3B). After reducing the less important elements, the final LDA model was conducted on 10 elements: Sr, Mn, As, Cs, Co, Cr, Pb, Be, Na, and U, which was obtained with a training accuracy of 92.69 % based on 2018 and 2019 samples, and a validation accuracy of

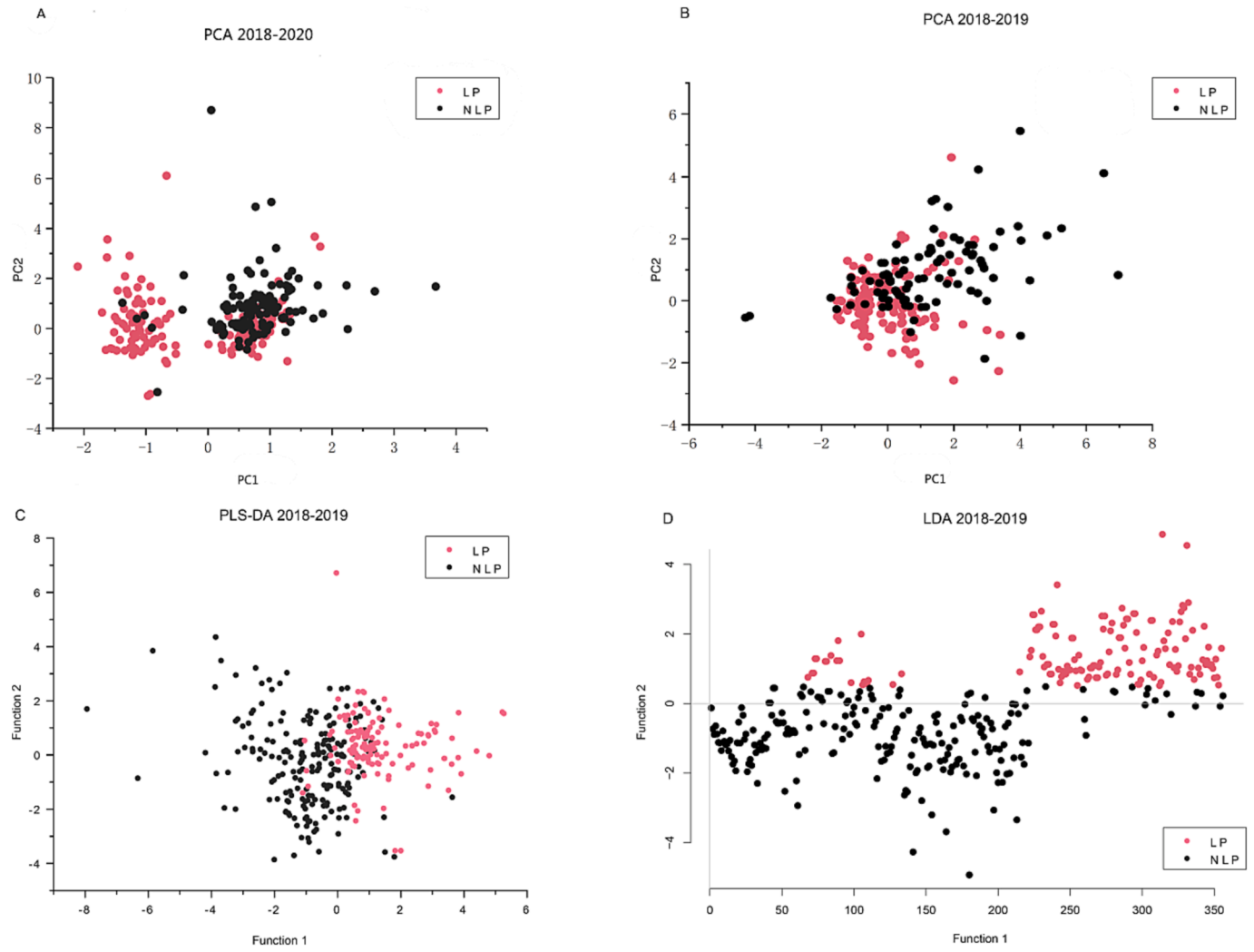


Fig. 2. PCA, PLS-DA and LDA of samples from Loess Plateau (LP) and non-LP (NLP) regions with the contents of mutielements. A: PCA of all samples showed the plot distribution between LP and NLP samples. B: PCA of samples from 2018 to 2019 showed overlaps between LP and NLP samples. C: PLS-DA of samples from 2018 to 2019 showed separation between LP and NLP samples. D: LDA of samples from 2018 to 2019 showed the separation between LP and NLP samples.

