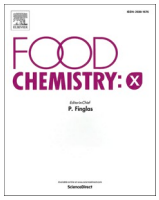
Food Chemistry: X 25 (2025) 102190

Contents lists available at ScienceDirect

Food Chemistry: X

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



Enhanced multivariate data fusion and optimized algorithm for comprehensive quality profiling and origin traceability of Chinese jujube

Peng Chen [a], Xiaoli Wang [e], Rao Fu [b], Xiaoyan Xiao [d], Yu Li [b], Tulin Lu [b], Tao Wang [a], Qiaosheng Guo [a,*], Peina Zhou [c,*], Chenghao Fei [a,*]

[a] Institute of Chinese Medicinal Materials, Nanjing Agricultural University, Nanjing 210095, China
[b] College of Pharmacy, Nanjing University of Chinese Medicine, Nanjing 210023, China
[c] Institute of Plant Resources and Chemistry, Nanjing Research Institute for Comprehensive Utilization of Wild Plants, Nanjing 210042, China
[d] Suzhou Liliangji Health Industry Co., Ltd., Suzhou 215000, China
[e] Changzhou Affiliated Hospital, Nanjing University of Chinese Medicine, Changzhou 213003, China

ARTICLE INFO

Keywords:
Chinese jujube
UF-GC-*E*-nose
GC–MS
Multivariate statistics
Intelligent algorithm
Traceability

ABSTRACT

Chinese jujube (CJ) is a nutritious food. Its authenticity has received increasing attention. This research utilized computer vision, ultrafast gas-phase electronic nose, and GC–MS technologies to collect jujube samples from various regions in China, including Xinjiang, Gansu, Shaanxi, Henan, Shandong, and Hebei. Multidimensional trait data, encompassing spectra, texture, and odour, were gathered. By employing multivariate statistical methods, 46 trait characteristic factors (VIP > 1, $P < 0.05$) were identified and utilized to rapidly differentiate jujube samples originating from different regions. The multivariate statistical analysis and support vector machine (SVM) classification were also combined to develop a novel artificial intelligence algorithm. The accuracy of this innovative method was significantly higher than that of conventional discriminant analysis methods, achieving a perfect 100.0 % accuracy. As a consequence of this research, more intelligent algorithms can be developed that trace the origin of food based on multidimensional data.

1. Introduction

Ziziphus jujuba (*Ziziphus jujuba* Mill.) is the dried ripe fruit of the jujube plant of the Rhamnaceae family, with more than 1000 subspecies identified (Zhao et al., 2023). As a widely distributed crop, Jujubes grow mainly in subtropical and tropical regions of Asia and the Americas, as well as in the Mediterranean region. Although Jujube plants require hot summers and plenty of water to bear fruit, they are extremely adaptable to a wide range of temperatures and rainfall conditions (Song et al., 2019). Chinese jujube (CJ), also known as Chinese red jujube, is a subspecies of jujube that has been cultivated and domesticated in the Yellow River basin of China for 4000 years (Chen et al., 2023). CJ is not only consumed as a homogenous medicinal food but also used as a traditional Chinese medicine, which is appreciated and sought after for its good taste, high nutritional value and medicinal benefits. Studies have shown that CJ has anticancer, antiepileptic, anti-inflammatory, anti-insomnia and neuroprotective effects (Ji et al., 2023).

CJ cultivation spans China, with an estimated 20,000 ha of land dedicated to its production. Cash crops are predominantly grown in northern China, particularly in the historic cultivation regions located in the eastern portion of the middle and lower Yellow River basin (Henan, Hebei, Shandong, and Shaanxi), as well as in the production regions within arid western environments (Xinjiang and Gansu) (Shen, Wang, & Lyu, 2023). Two common forms of CJ are fresh and dry, and 70 % of CJ is sold as dry fruit (Sun, Sun, Zhang, & Liu, 2016; Yang et al., 2021). Differences in natural environmental conditions such as temperature, light intensity, precipitation, and soil conditions lead to differences in the organoleptic quality, nutrient content, and flavour of jujubes obtained from different origins (Duan et al., 2023). There is confusion over the geographical origin of commercially available CJs and the phenomenon of impersonation. At the same time, growing awareness of health issues and concerns about adulteration, germplasm mixing and origin fraud suggest that food traceability is likely to be a prominent aspect of food in the future (Islam & Cullen, 2021). Therefore, CJ

Abbreviations: CJ, Chinese jujube; UF-GC-E-nose, ultra fast gas phase electronic nose; LBP, Local Binary Pattern; DBN, Deep Belief Network; PCA, principal component analysis; PLS-DA, Partial Least Squares Discriminant Analysis; VIP, variable importance factor; MDF, multivariate data fusion..

* Corresponding authors.
E-mail addresses: gqs@njau.edu.cn (Q. Guo), zhoupeina@163.com (P. Zhou), feichenghao@njau.edu.cn (C. Fei).

https://doi.org/10.1016/j.fochx.2025.102190
Received 19 October 2024; Received in revised form 7 January 2025; Accepted 13 January 2025
Available online 18 January 2025

requires a rapid, feasible, and reliable geographic traceability method.

Food origin is typically determined by appearance, color, and taste, but these methods are subjective and inaccurate (Shin, Kim, Choi, Kim, & Han, 2023). New technologies utilizing external features are now being used in food safety and quality control to meet modern demands. The use of near-infrared (NIR) techniques has been used for assessing the wine-steaming process of Polygonati Rhizome, predicting the moisture content of black tea during drying, and other studies (Chen et al., 2024). Despite its advantages, the NIR technique is limited by the requirement for samples to be prepowdered before analysis, resulting in potential damage to the sample. In contrast, machine vision analysis offers a nondestructive testing approach that allows for the acquisition of color, texture, and shape information from samples. UF-GC-*E*-nose technology combines headspace, electronic nose, and gas chromatography to capture subtle differences in food odours and has been used to reveal changes in volatile flavour compounds during the boiling of salted ducks, to characterize the flavour of raw milk from different regions of China, to differentiate barley malt from different origins, and to identify pu-erh tea with different storage ages (Chi, Zhang, Zheng, Wang, & Liu, 2023; Li, Al-Dalali, Wang, Xu, & Zhou, 2022; Ma et al., 2024; Rong et al., 2023). The complexity of food makes it difficult to judge food consumption using one method. Classification algorithms combine data from various sensory techniques for better judgement. At that time, Support Vector Machine (SVM) was a commonly used algorithm in this field for classification and regression tasks (Yan, Wang, & Lei, 2022; Zhang et al., 2023). The basic concept of SVM is to determine the hyperplane by maximizing the interclass interval, and support vectors play a key role in determining the optimal classification hyperplane, which performs powerfully and generalizes well in practical situations for high-dimensional datasets and small sample data. At present, the feasibility of optimizing the SVM classification algorithm has not yet been proven. This optimization aims to determine CJ origin traceability. It involves fusing spectral, texture, UF-GC-*E*-nose, and GC–MS multidimensional feature data.

The objective of this research was to gather multidimensional feature data, including spectrum, texture, and odour data, from CJ samples originating from various regions (Xinjiang, Gansu, Shaanxi, Henan, Shandong, and Hebei) via computer vision, UF-GC-E-nose and GC–MS. To analyse the collected data, multivariate statistical techniques and data fusion methods were utilized to establish a comprehensive database. Furthermore, CJ samples were traced to their respective origins by using multivariate data fusion and artificial intelligence algorithms. Methods for classifying and predicting CJ samples from different regions can be developed quickly.

2. Materials and methods

2.1. Sample collection

One hundred and twenty CJ samples from six producing areas in North China were collected as the plant materials. Twenty batches were collected from each producing area, including Xinjiang (XJ), Gansu (GS), Shaanxi (SN), Henan (HN), Hebei (HB) and Shandong (SD). The fruits at the mature stage were collected and uniformly dried. The samples were initially washed with water to remove any sediment and impurities. Subsequently, the samples were subjected to a drying process at 45 °C for 1 h in order to remove the surface moisture. In order to ensure sample stability, the prevention of component decomposition, as well as improvements in analytical accuracy and reproducibility. The dried jujube fruits were stored at −20 °C. Detailed information on the samples is given in Table S1. XJ-CJ, GS-CJ, SN-CJ, HN-CJ, HB-CJ, or SD-CJ are the unique geo-labels given to each sample. These latitudes and longitudes range from 33.63°N to 41.17°N and from 79.93°E to 117.97°E, respectively. Fig. 1-A shows details and photographs of the sampling locations. RESTEK Co., Ltd. provided us with a mixture of n-alkanes (nC6 ~ nC16, batch number A0142wq987).

2.2. Computer vision-based data collection

2.2.1. Python algorithm for extracting sample spectral information

To extract spectral information from CJ samples, a Python-based algorithm was developed. Initially, the CJ samples were captured using auto white balance and auto exposure settings, ensuring no color gain adjustments were made during the imaging process. Using a platform under the camera, samples were positioned directly below the camera, and images in PNG format were captured. For each sample batch, 100 images were taken. Red, green, and blue spectral data from these images were analysed using a Python script. Each color channel was assigned specific wavelength ranges, and the wavelength arrays were standardized in length. A total of 120 datasets were generated from this analysis, comprising one assay, 20 batches, and 6 categories. Table S2 details how the Python algorithm can be used to extract spectral information.

