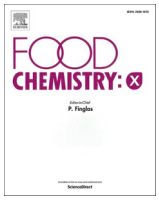
Food Chemistry: X 22 (2024) 101447

Contents lists available at ScienceDirect

Food Chemistry: X

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



The utilization of a data fusion approach to investigate fingerprint profiles of dark tea from China's different altitudes

Zhenhong Wang [a,1], Yuanxi Han [b,1], Liyou Zhang [a], Yongxiang Ye [b], Liping Wei [a], Liang Li [b,*]

[a] Resources & Environment College, Tibet Agriculture & Animal Husbandry University; Tea Industry Engineering Center of Tibet Agriculture and Animal Husbandry University, Nyingchi 860000, China
[b] Food Science College, Tibet Agriculture & Animal Husbandry University; R&D Center of Agricultural Products with Tibetan Plateau Characteristics; The Provincial and Ministerial Co-founded Collaborative Innovation Center for R&D in Tibet Characteristic Agricultural and Animal Husbandry Resources, Nyingchi 860000, China

ARTICLE INFO

Keywords:
Data fusion
High-performance liquid chromatography
Dark tea
Classification

ABSTRACT

Dark tea refers to a kind of post-fermented product, and its quality and price vary owing to the distinct altitudes at which it grows. In this study, a novel method based on high performance liquid chromatography with a diode-array detector (HPLC-DAD) and an evaporative light scattering detector (HPLC-ELSD) was proposed for the classification of dark teas from distinct altitudes in China. Through implementing a strategy fusing feature-level data to construct a combined dataset, the classification performance of dark teas from distinct altitudes in China was evaluated after preprocessing. The results suggested that, through the feature fusion strategy, the identification accuracy rate increased from <70% of a single detector to 76.923%. After the implementation of preprocessing, the identification accuracy rate was further improved. Typically, the model identification accuracy rate after short-time Fourier Transform (STFT) treatment reached 92.85%, and the AUROC value was higher than 0.84, exhibiting a favorable generalization ability. This study provides a new thinking for the identification technology of dark teas from different altitudes in China.

1. Introduction

Dark tea, one of the 6 substantial tea groups in China, is renowned for its distinctive post-fermentation process (Pan, Le, He, Yang, & Ling, 2023). Unlike other tea kinds, the solid-state fermentation process plays a pivotal role in the production of dark tea since microorganisms and a warm and humid environment facilitate complex biochemical reactions in the harvested tea leaves, causing the development of unique flavors and beneficiary compounds (Zheng et al., 2019). These features grant dark tea varied health-beneficiary features, such as antioxidative characteristics (Ma et al., 2022), a natural enemy of obesity (Qu et al., 2023), and a decrease in high blood lipids (Mao et al., 2018).

As the health-promoting functions of dark tea are researched further, the market demand increases, especially for dark tea cultivated in high-altitude plantations, which are favored for their abundant beneficial volatile compounds. The reports underline that the aroma profile and positive health effects of tea leaves are pertinent to the altitude at which they grow since high-altitude tea leaves contain more aromatic hydrocarbons, ketones, esters, and health-promoting volatile compounds (Jiang, Boorboori, Xu, & Lin, 2021; Kfoury et al., 2018; Wang et al., 2022; Wang et al., 2022; Wang, Liang, Ko, & Lin, 2022). Additionally, high-altitude tea cultivation in low-production environments normally follows natural farming practices, decreasing the administration of chemical fertilizers and pesticides, which is another reason for its acclaimed status (Wang, Li, et al., 2022; Wang, Liang, et al., 2022; Wang, Nie, et al., 2022). Consequently, high-altitude dark tea can be marketed with a higher price tag owing to favorable characteristics. However, it is difficult for customers to distinguish the quality status of high-altitude dark tea from other types of dark teas. Some retailers might have falsely labeled lower-quality products as high-altitude (Jiang et al., 2022). Such actions undermine the trust of the customers, leading to losses for both customers and producers. Thus, the development of efficient approaches to authenticate the origin of dark tea has been a substantial issue.