94.24 % for 2020 samples. The important features of elements ranked by LDA are shown in Fig. 3C.

3.4. Nonlinear models for discrimination of apples from LP and NLP regions

BP-ANN was carried out to discriminate the LP and NLP apples with a nonlinear methodology. Based on the samples from 2018, the BP-ANN obtained a training accuracy of 93.44 % (Table 2). The validation accuracy for the 2018 and 2019 samples reached 95.68 % and 93.53 %, respectively. The BP-ANN model of the 2019 samples had a training accuracy of 92.09 %, and its verification accuracy of the 2020 sample reached 93.53 %. The BP-ANN of the 2018–2019 samples obtained 94.26 % training accuracy, and its validation accuracy for 2020 samples was 94.24 %. The BP-ANN of all 2018–2020 samples was obtained with a training accuracy of 94.06 %. ROC analysis showed that the AUC was 0.94 on average, indicating that the model showed good performance (Table S5). The values of variable importance in projection (VIP) for 29 elements were calculated in the BP-ANN models (Table S6). Mn, Na, K, Cd, Co, Be, Pb, Ca, Sr, Li, and U gave significant VIP scores, suggesting that these elements have significant contributions to modelling (Fig. 4A).

RF was used to construct binary classification trees for sample discrimination. Based on the samples from 2018, RF obtained a training accuracy of 94.67 % (Table 2). The validation accuracy for the 2018 and 2019 samples reached 98.56 % and 98.57 %, respectively. The RF model of the 2019 samples had a training accuracy of 94.96 %, and its validation accuracy of the 2020 samples reached 96.40 %. The RF of the 2018–2019 samples obtained 95.30 % training accuracy, and its validation accuracy for the 2020 samples was 97.29 %. The RF of all 2018–2020 samples was obtained with a training accuracy of 95.79 %. ROC analysis showed that the AUC was 0.99, indicating that the model showed good performance (Table S5). The values of variable importance in projection (VIP) for 29 elements were calculated in the RF models (Table S6). Be, Co, Sr, Cd, Li, Tl, Pb, Ca, Na, and Ba gave significant VIP scores, suggesting that these elements have an important contribution to the model (Fig. 4A).

The nonlinear models were optimized to obtain an operational model for the discrimination of LP and NLP apples. The RF models showed better performance than those of BP-ANN in model reconstruction, which was also obtained with acceptable training and validation accuracies after reducing the less-important elements. The optimized RF model was conducted on 11 elements: Cd, Ca, Be, Mn, Na, Sr, Co, Li, Ba, Tl, and Pb. During tree construction, the cumulative error rates decreased to 0.09 and 0.06 for the LP and NLP samples, respectively (Fig. 4B). The classification accuracy was 95.30 % based on samples from 2018 and 2019, and the classification rate of cross-validated observations was 97.29 % for the 2020 samples. The important features

Table 2
Performance comparisons of commonly used discrimination models to classify the geographical origin of apples from Loess Plateau (LP) and non-LP (NLP) regions in China.