2.2.2. Python algorithm for extracting sample texture information

The same batch of CJ photographs used in section “2.2.1”. The GLCM is a method used to characterize the texture of an image. It extracts the texture information of an image by counting the frequency and distribution of occurrences of adjacent pixel pairs in the image. To compute the GLCM, it is first necessary to select a specific direction and distance and then process each pixel in the image. For each pixel, the frequency of occurrence of its surrounding pixels at the gray level is counted, and a

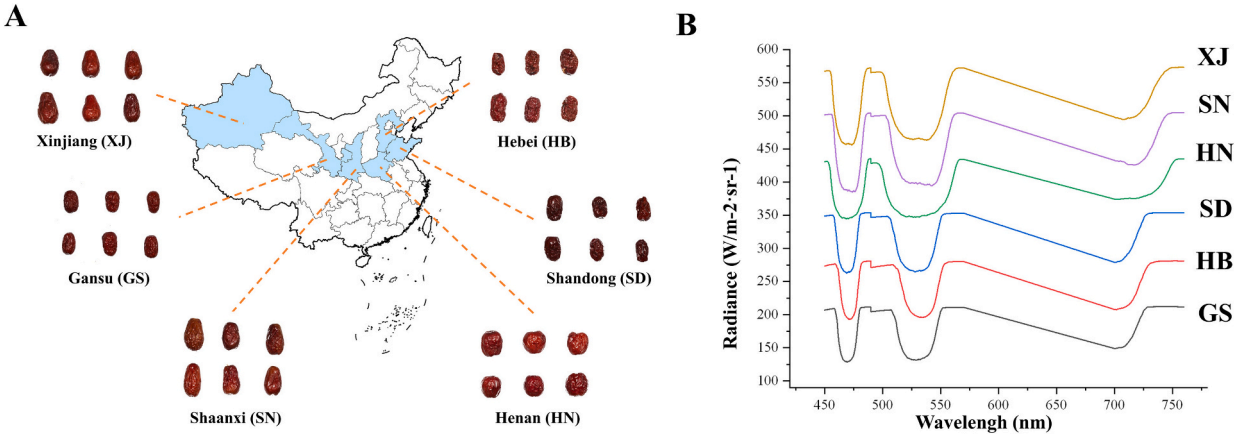


Fig. 1. Geographic information (A) and spectral characterization (B) of CJ samples. GS, Gansu; HB, Hebei; SD, Shandong; SN, Shaanxi; HN, Henan; XJ, Xinjiang.

matrix representing these frequencies of occurrence is constructed. Texture features of an image can be derived by computing statistical attributes of the GLCM, such as energy, contrast, correlation and other metrics. The GLCM can be applied in many fields, such as medical image analysis, geological image analysis, and remote sensing image analysis. By extracting the texture features of an image, the representation of the image can be better understood, thus enabling deeper analysis and application of the image. A total of 120 datasets (1 detection × 20 batches × 6 categories) were obtained for computer texture. In Table S3, the specific steps of the Python texture extraction algorithm are described.

2.3. GC–MS analysis

2.3.1. GC–MS analysis conditions

The CJ sample powder (2.0 g) was placed in a 10-mL headspace vial and equilibrated at 50 °C for 40 min. The automatic syringe extracted the gas above the headspace vial and injected it into the instrument for analysis. Using an Agilent 7890B gas chromatograph (Agilent Technologies, Santa Clara, CA, USA), the temperature of the column was programmed as follows: 40 °C was maintained for 3 min, increased to 120 °C at 5 °C/min, increased to 200 °C at 10 °C/min, and then maintained at a constant temperature for 5 min. Helium was used as the carrier gas at a flow rate of 1.0 mL/min. An Agilent 7000C triple quadrupole mass spectrometer (Agilent Technologies, Santa Clara, CA, USA) was used for the MS analyses, and the MS was operated in full scanning mode under the following conditions: ion source and transmission line at 250 °C, electron impact source at 70 eV, mass range of *m/z* 35–550, and scan rate of 2000 (Chen, Song, Bi, Meng, & Wu, 2018).

2.3.2. Identification of volatile organic compounds (VOCs)

Volatile compounds were initially identified using the NIST 14 mass spectral library and confirmed based on the 7500 Compound Ion Analysis database as well as literature spectra of date-specific compounds and reference standards. RI-NIST14 indicates that retention indices (RIs) were determined based on the NIST 14 library for the GC–MS system. The aroma profiles were based on the literature and web-based descriptor sites (http://www.thegoodscentscompany.com/) (Wu, Zhan, Tang, Li, & Duan, 2022).

2.4. UF-GC-E-nose analysis

2.4.1. Sample incubation

Incubation temperature, stirring rate, and incubation time were determined using automatic headspace injection. During incubation, the sample odour components evaporated and accumulated in the upper portion of a closed gas phase vial. To incubate the powder sample, a robotic arm was used to transport it to the incubator. A stirring speed of 500 rpm and a stirring time of 7 s were selected, each cycle lasting 5 s (5 s of stirring and 2 s of pausing), the temperature was set at 45 °C, and the incubation time was 20 min.

2.4.2. Sample detection

Sample detection parameters included 250 L injection volume, 125 L/s injection rate, 20 s of injection time, and 200 °C inlet and trap temperatures. The initial temperature of the column oven was 45 °C, and the programmed temperature increase rate was 5 °C per second to 80 °C, 2 °C per second to 120 °C, 1 °C per second to 160 °C, and 2 °C per second to 180 °C. The data acquisition time was 120 s, with an acquisition period of 0.01 s, and the carrier gas was hydrogen at a flow rate of 1 mL/min. Three measurements were taken for each batch of CJ samples, and the results were averaged.

2.4.3. Analysis of volatile compounds

DB-5 and DB-1701 Kovats retention indices were analysed for the sample peaks from the ArochemBase database, and the data were used to identify the corresponding compounds.

2.5. Analysis of support vector machine (SVM)

SVM is a commonly used supervised learning algorithm for classification and regression analysis (Yan et al., 2022). It uses a kernel function to map data points into a high-dimensional space, making them more linearly separable. An optimal hyperplane is then found to separate the classes and maximize the distance to the nearest data point. The support vectors determine the final classification boundaries. SVMs are effective for high-dimensional datasets and can handle situations where the feature dimensionality is greater than the number of samples. It can also handle nonlinear classification problems with the right kernel function. Memory usage can be reduced with support vectors during training. SVM is a powerful algorithm for small sample and high-dimensional datasets. Multivariate statistical analysis was used to quantify and optimize the multidimensional features in CJ samples, then used as an input layer in SVM classification. Six nominal output variables were used to accomplish the classification task: 1 for the GS-CJ samples, 2 for the HB-CJ samples, 3 for the SD-CJ samples, 4 for the HN-CJ samples, 5 for the SN-CJ samples, and 6 for the XJ-CJ samples.

2.6. Data analysis

For the extraction of spectra and texture features, Python software 3.6.8 (Waverley Software, USA) was used. A mean + standard deviation (SD) is used to express data. An analysis of the data was performed using SPSS software (version 22, SPSS Inc.). An analysis of variance was used to analyse the data, and Duncan's multiple-range test and the least significant difference test were used to identify statistically significant differences ($P < 0.05$). The Principal Component Analysis (PCA) was conducted using Simca software (version 14.1, Sartorius Group, SW), and the Partial Least Squares Discrimination Analysis (PLSDA) was performed using GraphPad Prism (version 9, GraphPad Software, San Diego, CA, USA). Support Vector Machine (SVM) algorithm analysis was carried out using MATLAB software (version R2023a, MathWorks, Inc., USA).

3. Results and discussion

3.1. Spectral characterization

Fig. 1-B presents the analysis of visible spectra from CJ samples originating from six distinct locations, it is evident that these samples exhibit consistent absorbance bands and display comparable patterns (Chen et al., 2024). Specifically, all samples exhibit three distinct peaks at wavelengths of 460–490 nm, 520–550 nm, and 690–720 nm, although variations in absorption intensity are observed among samples originating from different sources. At the lowest peaks of 460–490 nm and 520–550 nm wavelength ranges, CJ samples from the SX origin show the largest spectral fluctuations, while those from the GS origin exhibit the smallest fluctuations. In the spectral analysis of CJ samples from various origins, it was observed that the spectral fluctuation in the lower peak range of 690 to 720 nm was most pronounced in samples of SD origin and least pronounced in samples of HN origin. Furthermore, all CJ samples from the six origins exhibited maximum absorption peaks in the range of 480–510 nm. The fluctuation ranges of the blue and green absorption peaks were similar across samples of different origins, while the red absorption peak generally exhibited a greater intensity than did the blue and green absorption peaks. According to the RGB principle of three primary colours (Erba, Buzzelli, & Schettini, 2024), the CJ samples predominantly absorb blue and green light while reflecting red light. Therefore, the CJ samples appear red because the human eye perceives the reflected red light. This finding reinforces the reliability and practicality of utilizing Python algorithms for extracting spectral information from samples.