Currently, methods analyzing the trace of tea origins include component investigation methods such as gas chromatography–mass spectrometry (Yun et al., 2021), liquid chromatography-mass spectrometry, HPLC (Gu et al., 2022), element analysis methodologies such

* Corresponding author.
E-mail address: jwllok@sina.com (L. Li).
[1] Zhenhong Wang and Yuanxi Han are Co-first Authors

https://doi.org/10.1016/j.fochx.2024.101447
Received 2 January 2024; Received in revised form 21 April 2024; Accepted 5 May 2024
Available online 10 May 2024

as table isotope mass spectrometry (Jin et al., 2020), inductively coupled plasma mass spectrometry (Liu, Meng, Zhao, Ye, & Tong, 2021), spectral methodologies such as near-infrared spectroscopy (Zhang et al., 2023), nuclear magnetic resonance (Cui et al., 2023), and sensory analysis techniques called electronic nose (Fu, Liu, Chen, & Xing, 2023), electronic tongue (Li, Lei, Yang, & Liu, 2014)). HPLC was found to present characteristics of good reproducibility, high precision, and minimal reagent utilization, thus, making it broadly implemented in tea authentication processes (Wang, Li, et al., 2022; Wang, Liang, et al., 2022; Wang, Nie, et al., 2022). Nevertheless, to our knowledge, limited research focusing on distinguishing dark teas from distinct altitudes was conducted in the literature. Previous research showed that high-performance liquid chromatography combined with chemometrics was often employed by scholars, implementing limited non-volatile substances in tea such as catechins (Fang et al., 2019), amino acids (Wu et al., 2021), and caffeine (Su, Wu, Wan, & Ning, 2019) as metrics to identify the origin. The process of quantifying and qualifying chemical components to some extent was determined to be complex and costly (Sun et al., 2024). Note that the chemical composition of fresh tea leaves became intricate after the solid-state fermentation process (Zhu et al., 2020). Thus, employing limited chemical markers for distinguishing dark teas from varied altitudes in the market seemed challenging.

The detection of a single limited substance is insufficient to reflect all chemical components in complex samples, thus limiting the comprehensive exploration and examination of the chemical attributions of samples from distinct sources (Pinu, 2018). In recent years, scholars have increasingly implemented non-targeted HPLC algorithms to authenticate the food quality of coffee (Klikarova & Ceslova, 2022), chili pepper (Sun et al., 2023), honey (García-Seval, Martínez-Alfaro, Saurina, Núñez, & Sentellas, 2022), and other food products, presenting the potential of the algorithm to collect as much sample as possible. Additionally, as data analysis algorithms improve, it is possible to efficiently integrate non-targeted data from several sensors through the utilization of rational data fusion algorithms to attain more comprehensive and rich data (Deng, Chen, Fu, & Yun, 2023). By unifying these 2 approaches, the advantages of non-targeted HPLC approaches could be better leveraged to give comprehensive data to model authentication, potentially appearing as a new tool to identify the authenticity and safety of food products.

It is a fact that compounds in dark tea could exhibit distinct UV absorption characteristics. The research aimed to employ the combination of the HPLC, DAD, and ELSD to establish HPLC-DAD and HPLC-ELSD fingerprint profiles of dark teas from distinct altitudes. Thus, more detailed chemical information can be captured from samples. Through a strategy fusing feature-level data, the data from the 2 fingerprint profiles are integrated to attain a more refined and comprehensive dataset of dark teas from varied altitudes, bringing new perspectives and possibilities to authenticate dark tea from distinct origins.

2. Experimentation

2.1. Chemicals reagents and samples

Methanol of HPLC grade is sourced from Merck Chemical Technology in Shanghai, while acetonitrile and acetic acid are procured from Tianjin Comeo Chemical Reagent Co. in Tianjin, China. Ultra-pure water is obtained from Guangzhou Watson's Food & Beverage Co., Ltd., located in Guangzhou, China.