Group	Group 1 Training set 2018		Validation set 1 2019a		Validation set 2 2020a		Group 2 Training set 2019		Validation set 2020b		Group 3 Training set 2018–2019		Validation set 2020c		Group 4 Modelling samples 2018–2020	
Sample	LP	NLP	LP	NLP	LP	NLP	LP	NLP	LP	NLP	LP	NLP	LP	NLP	LP	NLP
	86	158	55	84	57	82	55	84	57	82	150	233	57	82	207	315
A:PLS-DA																
Valid	60	117	41	64	43	62	42	67	44	60	106	184	43	61	154	253
False	26	41	14	20	14	20	13	17	13	22	35	58	14	21	44	71
Accuracy %	69.76	74.05	74.55	76.19	75.44	75.61	76.36	79.76	77.19	73.17	75.18	76.03	75.44	74.39	77.77	78.09
Averaged %	72.54		75.54		75.53		78.42		74.82		75.72		74.82		77.97	
B: LDA																
Valid	81	143	54	82	55	77	49	77	51	76	130	225	55	76	180	298
False	5	15	1	2	2	5	6	7	6	6	11	17	2	6	18	26
Accuracy %	94.19	90.51	98.18	97.62	96.49	93.9	89.09	91.67	89.47	92.68	92.2	92.98	96.49	92.68	90.91	91.98
Averaged %	91.8		97.84		94.96		90.64		91.37		92.69		94.24		91.57	
C: BP-ANN																
Valid	79	149	53	80	54	76	50	78	53	77	133	228	53	78	187	304
False	7	9	2	4	3	6	5	6	4	5	8	16	4	4	11	20
Accuracy %	91.86	94.3	96.36	95.24	94.74	92.68	90.9	92.86	92.98	93.9	94.32	94.21	92.98	95.12	94.44	93.83
Averaged %	93.44		95.68		93.53		92.09		93.53		94.26		94.24		94.06	
D: RF																
Valid	82	149	55	82	56	81	53	79	55	79	134	231	56	79	192	308
False	4	9	0	2	1	1	2	5	2	3	7	11	1	3	6	16
Accuracy %	94.74	94.3	100	97.62	98.25	98.78	96.36	94.05	96.49	96.34	95.04	95.45	98.25	96.34	96.97	95.06
Averaged %	94.67		98.56		98.57		94.96		96.40		95.30		97.29		95.79	

PCA: Principal component analysis; LDA: Linear discriminant analysis; PLS-DA: Partial least squares discrimination analysis; BP-ANN: Back-propagation artificial neural networks; RF: Random forest. The accuracies of the validation datasets for 2019a, 2020a, and 2020b were predicted by using training datasets from 2018, 2018, and 2018–2019, respectively.

ranked by RF were also obtained (Fig. 4C).

4. Discussion

For the authentication of apples from the Loess Plateau in China, this study analyses 29 elements of 522 apples collected from different regions in 2018–2020. In this study, a validated method of ICP–MS detection was established, which can be used to determine the content of multiple elements in a sample through one-time sampling and detection. Compared with other analytical methods with atomic fluorescence and atomic absorption, ICP–MS detection is more convenient and has higher throughput and efficiency (Rapa et al., 2023). Apples are rich in various mineral elements, among which P, Fe, Mg, and Zn have important nutritional functions for human health (Acquavia et al., 2019; Zhang et al., 2022). In addition, this study reported the contents of various trace elements in apples from the main regions of China, which will be beneficial to the dietary mineral evaluation of apples. Therefore, this study used the ICP–MS method to report the content of 29 elements in apples from the main production regions in China, and the results can lay a foundation for the subsequent study of multiple elements in apples.

Compared with apples from other regions, LP apples have certain characteristics in multielement profiles, suggesting that the level of mineral elements in apples is affected by producing regions. The comparison shows that LP and NLP apples have significant differences in Ca, Mg, Fe, Cu, Mg, Sr, Ba, and Pb (Fig. S1). The mineral element characteristics of apples in LP production region are related to the soil and climate conditions. Previous research has shown that the soil in LP production region is consistent, mainly yellow cotton soil, cinnamon soil, and dark loessial soil, which are formed due to the accumulation of yellow sand brought by the northwest monsoon (Bu, Zhang, Wang, Yang, Shao, & Wu, 2016). In addition, many LP areas have relatively high geographical altitudes, sufficient sunlight, high diurnal temperatures, and dry climates. The NLP mainly contains the regions of the surrounding Bohai Bay, the ancient Yellow River original and the southwest cool highland. Specifically, the surrounding Bohai Bay and the ancient Yellow River original are hilly lands of lower altitude (<300 m) with sandy loam and brown soil, and their climates are hotter and moister (Shen, Zhang, Nie, Zhang, & Bacha, 2022). The southwest cool highland apple-producing region contains parts of Sichuan, Yunnan, and Tibet Provinces, which are mainly hilly land with high altitude and covered with red and low-pH soils (Zhang et al., 2022). Major areas from this region have a climate of high diurnal temperature, arid atmosphere, and intensive sunlight exposure. The differences in environment and climate may be the potential reasons for the mineral element characteristics of apples from different regions.