3.2. Characterization of texture

In order to describe the texture of an image, the GLCM is commonly used as a texture computation method (Yang, He, Li, Liu, & Lan, 2022). The Python platform was utilized to conduct sliding window computation for the successful extraction of GLCM texture features from the CJ samples. Throughout the investigation, a 3 × 3 computation window was employed to maintain the precision and dependability of texture extraction.

The GLCM provides a method for describing the texture features of an image by calculating the statistical information between different gray levels to analyse the texture features of an image. Where Contrast is the contrast, Dissimilarity is the degree of difference, Homogeneity is the degree of similarity, Energy is the degree of energy, Correlation is the correlation between pixels, and ASM is a measure of texture thickness. These parameters can help us to analyse and understand the texture characteristics of an image, leading to a deeper understanding and application in image processing and analysis. The stacked plots of the GLCM feature parameters extracted from CJ samples of six different origins (XJ, SN, HN, SD, HB, and GS) are shown in Fig. 2. As seen from the figure, by comparing the contrast, dissimilarity, energy and ASM values, it is found that the larger the values of the above parameters are, the deeper the grooves representing the sample texture and the rougher the sample surface. The specific magnitudes of contrast, dissimilarity, energy and ASM values for CJ samples of different origins are as follows: SD > HB > GS > HN > XJ > SN; HB > GS > SD > XJ > SN > HN; HB > GS > SD > HN > SN > XJ; and HB > GS > SD > HN > SN > XJ. This indicates that the surface texture of the CJ samples from the SD, HB and GS origins is rougher and that the texture grooves are deeper than those of the CJ samples from the HN, XJ and SN origins. In addition, based on the analysis of homogeneity and correlation values of CJ samples of different origins, the order of homogeneity values of CJ samples of different origins was GS > HB > SD > SN > XJ > HN, and the order of correlation values was GS > SN > XJ > HN > HB > SD. The smaller the difference between the above two parameters of CJ samples between origins is, the more similar the samples' surface textures are, and the larger the difference is, the more different the samples' surface textures are. According to the results, the surface textures of CJ samples from SD, HB, and GS origins are more similar to each other. However, they are rougher compared to the surface textures of CJ samples from HN and XJ origins. There were significant differences in the surface texture characteristics of CJ samples from different origins, which may be influenced by the growing environment and treatment methods. These results provide an important reference for CJ quality evaluation and scientific data support for CJ origin identification and quality control.

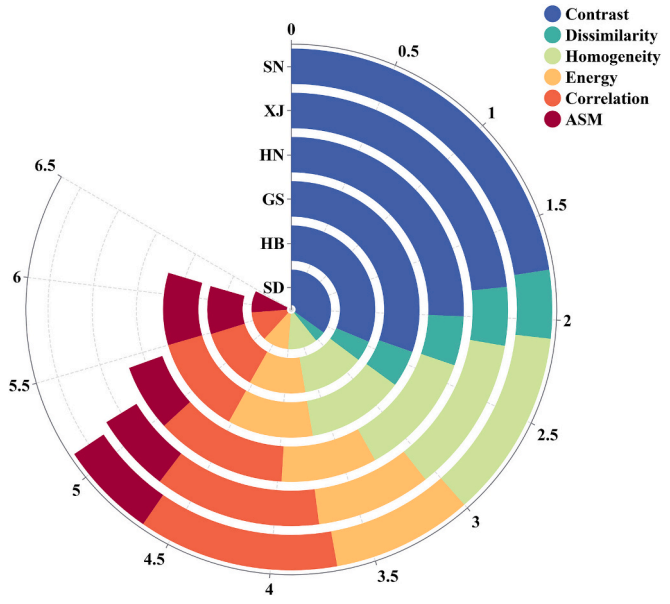


Fig. 2. Texture features of CJ samples. GS, Gansu; HB, Hebei; SD, Shandong; SN, Shaanxi; HN, Henan; XJ, Xinjiang.

3.3. GC–MS-based quantitative analysis of odours in CJs from various origins

A total of 120 batches of CJ from various sources were subjected to analysis using GC–MS under the specified test conditions outlined in section "2.3". The chromatograms acquired through GC–MS total ion flow analysis are illustrated in Fig. S1. The quantities of various components in CJ samples from different origins analysed by GC–MS are shown in Fig. 3-A. As shown in the figure, in CJ samples from GS, the highest number of acid components (12) was found, followed by esters (4), and the lowest number of alcohols, aldehydes and ketones (3) were detected. In the CJ samples from HB, the greatest number of acids (8), followed by alcohols and aldehydes (5), 4 esters and 3 ketones were also found. Among the CJ samples of SD, the highest number of ester components (5), followed by alcohol and acid components (4), and the lowest number of aldehydes and ketones (3) were detected. Among the CJ samples from HN, the greatest number of acid components (8), followed by alcohols, aldehydes and esters (3), and the lowest number of ketones (1) were detected. Among the CJ samples from SN, the highest number of acid components (12), followed by aldehydes and esters (3), and the lowest number of alcohols and ketones (2) were detected. In the CJ samples from XJ, the highest number of acid components (9), followed by ester components (4), 3 alcohol components were also found, and the lowest number of aldehydes and ketones (2) were detected. With 25 components identified each, the CJ samples from GS and HB had the highest number of volatile components. Only 17 volatile components were detected in the CJ samples from the SD group.

A Venn diagram was constructed based on the specifics of the volatile components of CJ from each origin by GC–MS, and the results are shown in Fig. 3-B and Table S4. The analysis of the figure revealed that the CJ samples from six different origins contained a total of 4 shared volatile components, namely, isobutyric acid, butyric acid, 2-furanmethanol and pyran-2-one. Among the six samples, only the CJ samples from GS exhibited distinct volatile components. Specifically, the CJ samples from GS contained the unique volatile component 2,3-dihydro-3,5-dihydroxy-6-methyl-4H-pyran-4-one. The volatile components shared between the two origins were as follows: GS and HB shared 2,3-butanediol and 3-hydroxy-2H-pyran-one. The volatile components shared between the three origins were as follows: GS, SD and HN shared propyl pentanoate; GS, SD and XJ shared decyl butanoate; GS, HB and HN shared 2-butenal; XJ, HB and HN shared nonanal; GS, HB and XJ shared heptanoic acid; GS, HB and SN shared 2-methylbutyric acid and acetic acid; GS, HN and SN shared dodecanoic acid and hexanoic acid; and GS, SN and XJ shared benzoic acid. The volatile components shared between the four origins were as follows: HB, SD, SN and XJ shared methyl decanoate, undecanal, 2-octenoic acid and 2,3-dihydro-3,5-dihydroxy-6-methyl-4H-p[yran-4-one; HB, SD, SN and GS shared hexanal and (Z)-2-penten-1-ol; HB, SD, GS and XJ shared methyl laurate; HB, SD, HN and XJ shared oct-1-en-3-ol and 1,2-ethylene glycol; HB, GS, HN and SN shared methyl hexanoate and 5-hydroxymethyl-2-furfural; and GS, HN, SN and XJ shared decanoic acid. The volatile components shared between the five origins were as follows: GS, SD, HN, SN and XJ shared octanoic acid; HB, SD, HN, SN and XJ shared methyl octylate; and HB, GS, HN, SN and XJ shared propanoic acid and 3-hydroxybutyric acid.

As shown in Fig. 3-C, the volatile components represent modular units in the pie chart comparing CJ samples from six different origins by gas chromatography/mass spectrometry. In the CJ sample from GS, the most dominant volatile component was aldehydes (68.94 %), followed by acids (25.86 %) and esters (5.04 %). In the CJ sample from HB, the dominant volatile component was aldehydes (49.29 %), followed by