Dark tea samples (n = 48) are gathered from commercial tea products based on the distinct production plantations in China (Fig. 1). The dark tea from distinct regions is categorized into production areas low-altitude (n = 16), medium-altitude (n = 16), and at high-altitude (n = 16), respectively with the average altitudes ranging between 200 and

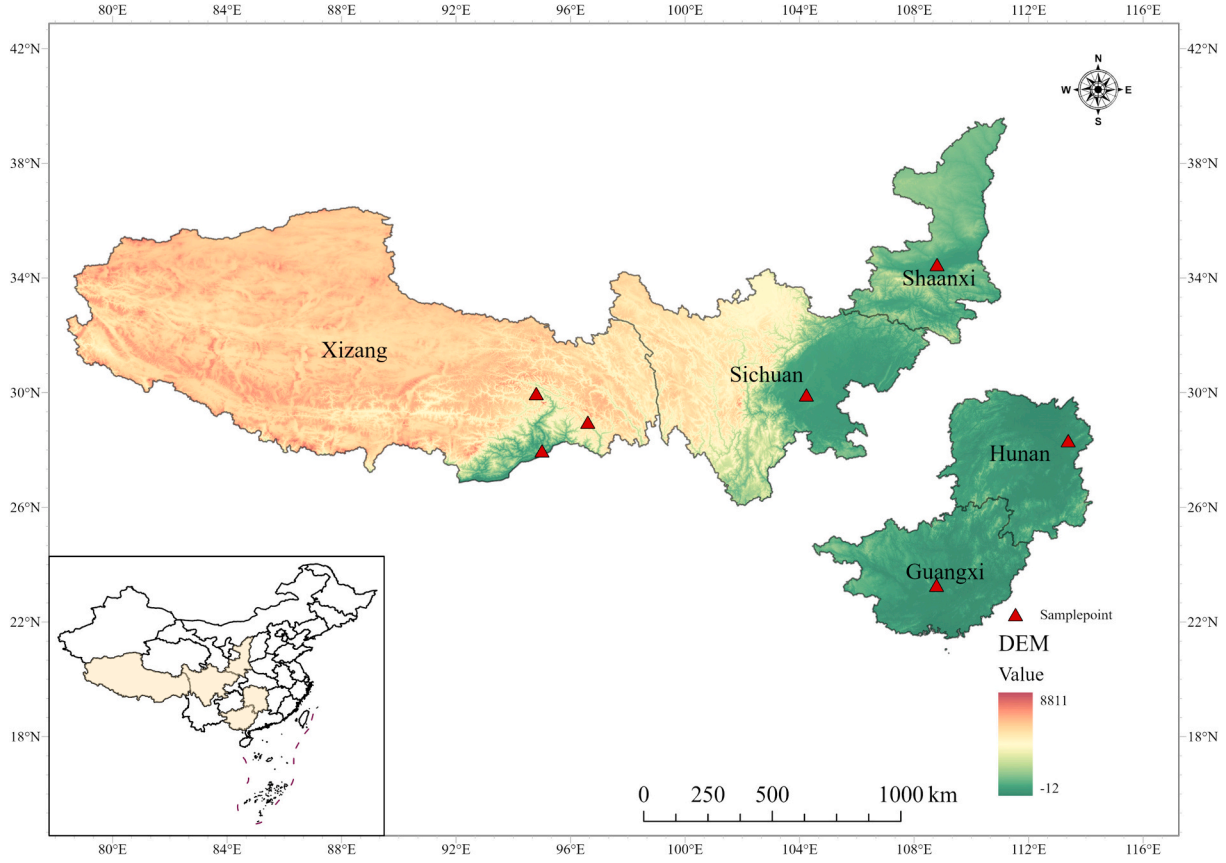


Fig. 1. Origins of the dark tea.

1000 m, 1100 to 1500 m, and 2500 to 3000 m, respectively, based on the mean altitude at which tea is grown. Then, each sample is preserved in an aluminum foil package and kept at −20 °C in a refrigerator. To analyze an accurate quantity of powdered sample (passed through a 40-mesh sieve) is weighed. The sample weighing 0.20 g is transferred into a volumetric flask (50 mL capacity) and diluted with a solvent consisting of 50% methanol to be used in the extraction process. The mixture is subjected to ultrasound-assisted extraction at a constant temperature of 60 °C for 60 min by utilizing a water bath. The extraction solution is allowed to reach ambient temperature before 10 min of centrifuge at 13000 rpm to detach the supernatant. This is then passed through a 0.22 μm membrane filter. The resulting filtrate is reaped for subsequent chromatography testing.