The results from the discriminate analysis revealed the practicability of LP apple authentication based on inherent differences in multielement profiles. In PCA, the first five PCs scored approximately 60 % of the total variance, which reflects that the main PCs explained the general difference in the data. From the elemental PC scores (Table S4), several elements contributed significantly to the main PCs, indicating that the data variations were correlated to the changes in these elements. PLS-DA, LDA, BP-ANN, and RF were used to construct the discriminate model of LP apples. From the linear models, the training and validation accuracies of the LDA model were generally higher than those of the PLS-DA, indicating that the LDA is more suitable than the PLS-DA for sample classification. In addition, the two linear models obtained a basically consistent ranking of element contribution values (Fig. 3). The elements Be, Co, Mn, Sr, Pb, and Na were critical contributors to building these linear models, which provide a foundation for LDA model optimization. BP-ANN and RF represented the nonlinear models and showed better discrimination rates than those of linear models. Our previous studies also showed that nonlinear models are more practical for recognizing actual data differences with multiple parameters and complex iterative calculations (Zhang et al., 2019; Zhang et al., 2022). The training and validation accuracies of RF were slightly higher than

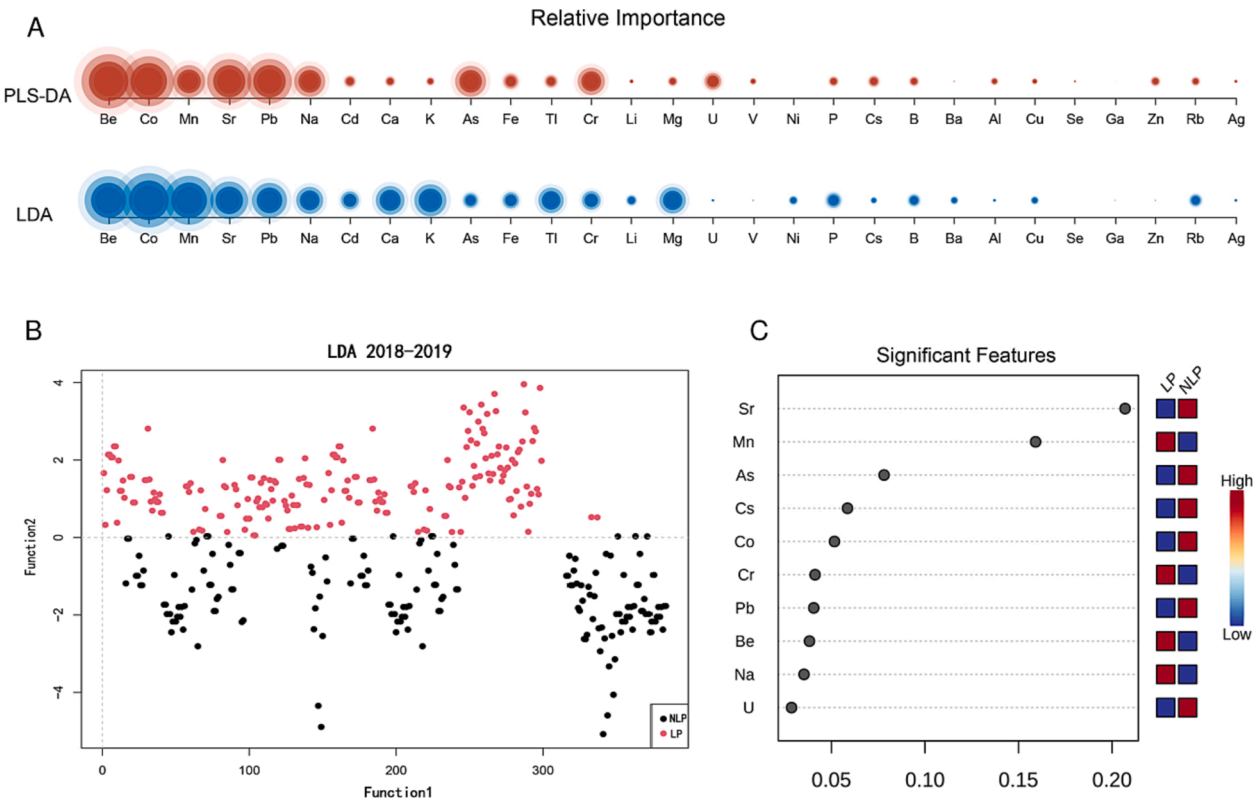


Fig. 3. Linear models for discrimination of apples from Loess Plateau (LP) and non-LP (NLP) regions. A: Relative importance of the elements used in the LDA and PLS-DA models. The relative importance values of elements were obtained from the discriminate modelling based on all elements, which were represented by node area size. B: Optimized LDA of samples from 2018 to 2019 showed separation between LP and NLP samples. C: Relative importance of the elements Sr, Mn, As, Cs, Co, Cr, Pb, Be, Na, and U in the optimized LDA model.

those of BP-ANN, mainly because of the differences in methodology and the different parameters assigned to elements (Maxwell et al., 2018; Wiesmeier, Barthold, Blank, & Kogel-Knabner, 2011). In addition, the critical elements for the RF and BP-ANN models were similar, but the RF model was more focused on Be, Co, Cd, Sr, Li, Pb, Ba, and Ti (Fig. 4). Therefore, linear and nonlinear discriminate models were constructed for the authentication of LP apples, and the comparisons between models were substantial for obtaining an optimized model.