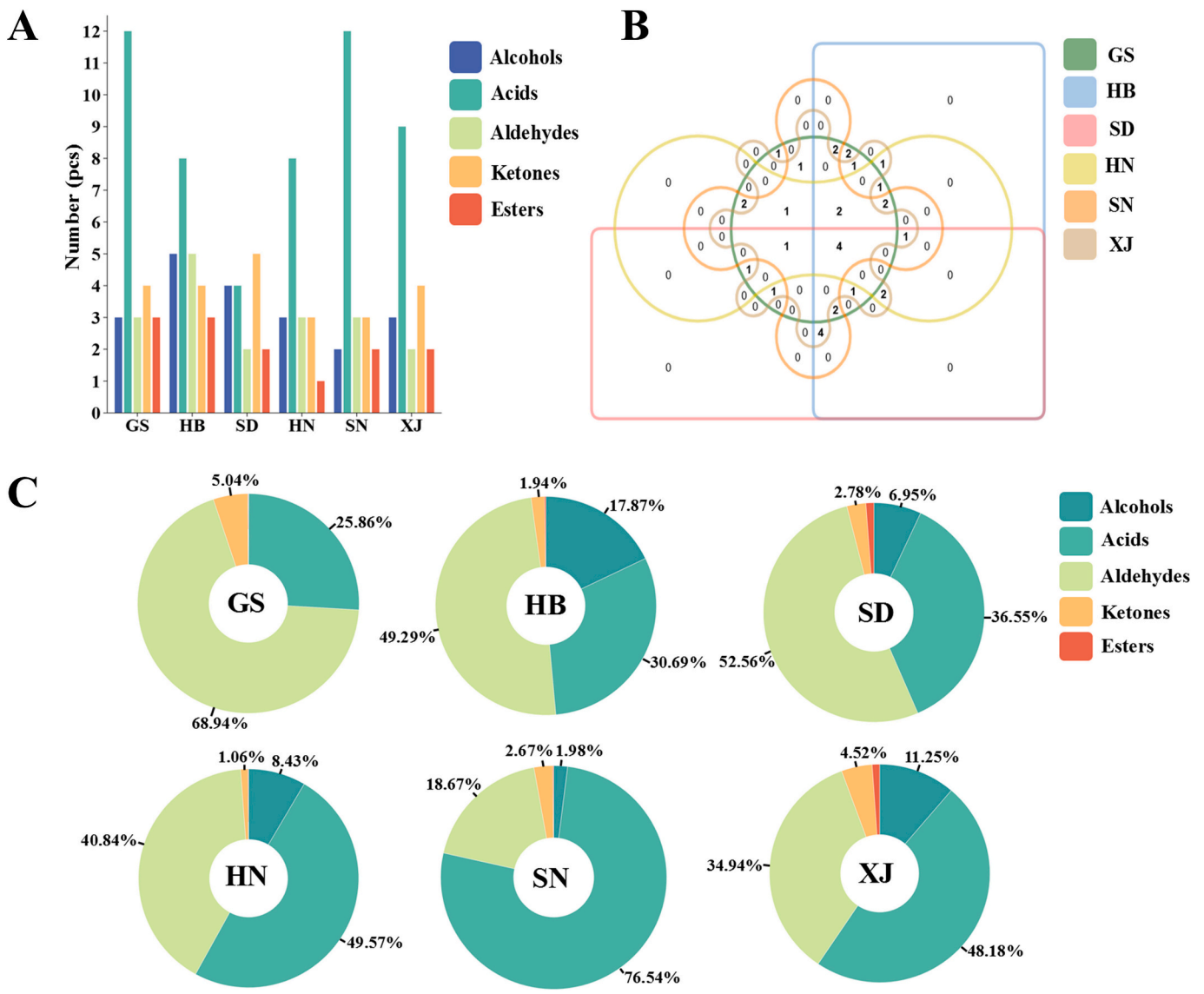


Fig. 3. Histogram of the number of each type of constituent (A), pie charts (A) and Wayne plots (B) of the relative amounts of odour components in CJ samples by GC–MS. GS, Gansu; HB, Hebei; SD, Shandong; SN, Shaanxi; HN, Henan; XJ, Xinjiang.

acids (30.69 %) and esters (17.87 %). In the CJ sample from SD, the most dominant volatile component was aldehydes (52.56 %), followed by acids (36.55 %) and esters (6.95 %). In the CJ samples from HN, the three most abundant volatile components were acids (49.57 %), aldehydes (40.84 %) and alcohols (8.43 %). In the CJ samples from SN, the three categories with the highest relative contents of volatile components were acids (76.54 %), aldehydes (18.67 %) and esters (2.67 %). The CJ sample from XJ was also dominated by acid volatiles (48.18 %), followed by aldehydes (34.94 %) and alcohols (11.25 %). The combined relative content of aldehydes and acids exceeded 80 % of the total composition across the six sample types. CJ samples from different origins, including GS, HB, SD, HN, SX and XJ, had different aldehyde and acid compositions. Specifically, samples from the GS, HB and SD origins had higher relative aldehyde contents due to their smaller size, while samples from the HN, SX and XJ origins had higher relative acid contents due to their larger size (Shi et al., 2022).

3.4. *Analyzing the odours of CJs of different origins using UF-GC-E-nose*

A total of 120 batches of CJ from various sources were subjected to analysis using the UF-GC-E-nose under the specified test conditions outlined in section “2.4”. The chromatograms obtained on the MXT-5 and MXT-1701 columns of the UF-GC-E-nose are shown in Fig. S2. UF-GC-E-nose analyses of the quantities of various components in CJ samples from different origins are shown in Fig. 4-A. In the CJ samples from GS, the number of esters was the highest (8), followed by acids (6), ketones (5), alcohols (4) and aldehydes (4), and the number of other components was the lowest (2). In the CJ samples of HB, the number of esters and acids was the greatest (7), followed by alcohols, aldehydes and ketones (5), and the lowest number of other components (4). Among the CJ samples of SD, the highest number of esters (10) followed by acids (7), alcohols (4) and aldehydes (4) and the lowest number of ketones (3). In the CJ samples of HN, the greatest number of ester components (12) followed by acid components (10), alcohols with ketones (5), and then aldehydes (4), and the least number of other components (2) were detected. In the CJ samples of SN, the highest number of ester components (10) followed by acid components (8), ketone components (5), and the lowest number of alcohols, aldehydes and others (4) were detected. In the CJ samples from XJ, the most ester components (11) were followed by acid components (8), alcohols, aldehydes and ketones (3), and the least number of other components (2). In addition, the total number of volatile components in the CJ samples from HN was greatest, with 38

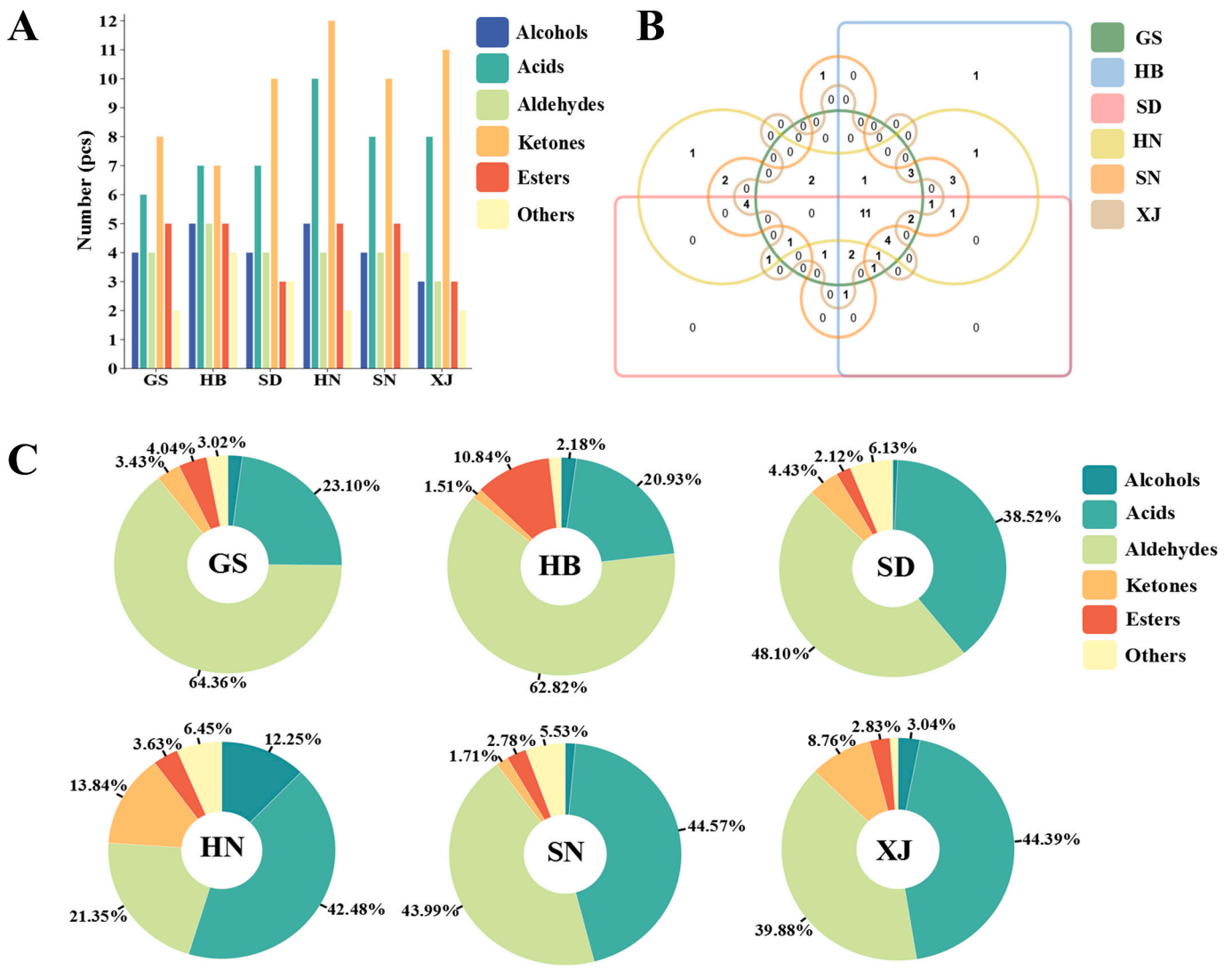


Fig. 4. Histogram of the number of each type of constituent (A), pie charts (A) and Wayne plots (B) of the relative amounts of odour components of CJ samples by UF-GC-E-nose. GS, Gansu; HB, Hebei; SD, Shandong; SN, Shaanxi; HN, Henan; XJ, Xinjiang.

components identified. CJ samples from GS had the lowest total number of volatile components, with only 29.