2.2. Chromatographic analysis

The analysis is conducted by employing an HPLC with 2 distinct detectors. The first one used in the analysis is the Agilent 1260 Infinity HPLC system (Agilent Technologies, USA). It is composed of a G1311C quaternary gradient pump, a G1329B autosampler, a G1316A column oven, and a G4212B diode array detector (DAD). The employed chromatographic column is the Agilent ZORBAX Plus-C18 (250 mm × 4.6 mm, 5 μm), operated at 35 °C with a flow ratio of 0.3 mL/min. Detection is performed at a wavelength of 280 nm, utilizing an injection volume of 20 μL. The mobile phase is composed of 5% acetic acid water (A) and acetonitrile (B). Before the sample is used, both mobile phases are filtered through a 0.2 μm membrane and degassed by employing an ultrasound. The sample is subjected to the following gradient elution: 85% A (0.00 min to 8.00 min), 85% A to 75% A (8.00 min to 25.00 min), 75% A to 60% A (25.00 min to 35.00 min), 60% A to 40% A (35.00 min to 45.00 min), 40% A to 20% A (45.00 min to 55.00 min), 90% A (55.00 min to 60.00 min). After each run, an equilibration period of 15 min is performed under initial requirements (85% A and 15% B). The approach is called HPLC-DAD.

The second detector is the Nexera LC20A HPLC system (Shimadzu Corporation, Japan). It consists of a CBM-20Alite system controller, a SIL-20A autosampler, a CTO-20A column oven, and an ELSD-LT II evaporative light scattering detector. The utilized chromatographic column is the Shim-pack GIS C18 (250 mm × 4.6 mm, 5 μm), operated in non-split mode with a drift tube temperature of 50 °C and an injection volume of 20 μL. The mobile phase consists of 5% acetic acid water (A) and acetonitrile (B). Like the previous system, both mobile phases are filtered through a 0.2 μm membrane and degassed utilizing ultrasound before the sample is used. The sample is subjected to the following gradient elution: 99% A (0.00 min to 5.00 min), 99% A to 90% A (5.00 min to 23.00 min), 90% A to 82% A (23.00 min to 50.00 min), 82% A to 10% A (50.00 min to 60.00 min), 10% A (60.00 min to 75.00 min). After each run, an equilibration period of 15 min is performed under initial conditions (85% A and 15% B). The approach is called HPLC-ELSD.

2.3. The analysis of the data

2.3.1. The demonstration of the experimentation

Fig. 2 depicts the primary steps of data analysis in the research. First, the fingerprint profiles of dark teas from distinct altitudes are constructed by implementing the raw chromatographic signals collected by the chemical investigation approaches from varied detectors, and representative characteristic signals are picked. The chosen signals from distinct detectors are then integrated through attribute fusion to construct a new dataset. To advance the data quality of this newly constructed dataset, 3 distinct preprocessing approaches are implemented before the classification approach is run. Lastly, the method conducting both classification and discrimination that yields the best traceability outcomes is assessed.

2.3.2. Feature selection and data fusion

Data fusion schemes enable the procurement of complementary information from a range of analytical instruments (Borras et al., 2015), thus facilitating a more comprehensive and precise chemical depiction of the sample (Deng et al., 2023). Based on the fusion hierarchy, it is normally grouped into 3 types: fusion at the data level, the feature level, and the decision level, respectively.

2.3.3. Data preprocessing

In this scenario, 3 preprocessing methods are implemented to enhance the caliber of the integrated data collection. The STFT is an ideal analytical approach and provides high resolution for signal frequencies in digital signal processing, enabling a clear extraction of frequency components (spectra) within the spectral signal. The utilization of the STFT efficiently decreases random noise and interference within the chromatic data, thus improving the precision of classifications in subsequent approaches.

Conversely, the Hilbert Transform (HT) as a methodology investigates time-frequency domains, depending on signal-specific local traits. It proves especially effective to examine signals that are characterized as non-stationary and nonlinear. This approach decomposes complex non-stationary signals into the components of the multiple intrinsic mode function, thus, allowing for the description of local characteristics of signals on the time-frequency plane. When compared to other analysis approaches like the STFT, it exhibits better energy concentration and could reflect the local attributions of non-stationary signals more precisely, thus enabling more accurate attribute derivation of such signals.