The optimized RF model obtained sufficient efficiency for LP and NLP apple discrimination. From the linear models, LDA obtained high training and validation accuracies based on 29 elements. The reconstructed LDA model can obtain a comparable discriminant rate based on the selected 10 elements (Fig. 3B and C). Therefore, the LDA model is dependent on these important elements involved in modelling (Suzuki, Chikaraishi, Ogawa, Ohkouchi, & Korenaga, 2008). The RF model showed better tolerance to the removal of the less-important elements and obtained acceptable discriminant accuracy based on several important elements. The final RF model was constructed based on 11 elements, which obtained approximately 95 % test accuracy and 97 % verification accuracy. Therefore, this study obtained an optimized RF model for the discrimination of apples from LP and NLP regions from two consecutive years, 2018 and 2019, which was sufficient to distinguish LP and NLP apples in the coming year of 2020. This study obtained a multielement-based RF model for the authentication of apples from the Loess Plateau in China, which laid the foundation for fruit quality and safety control.

5. Conclusion

This study established multielement-based discriminate models for the authentication of LP apples. Comparative analysis showed the different multielement profiles between LP and NLP apples. The LDA, PLS-DA, BP-ANN, and RF discriminate models showed different training and validation accuracies for the discrimination of LP apples. The LDA and RF models were reconstructed based on the values of element contribution. The RF model showed better tolerance to remove the less-important elements. The final RF model was optimized based on 11 elements: Cd, Ca, Be, Mn, Na, Sr, Co, Li, Ba, Tl, and Pb, which obtained a classification accuracy of 95.30 % for 2018 and 2019 apples and a validation accuracy of 97.29 % for 2020 samples. Further studies related to the regulatory mechanism of environmental conditions on apple multielement fingerprints should be explored for theoretical confirmation. This study has successfully established a multielement-based RF model for the geographical authentication of LP apples, which provides a foundation for the quality and safety control of apples and further geographical traceability studies.

CRediT authorship contribution statement

Jianyi Zhang: Methodology, Software, Formal analysis. **Youming Shen:** Conceptualization, Writing – review & editing. **Ning Ma:** Methodology, Software, Writing – review & editing. **Guofeng Xu:** 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.