Moreover, a Venn diagram was constructed based on the specifics of the volatile components of CJ from each origin by the UF-GC-*E*-nose, and the results are shown in Fig. 4-B and Table S5. The analysis of the figure revealed that the CJ samples from six different origins contained a total of 12 shared volatile components, including γ-decalactone, decyl butanoate, butyl cinnamate, octanal, hexanoic acid, octanoic acid, ethanol, 2-methyl-1-pentnol, 2-furanmethanol, 2-dodecanone, 2-tridecanone and rheosmin. Among the six samples, only the CJ samples from HN exhibited distinct volatile components. Specifically, the CJ samples from HN contained the unique volatile component nonanoic acid. The volatile components shared between the two origins were as follows: HN and SN shared hexyl pentanoate, and HN and HN shared (E)-2-penten-1-ol. The volatile components shared between the three origins were as follows: SD, HN and XJ shared propyl pentanoate; HB, HN and SN shared decanoic acid, guaiacol, 3-methylbutanal and pentanoic acid; and GS, HB and SD shared 2-butenal. The volatile components shared between the four origins were as follows: HN, SD, SN and XJ shared delta-octalactone, undecanal, dodecanoic acid and heptanoic acid; HN, SD, GS and XJ shared benzoic acid; GS, SD, SN and XJ shared methyl laurate; HN, SN, GS and XJ shared 2-butenoic acid-hexyl ester and ethyl pentadecanoate; HN, SN, GS and HB shared hexanal, 3-pentanone and 5-ethyl-3-hydroxy-4-methyl-2(5H)-furanone; HB, SD, SN and XJ shared β-pinene; GS, HB, SD and XJ shared vanillin; and HB, SD, SN and HN shared methyl hexanoate. The volatile components shared between the five origins were as follows: HB, HN, SN, GS and XJ shared propanoic acid; HB, SD, SN, GS and XJ shared methyl eugenol and 2-methylbutyric acid; HB, HN, SD, GS and XJ shared ethyl 3-phenylpropionate, delta-hexalactone and phenylacetic acid; GS, HB, SD, HN and SN shared 2-ethyl furan and propylene glycol; and XJ, HB, SD, HN and SN shared methyl decanoate.

As shown in Fig. 4-C, in the CJ sample from GS, the most dominant volatile component was aldehydes (64.36 %), followed by acids (23.10 %) and ketones (4.04 %). In the CJ sample from HB, the dominant volatile component was aldehydes (62.82 %), followed by acids (20.93 %) and ketones (10.84 %). In the CJ samples from SD, the most dominant volatile component was aldehydes (48.10 %), followed by acids (38.52 %) and others (6.13 %). In the CJ samples from HN, the three most abundant volatile components were acids (42.48 %), aldehydes (21.35 %) and esters (13.84 %). In the CJ samples from SN, the three categories with the highest relative contents of volatile components were acids (44.57 %), aldehydes (43.99 %) and others (5.53 %). The CJ sample from XJ was also dominated by acid volatile components (44.39 %), followed by aldehydes (39.88 %) and esters (8.76 %). Similar to the results of the GC–MS analyses, the relative content of aldehydes and

acids in the CJ samples from the six origins analysed by UF-GC-*E*-nose combined exceeded 80 % of the total composition. Samples from GS, HB and SD had relatively high aldehyde contents, while samples from HN, SX and XJ had relatively high acid contents.

3.5. Analysis of the aroma characteristics of volatile substances

The flavour information of the volatile components in the CJ samples identified by GC–MS and UF-GC-E-nose is presented in Table S4 and Table S5. Based on the results obtained by GC–MS and UF-GC-*E*-nose, acids and aldehydes accounted for more than 60 % of the total components in the CJ samples from the six origins, with relative contents ranging from 79.98 % - 95.21 % and 63.83 % - 88.56 %, respectively. Aldehydes and acids have relatively low odour thresholds and play an important role in the overall flavour composition of the CJ samples. Aldehydes produced by fatty acid oxidation and amino acid metabolism are believed to have a significant impact on the flavour profile of date palm fruits (Verma & Srivastav, 2020). Based on the GC–MS results, 5-hydroxymethyl-2-furfural was the most abundant aldehyde relative to the caramel flavour in the GS and SN samples (Shakoor, Zhang, Xie, & Yang, 2022). Hexanal was the most abundant aldehyde relative to the caramel flavour in the HB and SD samples (28.86 % and 51.65 %, respectively), which was responsible for the fresh fruity aroma of these samples (Acquaticci et al., 2024). In addition, nonanal was the predominant aldehyde in the HN and SD samples at 32.07 % and 33.75 %, respectively, with a fresh orange peel aroma (Ning, Zhou, Fan, El-Kassaby, & Jian Bian, 2024). Based on the results of the UF-GC-*E*-nose assay, octanal was found in all six CJ samples, giving them fruity aromas of orange and lemon. It was also the most abundant aldehyde in relative terms in the HB and XJ samples. In relative terms, vanillin was the most abundant aldehyde in the GS and SD samples, at 32.50 % and 17.62 %, respectively, which contributed to the pleasant vanilla odour of these samples (Boerzhijin, Osafune, Kishimoto, Hisatsune, & Isogai, 2024). In addition, hexanal was the most dominant aldehyde in the HN and SN samples, with 15.41 % and 27.88 %, respectively, exhibiting a fresh fruity flavour. Moreover, undecanal, 2-butenal and hexanal (GC–MS) were the aldehyde components detected by both GC–MS and UF-GC-*E*-nose. Acids in date fruit play a key role in determining the aroma of the samples. Based on the GC–MS results, isobutyric acid and butyric acid were found in all six CJ samples, giving the CJ samples a cheese and butter aroma. Octanoic acid was the most relatively abundant acid in the GS samples, producing fresh fruity and cheesy aromas (Lomborg, Wiebe, & Esbensen, 2008). Acetic acid was the most relatively abundant acid in the HB samples, giving the samples a vinegar-like acidic flavour. 2-Octenoic acid was the most abundant acid relative to the SD and XJ samples, at 31.23 % and 27.83 %, respectively, which contributed to the cheese-like aroma of these samples. In addition, hexanoic acid was the most dominant acid in the HN samples (34.46 %), which had a sweet and sour cheese flavour. 3-Hydroxybutyric acid was the most dominant acid in the SN samples, at 29.69 %. Based on the results of the UF-GC-*E*-nose assay, hexanoic acid and octanoic acid were found in all six CJ samples, which gave the samples fresh fruity and cheese aromas (Cao et al., 2024). Moreover, hexanoic acid was the most abundant acid in the GS, HB, SD, HN and SN samples, with relative contents of 15.06 %, 15.33 %, 15.59 %, 31.74 % and 11.16 %, respectively. Hexanoic acid was the most abundant acid in the XJ samples, with a relative content of 12.18 %. Moreover, hexanoic acid, dodecanoic acid, propanoic acid, heptanoic acid, decanoic acid, 2-methylbutyric acid, octanoic acid and benzoic acid were detected by both GC–MS and UF-GC-Nose. In addition, based on the results obtained by GC–MS and UF-GC-*E*-nose, an interesting phenomenon was that the CJ samples from GS, HB and SD were usually smaller in size and mainly had aldehydic fruity flavours. On the other hand, the CJ samples from HN, SX and XJ were usually larger in size, and the overall flavour was dominated by acidic cheese and butter notes.