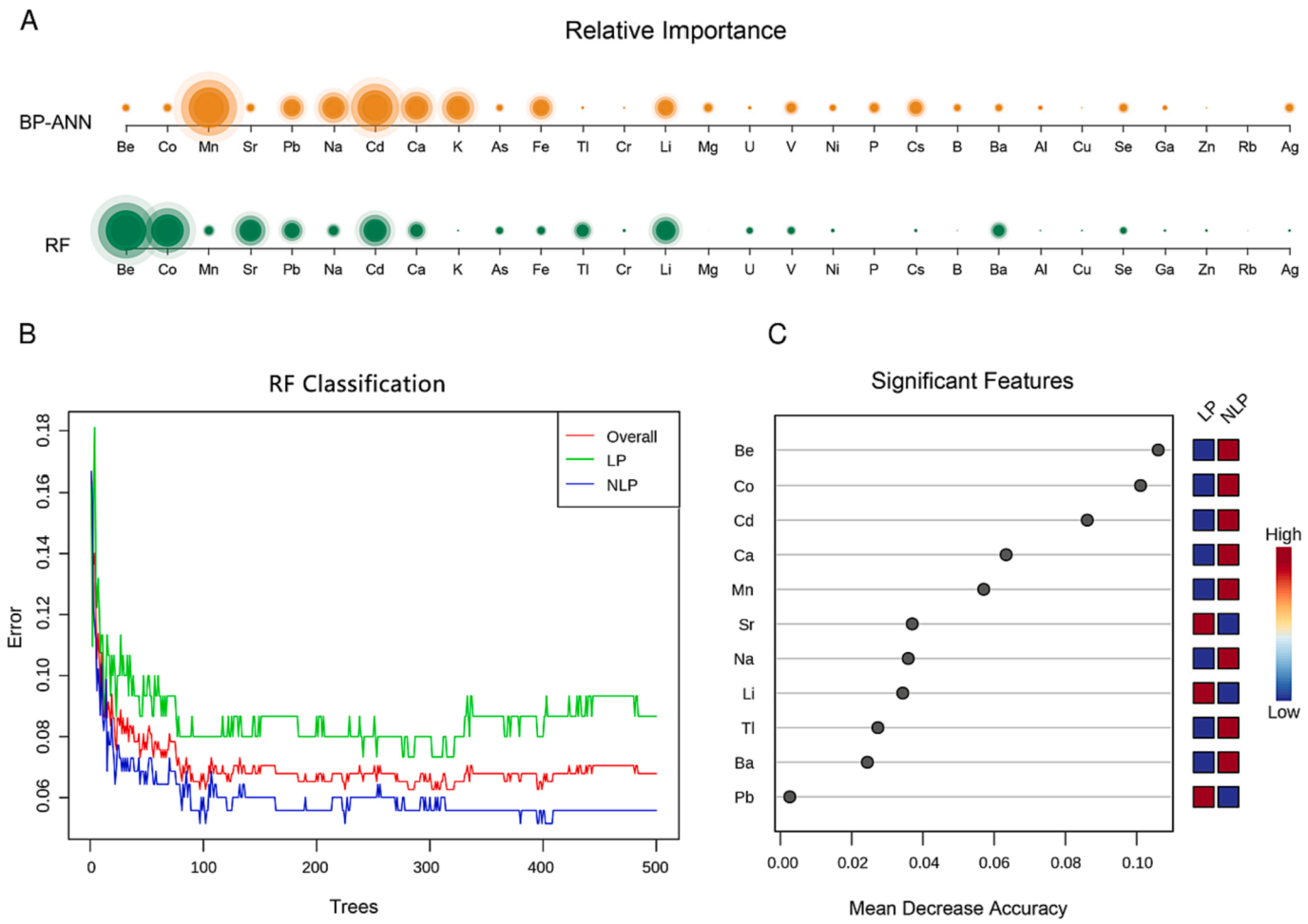


Fig. 4. Nonlinear linear models for discrimination of apples from Loess Plateau (LP) and non-LP (NLP) regions. A: Relative importance of the elements used in the BP-ANN and RF models. B: Optimized RF trees showed the stability of sample classification. C: Relative importance of the elements Cd, Ca, Be, Mn, Na, Sr, Co, Li, Ba, Tl, and Pb in the optimized RF model.

Data availability

No data was used for the research described in the article.

Acknowledgement

We acknowledge Zhen Haidong and other colleagues for their assistance in field sampling and laboratory analyses. This work was supported by the Central Public Interest Scientific Institution Basal Research Fund (Y2023QC) and the Agricultural Science and Technology Innovation Program (CAAS-ASTIP).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.fochx.2023.100948.