The relative content and amount of alcohols in the CJ samples played an important role in determining the flavour profile. Saturated alcohols have a high threshold and contribute minimally to the flavour profile, while unsaturated alcohols have a low threshold and significantly enhance the flavour profile of the samples (Segelke, von Wuthenau, Neitzke, Müller, & Fischer, 2020). 2-Furanmethanol, an unsaturated alcohol component detected by both GC–MS and UF-GC-*E*-nose and common to all six CJ samples, imparts a caramel flavour to the samples. Based on the GC–MS results, oct-1-en-3-ol dominated the HB, SD, HN, and XJ yields to varying degrees (17.05 %, 3.63 %, 8.03 %, and 9.58 %, respectively). Oct-1-en-3-ol is a degradation product of linoleate or linolenate hydroperoxides and is considered to be a major volatile flavouring compound with an odour activity value (OAV) greater than 5. Its unsaturated nature imparts a rosy aroma to the samples (Guo et al., 2019). Based on the results of the UF-GC-*E*-nose assay, (E)-2-penten-1-ol unsaturated alcohol had the highest relative content of 7.57 % at the origin of the HN, indicating that the samples had a mushroom-like flavour. In addition, the saturated alcohol ethanol dominated the GS, HB, SD, SN and XJ production sites with varying percentages (1.50 %, 1.32 %, 0.50 %, 1.06 % and 2.57 %, respectively), which had a lesser effect on the overall flavour. Ester and ketone compounds had lower odour thresholds, with esters producing aromatic odours and ketones producing fruity flavours and enhancing the flavours of other compounds. Esters are formed through the esterification of organic acids and alcohols in the lipoxygenase metabolic pathway (Padilla-Jiménez et al., 2021). Ketones can be produced by the oxidative degradation of linoleic acid to generate hydroperoxides (Gong et al., 2023). In addition, the UF-GC-E-nose-based assay treated three other classes of compounds, β-pinene, guaiacol and methyl eugenol, which overall conferred woody, vanilla and floral aromas to the samples.

3.6. Multivariate statistical analysis

A database was created with data integrated from multidimensional features such as spectral, textural, and volatile components in order to ensure the traceability of the CJ samples. Generally, principal component analysis (PCA) is utilized for reducing the dimensionality of data. In recent years, PCA has been extensively used for differentiating agricultural products based on their variety and geographic origin (Zhou, Li, Wang, Wang, & Yu, 2024; Cao, Li, Ma, & Tian, 2024). PCA was used to analyse the multidimensional data of CJ samples from various origins, revealing the variability within and between groups. The results are displayed in Fig. 5, providing a general overview of the dataset and identifying patterns within the complex experimental data. Based on the data presented in Fig. 5-A, it is evident that the first and second principal components account for 45.2 % and 15.7 % of the variance, respectively, with a cumulative contribution of 60.9 %. The CJ samples from six different origins exhibit distinct clustering patterns, with samples from SN and XJ showing poor separation, as do samples from SD, HB, and GS. Additionally, there is partial overlap among these samples. Notably, larger CJ samples from SN, XJ, and HN demonstrated clear differentiation. Geological and climatic factors may account for the overlap between these sample groups.

As opposed to PCA, PLS-DA uses supervised methods to analyse discriminant data (Chen et al., 2024). The method can be used to separate overlapping information and group indicators that contribute significantly to principal components by imputing the importance of variables to those factors. It is possible to mitigate interference and reduce dimensionality by analyzing these indicators (Lu et al., 2021). Fig. 5-B shows that the discrimination rate for t[1] was 51.2 % while the discrimination rate for t[2] was 17.6 %, which results in a cumulative discrimination rate of 68.8 %. The clustering of 2D score scatter plots for CJ samples from diverse origins demonstrated significantly improved grouping compared to PCA, with reduced dispersion within sample clusters and enhanced separation between groups. These findings suggest successful reduction of interference and analysis dimensions. The PLS-DA model effectively traced the origin of CJ samples, showcasing strong adaptability and predictive capability. To prevent overfitting in

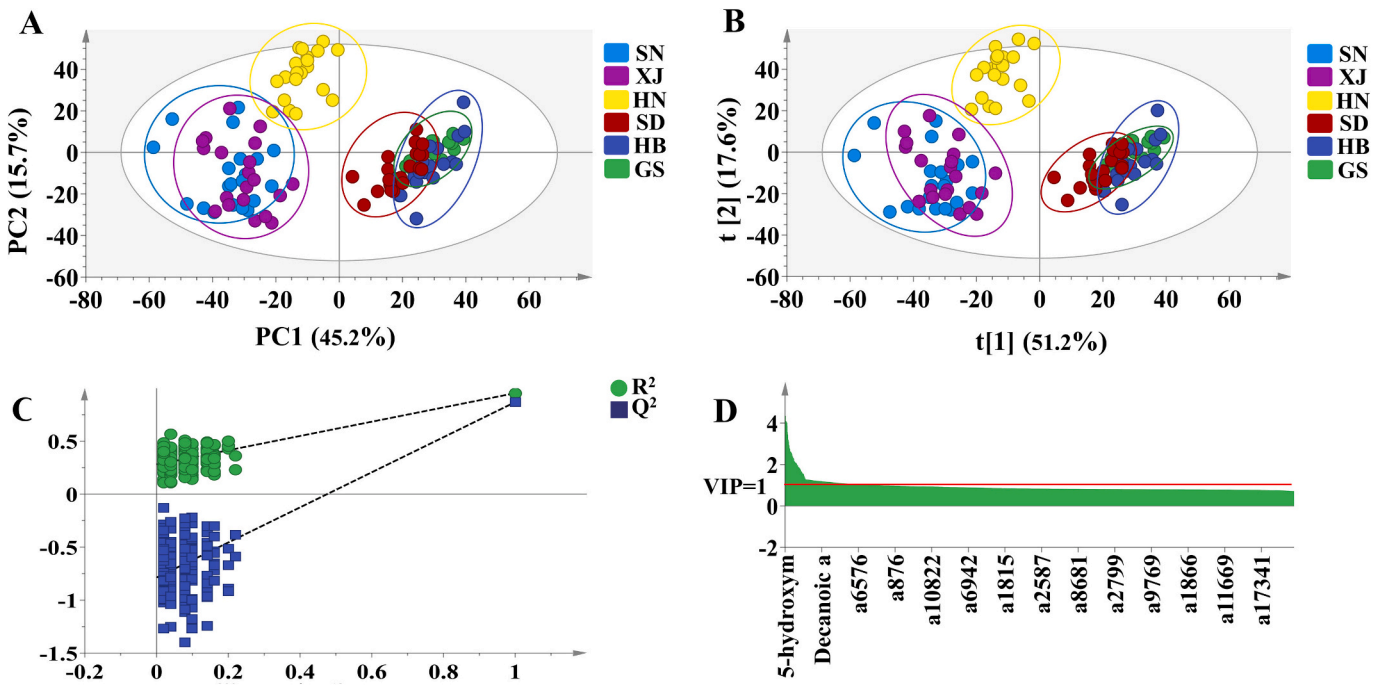


Fig. 5. Score plots of the PCA model (A); score plots of the PLS-DA model f (B); cross-validation results with 200 calculations using a permutation test (C); VIP plots (D). GS, Gansu; HB, Hebei; SD, Shandong; SN, Shaanxi; HN, Henan; XJ, Xinjiang. PCA, principal component analysis; PLS-DA, partial least squares discrimination analysis; VIP, variable importance for projecting.

the modeling process, a permutation test was employed to assess the validity of the supervised model. In Fig. 5-C, the permutation test plot for the PLS—DA model is shown for the comparison group. Permutation retention rate reductions are accompanied by decreases in the R2 and Q2 values of the stochastic model, indicating robustness and no overfitting in the model.

VIPs derived from the PLS-DA model are useful in measuring the intensity of multidimensional features (Xu et al., 2024). Model interpretation is generally considered to be highly dependent on trait factors with aVIP greater than 1. To identify the key differential traits affecting the CJ in the different provenances, we employed the VIP to identify the predominant trait factors in the CJ (Fig. 5-D). It was determined that 46 potential signature trait factors can be used to differentiate CJ from different origins when VIP $>$ 1 and P 0.05 were found. These factors include spectral bands ranging from 455.43 to 464.93 nm, 472.03–477.41 nm, 500.87–519.75 nm, 532.46–542.18 nm, and 700.11–714.13 nm and 4 textural features, including contrast, correlation, homogeneity, dissimilarity, 24 volatile components identified on the basis of UF-GC-*E*-nose, including: hexanoic acid, 2-butenal, dodecanoic acid, hexanal, vanillin, propanoic acid, nonanoic acid, 2-methylbutyric acid, pentanoic acid, octanoic acid, β-pinene, propyl pentanoate, octanal, delta-hexalactone, guaiacol, methyl hexanoate, benzoic acid, hexyl pentanoate, methyl eugenol, rheosmin, decyl butanoate, methyl decanoate, methyl laurate, ethyl pentadecanoate, and 13 volatile components identified on the basis of GC–MS, including: methyl decanoate, 5-hydroxymethyl-2-furfural, octanoic acid, dodecanoic acid, 2-butenal, decyl butanoate, methyl hexanoate, hexanal, propanoic acid, butyric Acid, 2-methylbutyric acid, hexanoic acid and benzoic acid.

3.7. Construction of the SVM classification algorithm for discriminating between CJ varieties and origins

The SVM algorithm effectively separates different classes of data points by determining the optimal hyperplane and maximizing the margins between them and is commonly used in supervised learning tasks. This study optimized the SVM classification algorithm by collecting and integrating multidimensional feature data (spectral, textural, and volatile components) using multivariate statistical analysis. As a result, CJ samples from various sources could be effectively differentiated based on the processed information. Using SVM classification, Fig. S3 shows the optimization process for differentiating CJs based on their origins.