References

Acquavia, M. A., Pascale, R., Foti, L., Carlucci, G., Scrano, L., Martelli, G., Brienza, M., Coviello, D., Bianco, G., & Lelario, F. (2021). Analytical methods for extraction and identification of primary and secondary metabolites of apple (*Malus domestica*) fruits: a review. *Separations, 8*(7). Article 91. https://doi.org/10.3390/separations8070091.

Bat, K. B., Eler, K., Mazej, D., Vodopivec, B. M., Mulic, I., Kump, P., & Ogrinc, N. (2016). Isotopic and elemental characterisation of Slovenian apple juice according to geographical origin: Preliminary results. *Food Chemistry, 203*, 86–94. https://doi.org/10.1016/j.foodchem.2016.02.039

Bu, C. F., Zhang, P., Wang, C., Yang, Y. S., Shao, H. B., & Wu, S. F. (2016). Spatial distribution of biological soil crusts on the slope of the Chinese Loess Plateau based on canonical correspondence analysis. *Catena, 137*, 373–381. https://doi.org/10.1016/j.catena.2015.10.016

Cruz, S. M., Schmidt, L., Dalla Nora, F. M., Pedrotti, M. F., Bizzi, C. A., Barin, J. S., & Flores, E. M. M. (2015). Microwave-induced combustion method for the determination of trace and ultratrace element impurities in graphite samples by ICP-OES and ICP-MS. *Microchemical Journal, 123*, 28–32. https://doi.org/10.1016/j.microc.2015.05.008

Dou, X., Zhang, L., Yang, R., Wang, X., Yu, L., Yue, X., … Li, P. (2022). Mass spectrometry in food authentication and origin traceability. *Mass Spectrometry Reviews*. , Article e21779. https://doi.org/10.1002/mas.21779

Giorgia Potorti, A., Francesco Mottese, A., Rita Fede, M., Sabatino, G., Dugo, G., Lo Turco, V., … Di Bella, G. (2022). Multielement and chemometric analysis for the traceability of the Pachino protected geographical indication (PGI) cherry tomatoes. *Food Chemistry, 386*, Article 132746. https://doi.org/10.1016/j.foodchem.2022.132746

Guo, J., Yuan, Y., Dou, P., & Yue, T. (2017). Multivariate statistical analysis of the polyphenolic constituents in kiwifruit juices to trace fruit varieties and geographical origins. *Food Chemistry, 232*, 552–559. https://doi.org/10.1016/j.foodchem.2017.04.037

Jandric, Z., Tchaikovsky, A., Zitek, A., Causon, T., Stursa, V., Prohaska, T., & Hann, S. (2021). Multivariate modelling techniques applied to metabolomic, elemental and isotopic fingerprints for the verification of regional geographical origin of Austrian carrots. *Food Chemistry, 338*, Article 127924. https://doi.org/10.1016/j.foodchem.2020.127924

Juniper, B. E., Watkins, R., & Harris, S. A. (1998). The origin of the apple. In Eucarpia Symposium on Fruit Breeding and Genetics. St Catherine Coll, Oxford, England. *484*, 27-33. https://doi.org/10.17660/ActaHortic.1998.484.1.

Kuang, L., Nie, J., Li, Y., Cheng, Y., & Shen, Y. (2020). Quality evaluation of 'Fuji' apples cultivated in different regions of China. *Scientia Agricultura Sinica, 53*(11), 2253–2263. https://doi.org/10.1016/10.3864/j.issn.0578-1752.2020.11.011

Li, C., Kang, X., Nie, J., Li, A., Farag, M. A., Liu, C., … Yuan, Y. (2023). Recent advances in Chinese food authentication and origin verification using isotope ratio mass

spectrometry. *Food Chemistry, 398*, Article 133896. https://doi.org/10.1016/j.foodchem.2022.133896

Longobardi, F., Casiello, G., Centonze, V., Catucci, L., & Agostiano, A. (2017). Isotope ratio mass spectrometry in combination with chemometrics for characterization of geographical origin and agronomic practices of table grape. *Journal of the Science of Food and Agriculture, 97*(10), 3173–3180. https://doi.org/10.1002/jsfa.8161