An SVM analysis was used to identify the origin of samples from CJ based on a 7:3 sample ratio in the training and testing groups. In the training group, 84 samples were included, including 14 GS samples, 10 HB samples, 15 SD samples, 15 HN samples, 17 SN samples, and 13 XJ samples. In the test group, 36 samples were included, of which 7 GS samples, 10 HB samples, 5 SD samples, 5 HN samples, 3 SN samples, and 7 XJ samples. Based on the results of the analysis, the SVM classification algorithm model successfully distinguishes CJ samples from a variety of sources, with a 100 % success rate in both training and validation sets. Additionally, by setting the maximum number of iterations to 100, the fitness curve of the SVM classification algorithm stabilizes after 71 iterations, resulting in a discrimination rate of 100 %. The optimized SVM algorithm exhibits superior performance compared to the traditional SVM classification algorithm. Furthermore, five batches of CJ samples were obtained from each of the six origins, with data on visible spectra, LBP texture, and volatile components being gathered for each sample. Subsequently, the collected data underwent back-validation utilizing the established SVM classification model, demonstrating a discrimination rate of 100 %. The results of this analysis are shown in Fig. 6. These results confirm that the SVM model is stable and viable.

4. Conclusions

This study utilized computer vision, ultrafast gas phase electronic nose, and GC–MS technologies to collect samples of jujube from various regions in China. A multivariate statistical analysis was used to gather data on traits like spectra, texture, and odour. A SVM classification algorithm was used to integrate the data and achieve 100 % accuracy in the classification of jujube origin traceability. The purpose of this research is to lay the foundation for developing intelligent algorithms

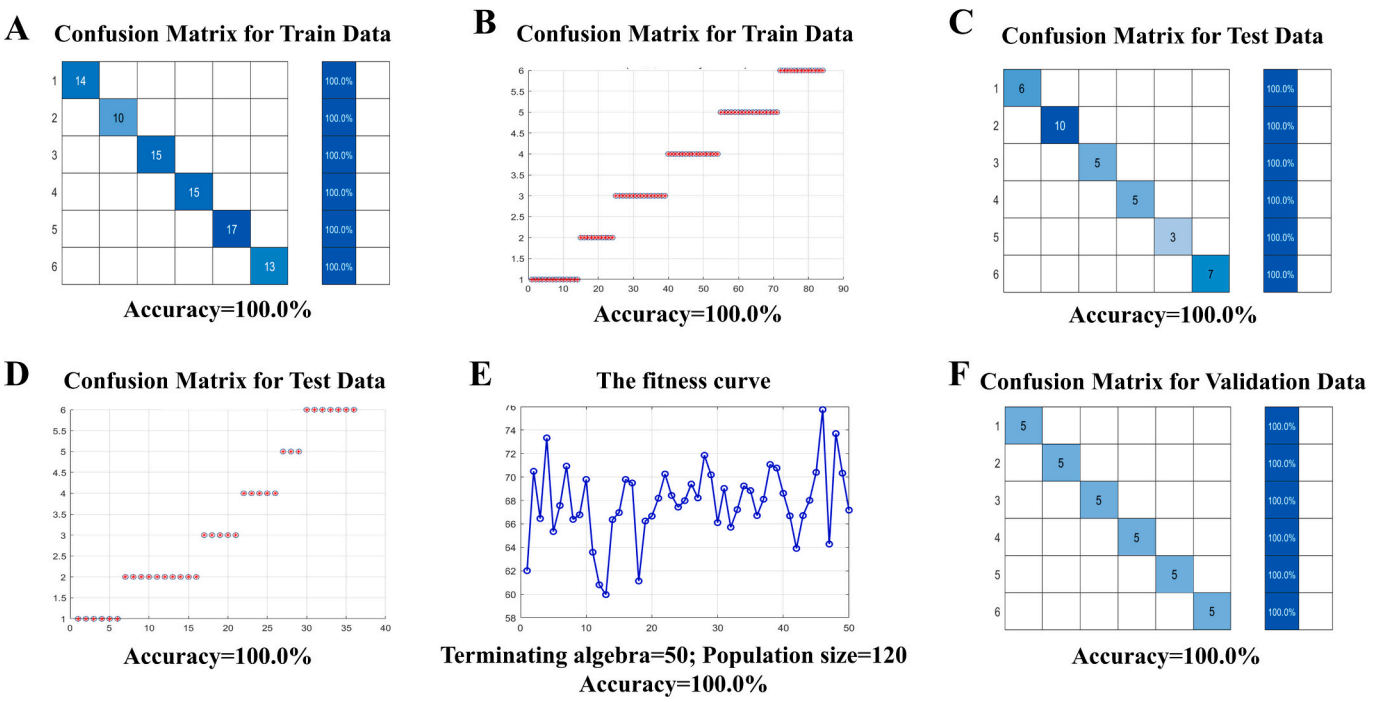


Fig. 6. Results of SVM training sets with different CJ origins (A-B); results of SVM test sets with different CJ origins (C—D); results of SVM fitness curves (E); results of SVM validation sets with different CJ origins (F). GS, Gansu; HB, Hebei; SD, Shandong; SN, Shaanxi; HN, Henan; XJ, Xinjiang.

for identifying food products and locating their origins. To combat food adulteration, future applications should focus on developing lightweight instruments and exploring multidimensional anticounterfeiting detection techniques. Innovative noninvasive detection methods can increase food quality and safety, increase industry support, and protect consumer interests.

CRediT authorship contribution statement

Peng Chen: Writing – original draft, Software, Methodology, Conceptualization. **Xiaoli Wang:** Validation, Methodology, Investigation, Formal analysis. **Rao Fu:** Validation, Methodology, Investigation, Formal analysis. **Xiaoyan Xiao:** Data curation. **Yu Li:** Data curation. **Tulin Lu:** Software, Methodology, Funding acquisition. **Tao Wang:** Data curation. **Qiaosheng Guo:** Writing – review & editing, Data curation. **Peina Zhou:** Writing – review & editing, Software, Methodology, Funding acquisition. **Chenghao Fei:** Writing – review & editing, Methodology, Funding acquisition, Formal analysis.

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.

Acknowledgements

This research study was supported by the Start-up Research Fund of Nanjing Agricultural University (130-804141), the National Key Research and Development Program Projects (No. 2023YFC3504200) and the National Natural Science Foundation of China (Grant No. 82404863, 82404814).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.fochx.2025.102190.

Data availability

Data will be made available on request.

References

Acquaticci, L., Schouten, M. A., Angeloni, S., Caprioli, G., Vittori, S., & Romani, S. (2024). Influence of baking conditions and formulation on furanic derivatives, 3-methylbutanal and hexanal and other quality characteristics of lab-made and commercial biscuits. *Food Chemistry, 437*(Pt 1), Article 137791. https://doi.org/10.1016/j.foodchem.2023.137791

Boerzhijin, S., Osafune, Y., Kishimoto, T., Hisatsune, Y., & Isogai, A. (2024). Quantitative determination of vanillin and its detection threshold in sake. *Food Chemistry, 458*, Article 140224. https://doi.org/10.1016/j.foodchem.2024.140224

Cao, Y., Huang, C., Guo, Y., Xu, Y., Gong, S., Chu, Q., & Chen, P. (2024). Unraveling the contributing factors of stale odor in Longjing tea through a sensomics approach. *Food Chemistry, 441*, Article 138301. https://doi.org/10.1016/j.foodchem.2023.138301

Cao, Z. X., Li, Y. X., Ma, A. J., & Tian, Y. L. (2024). Analysis and comparison of staminate flowers components in five Chinese walnut varieties. *Food & Medicine Homology, 1* (1), 9420005. https://doi.org/10.26599/FMH.2024.9420005

Chen, M., Wang, Z., Yu, J., Wang, J., Xu, H., & Yue, X. (2023). Effects of electron beam irradiation and ultrahigh-pressure treatments on the physicochemical properties, active components, and flavor volatiles of jujube jam. *LWT, 187*, Article 115292. https://doi.org/10.1016/j.lwt.2023.115292

Chen, P., Fei, C., Fu, R., Xiao, X., Qin, Y., Li, X., Guo, Z., Huang, J., Ji, D., Li, L., Lu, T., Guo, Q., & Su, L. (2024). Polygonati Rhizoma varieties and origins traceability based on multivariate data fusion combined with an artificial intelligence classification algorithm. *Food Chemistry, 460*(Pt 1), Article 140350. https://doi.org/10.1016/j.foodchem.2024.140350

Chen, P., Fu, R., Shi, Y., Liu, C., Yang, C., Su, Y., Lu, T., Zhou, P., He, W., Guo, Q., & Fei, C. (2024). Optimizing BP neural network algorithm for Pericarpium Citri Reticulatae (Chenpi) origin traceability based on computer vision and ultra-fast gas-phase electronic nose data fusion. *Food Chemistry, 442*, Article 138408. https://doi.org/10.1016/j.foodchem.2024.138408

Chen, Q., Song, J., Bi, J., Meng, X., & Wu, X. (2018). Characterization of volatile profile from ten different varieties of Chinese jujubes by HS-SPME/GC-MS coupled with E-nose. *Food Research International, 105*, 605–615. https://doi.org/10.1016/j.foodres.2017.11.054

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

Chi, X., Zhang, Y., Zheng, N., Wang, J., & Liu, H. (2023). HS-GC-IMS and HS-SPME/GC-MS coupled with E-nose and E-tongue reveal the flavors of raw milk from different regions of China. *Current Research in Food Science, 8*, Article 100673. https://doi.org/10.1016/j.crfs.2023.100673

Duan, Y., Liu, S., Zhu, Y., Wang, Y., Yan, F., Liu, Z., Shi, X., Liu, P., & Liu, M. (2023). *The influences of soil and meteorological factors on the growth and fruit quality of Chinese jujube (Ziziphus jujuba mill.). Plants (Basel, Switzerland), Vol. 12(24), 4107*. https://doi.org/10.3390/plants12244107

Erba, I., Buzzelli, M., & Schettini, R. (2024). RGB color constancy using multispectral pixel information. *Journal of the Optical Society of America. A, Optics, Image Science, and Vision, 41*(2), 185–194. https://doi.org/10.1364/JOSAA.506186

Gong, C., He, N., Zhu, H., Anees, M., Lu, X., & Liu, W. (2023). Multi-omics integration to explore the molecular insight into the volatile organic compounds in watermelon. *Food Research International, 166*, Article 112603. https://doi.org/10.1016/j.foodres.2023.112603

Guo, Q., Yu, J., Zhao, Y., Liu, T., Su, M., Jia, Z., Zhao, Y., Mu, Z., & Yang, M. (2019). Identification of fishy odor causing compounds produced by Ochromonas sp. and Cryptomonas ovate with gas chromatography-olfactometry and comprehensive two-dimensional gas chromatography. *The Science of the Total Environment, 671*, 149–156. https://doi.org/10.1016/j.scitotenv.2019.03.370

Islam, S., & Cullen, J. M. (2021). Food traceability: A generic theoretical framework. *Food Control, 123*(3), Article 107848. https://doi.org/10.1016/j.foodcont.2020.107848

Li, C., Al-Dalali, S., Wang, Z., Xu, B., & Zhou, H. (2022). Investigation of volatile flavor compounds and characterization of aroma-active compounds of water-boiled salted duck using GC-MS-O, GC-IMS, and E-nose. *Food Chemistry, 386*, Article 132728. https://doi.org/10.1016/j.foodchem.2022.132728

Lomborg, C. J., Wiebe, L., & Esbensen, K. H. (2008). At-line determination of octanoic acid in cultivation broth–an electronic tongue (ET) feasibility study. *Journal of Biotechnology, 133*(1), 162–169. https://doi.org/10.1016/j.jbiotec.2007.09.013

Ma, M., Chen, Z., Huang, B., Chen, X., Liu, H., Peng, Z., Dong, P., Lu, J., & Wu, D. (2024). Characterizing the key aroma compounds of barley malt from different origins using GC-E-nose, HS-SPME-GC-MS, and HS-GC-IMS. *Food Bioscience, 58*, Article 103707. https://doi.org/10.1016/j.fbio.2024.103707

Ning, K., Zhou, T., Fan, Y., El-Kassaby, Y., & Jian Bian, J. (2024). Discrimination of tulip cultivars with different floral scents using sensory assessment, electronic nose, and gas chromatography–mass spectrometry. *Industrial Crops and Products, 219*, Article 118996. https://doi.org/10.1016/j.indcrop.2024.118996

Padilla-Jiménez, S. M., Angoa-Pérez, M. V., Mena-Violante, H. G., Oyoque-Salcedo, G., Montañez-Soto, J. L., & Oregel-Zamudio, E. (2021). Identification of organic volatile markers associated with aroma during maturation of strawberry fruits. *Molecules (Basel, Switzerland), 26*(2), 504. https://doi.org/10.3390/molecules26020504

Rong, Y., Xie, J., Yuan, H., Wang, L., Liu, F., Deng, Y., Jiang, Y., & Yang, Y. (2023). Characterization of volatile metabolites in Pu-erh teas with different storage years by combining GC-E-nose, GC-MS, and GC-IMS. *Food Chemistry: X, 18*, Article 100693. https://doi.org/10.1016/j.fochx.2023.100693

Segelke, T., von Wuthenau, K., Neitzke, G., Müller, M. S., & Fischer, M. (2020). Food authentication: Species and origin determination of truffles (tuber spp.) by inductively coupled plasma mass spectrometry and Chemometrics. *Journal of Agricultural and Food Chemistry, 68*(49), 14374–14385. https://doi.org/10.1021/acs.jafc.0c02334

Shakoor, A., Zhang, C., Xie, J., & Yang, X. (2022). Maillard reaction chemistry in formation of critical intermediates and flavour compounds and their antioxidant properties. *Food Chemistry, 393*, Article 133416. https://doi.org/10.1016/j.foodchem.2022.133416

Shen, H., Wang, J., & Lyu, B. (2023). A study on factors influencing Chinese farmers' willingness to input in jujube planting. *Heliyon, 9*(11), Article e21470. https://doi.org/10.1016/j.heliyon.2023.e21470

Shi, Q., Han, G., Liu, Y., Jiang, J., Jia, Y., & Li, X. (2022). Nutrient composition and quality traits of dried jujube fruits in seven producing areas based on metabolomics analysis. *Food Chemistry, 385*, Article 132627. https://doi.org/10.1016/j.foodchem.2022.132627

Shin, D. M., Kim, Y. J., Choi, Y. S., Kim, B. K., & Han, S. G. (2023). Duck fat: Physicochemical characteristics, health effects, and food utilizations. *LWT, 188*, Article 115435. https://doi.org/10.1016/j.lwt.2023.115435

Song, L., Zheng, J., Zhang, L., Yan, S., Huang, W., He, J., & Liu, P. (2019). *Phytochemical profiling and fingerprint analysis of Chinese jujube (Ziziphus jujuba mill.) leaves of 66 cultivars from Xinjiang Province. Molecules (Basel, Switzerland), Vol. 24(24), 4528*. https://doi.org/10.3390/molecules24244528

Sun, Y., Sun, S., Zhang, J., & Liu, S. (2016). Reconstruction of wind velocity distribution using POD model. *Energy Procedia, 100*, 137–140. https://doi.org/10.1016/j.egypro.2016.10.155

Verma, D. K., & Srivastav, P. P. (2020). A paradigm of volatile aroma compounds in rice and their product with extraction and identification methods: A comprehensive review. *Food research international (Ottawa, Ont.), 130*, Article 108924. https://doi.org/10.1016/j.foodres.2019.108924

Wu, W., Zhan, J., Tang, X., Li, T., & Duan, S. (2022). Characterization and identification of pork flavor compounds and their precursors in Chinese indigenous pig breeds by volatile profiling and multivariate analysis. *Food Chemistry, 385*, Article 132543. https://doi.org/10.1016/j.foodchem.2022.132543

Xu, S. Y., Bai, C. H., Chen, Y. L., Yu, L. L., Wu, W. J., & Hu, K. F. (2024). Comparing univariate filtration preceding and succeeding PLS-DA analysis on the differential variables/metabolites identified from untargeted LC-MS metabolomics data. *Analytica Chimica Acta, 1287*, Article 342103. https://doi.org/10.1016/j.aca.2023.342103

Yan, Y., Wang, Y., & Lei, Y. (2022). Micro learning support vector machine for pattern classification: A high-speed algorithm. *Computational Intelligence and Neuroscience, 2022*, 4707637. https://doi.org/10.1155/2022/4707637

Yang, G., He, Y., Li, X., Liu, H., & Lan, T. (2022). Gabor-GLCM-based texture feature extraction using flame image to predict the O2 content and NO x. *ACS Omega, 7*(5), 3889–3899. https://doi.org/10.1021/acsomega.1c03397

Yang, L., Wang, X., He, S., Luo, Y., Chen, S., Shan, Y., Wang, R., & Ding, S. (2021). Heat shock treatment maintains the quality attributes of postharvest jujube fruits and delays their senescence process during cold storage. *Journal of Food Biochemistry, 45*(10), Article e13937. https://doi.org/10.1111/jfbc.13937

Zhao, X., Zhang, B., Luo, Z., Yuan, Y., Zhao, Z., & Liu, M. (2023). Composition analysis and nutritional value evaluation of amino acids in the fruit of 161 jujube cultivars. *Plants, 12*(9), 1744. https://doi.org/10.3390/plants12091744

Zhou, Z. R., Li, J. L., Wang, Y. X., Wang, Z. Q., & Yu, Y. T. (2024). Raman identification of adulteration in poly-alpha-olefin synthetic lubricant using principal component analysis and two-dimensional correlation spectroscopy. *Journal of Molecular Structure, 1295*(2), Article 136677. https://doi.org/10.1016/j.molstruc.2023.136677