Luykx, D. M. A. M., & Van Ruth, S. M. (2008). An overview of analytical methods for determining the geographical origin of food products. *Food Chemistry, 107*(2), 897–911. https://doi.org/10.1016/j.foodchem.2007.09.038

Maione, C., & Barbosa, R. M. (2019). Recent applications of multivariate data analysis methods in the authentication of rice and the most analyzed parameters: A review. *Critical Reviews in Food Science and Nutrition, 59*(12), 1868–1879. https://doi.org/10.1080/10408398.2018.1431763

Maxwell, A. E., Warner, T. A., & Fang, F. (2018). Implementation of machine-learning classification in remote sensing: An applied review. *International Journal of Remote Sensing, 39*(9), 2784–2817. https://doi.org/10.1080/01431161.2018.1433343

Rapa, M., Ferrante, M., Rodushkin, I., Paulukat, C., & Conti, M. E. (2023). Venetian protected designation of origin wines traceability: Multi-elemental, isotopes and chemometric analysis. *Food Chemistry, 404*, Article 134771. https://doi.org/10.1016/j.foodchem.2022.134771

Shen, Y., Zhang, J., Nie, J., Zhang, H., & Bacha, S. A. S. (2022). Apple microbial communities and differences between two main Chinese producing regions. *Food Quality and Safety, 6*, Article fyab033. https://doi.org/10.1093/fqsafe/fyab033

Suzuki, Y., Chikaraishi, Y., Ogawa, N. O., Ohkouchi, N., & Korenaga, T. (2008). Geographical origin of polished rice based on multiple element and stable isotope analyses. *Food Chemistry, 109*(2), 470–475. https://doi.org/10.1016/j.foodchem.2007.12.063

von Wuthenau, K., Segelke, T., Müller, M.-S., Behlok, H., & Fischer, M. (2022). Food authentication of almonds (Prunus dulcis mill.). Origin analysis with inductively coupled plasma-mass spectrometry (ICP-MS) and chemometrics. *Food Control, 134*, Article 108689. https://doi.org/10.1016/j.foodcont.2021.108689

Wiesmeier, M., Barthold, F., Blank, B., & Kogel-Knabner, I. (2011). Digital mapping of soil organic matter stocks using Random Forest modeling in a semi-arid steppe ecosystem. *Plant and Soil, 340*(1–2), 7–24. https://doi.org/10.1007/s11104-010-0425-z

Xu, L., Xu, Z., Wang, X., Wang, B., & Liao, X. (2020). The application of pseudotargeted metabolomics method for fruit juices discrimination. *Food Chemistry, 316*, Article 126278. https://doi.org/10.1016/j.foodchem.2020.126278

Zhang, J. Y., Nie, J. Y., Kuang, L. X., Shen, Y. M., Zheng, H. D., Zhang, H., … Asim, S. (2019). Geographical origin of Chinese apples based on multiple element analysis. *Journal of the Science of Food and Agriculture, 99*(14), 6182–6190. https://doi.org/10.1002/jsfa.9890

Zhang, J. Y., Nie, J. Y., Zhang, L. B., Xu, G. F., Zheng, H. D., Shen, Y. M., … Zhang, H. (2022). Multielement authentication of apples from the cold highlands in southwest China. *Journal of the Science of Food and Agriculture, 102*(1), 241–249. https://doi.org/10.1002/jsfa.11351

Zhang, Q., Zhou, B. B., Li, M. J., Wei, Q. P., & Han, Z. H. (2018). Multivariate analysis between meteorological factor and fruit quality of Fuji apple at different locations in China. *Journal of Integrative Agriculture, 17*(6), 1338–1347. https://doi.org/10.1016/S2095-3119(17)61826-4

Zhao, H. Y., Tang, J., & Yang, Q. L. (2021). Effects of geographical origin, variety, harvest season, and their interactions on multi-elements in cereal, tuber, and legume crops for authenticity. *Journal of Food Composition and Analysis, 100*, Article 103900. https://doi.org/10.1016/j.jfca.2021.103900