The Infinite Impulse Response (IIR) transform is an extremely effective filter in terms of computational efficacy. It functions by changing the relative proportions of frequency components or by filtering out specific frequency components in the input signal, designating advantageous higher precision and stability.

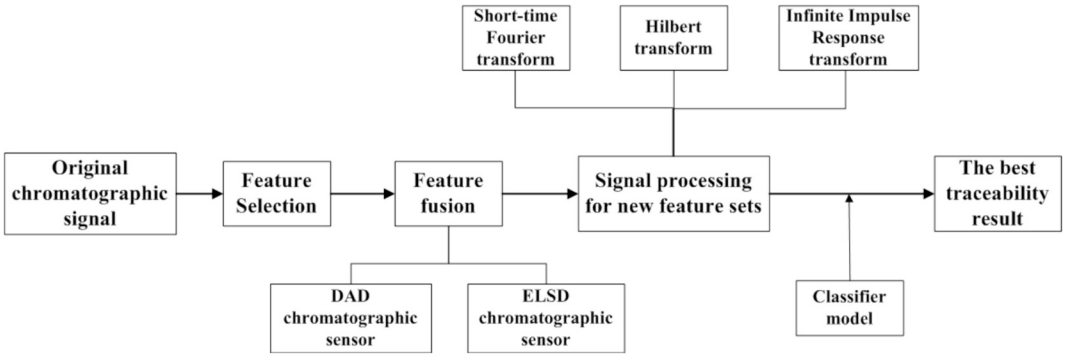


Fig. 2. The basic flow chart of experimental processing.

2.3.4. Chemometrics

Partial Least Squares Discriminant Analysis (PLS-DA), a linear classification approach based on partial least squares regression, is broadly implemented to validate the authenticity of food products (Jiménez-Carvelo, Martín-Torres, Ortega-Gavilán, & Camacho, 2021). An association between the predictor variables, X, and the outcome matrix, Y, is represented by a partial least squares regression. The matrix Y, the dependent variables, accepts binary coding where '1' and '0' denote sample inclusion and non-inclusion in a particular class, respectively. The objective is to pinpoint dark tea at low, medium, and high altitudes, respectively. The matrix's rows of dependent variables, Y, are represented by (1,0,0), (0,1,0), and (0,0,1).

2.3.5. Software

Chromatographic data of dark tea samples are captured by 2 detectors and converted into CSV format. Liquid chromatography fingerprint profiles of dark teas from varied altitudes are produced by employing Original 2021 (Origin Lab Corporation, Roundhouse Plaza., Northampton, MA, USA), while MATLAB R2018b (The MathWorks, Inc., Natick, MA, USA) software is utilized to extract features, preprocess of chromatographic data, and construct fusion models.

3. Results and discussion

3.1. Classification of single detector data based on PLS-DA

PLS-DA, a prediction and discrimination approach, leverages high-dimensional data for classification predictions, strategically maximizing intergroup variances based on predefined classes. Núñez, Martínez, Saurina, and Núñez (2021) utilized high-performance liquid chromatography with fluorescence detection fingerprint profile spectrum in conjunction with PLS-DA to run traceability research on coffee. The findings reveal that the constructed PLS-DA using statistically significant attributes exhibited high discriminative capability. This experimentation acted as a crucial reference to identify dark tea from varied altitudes. Consequently, it was suggested that the PLS-DA proved to perform the traceability on the fingerprint profile spectrum of dark tea. The fingerprint profiles spectra of randomly selected dark tea samples from different altitudes were shown in Fig. 3A and C, respectively. Dark teas from distinct altitudes exhibited varied fingerprint spectrum characteristics when both DAD and ELSD detectors were implemented. These outcomes formed 48*7201 and 48*2603 dimensional matrices, respectively, which were implemented to construct the PLS-DA. Table 1 depicts the overall predictive precision of the PLS-DA for the dark tea fingerprint profile spectrum that is attained from DAD and ELSD

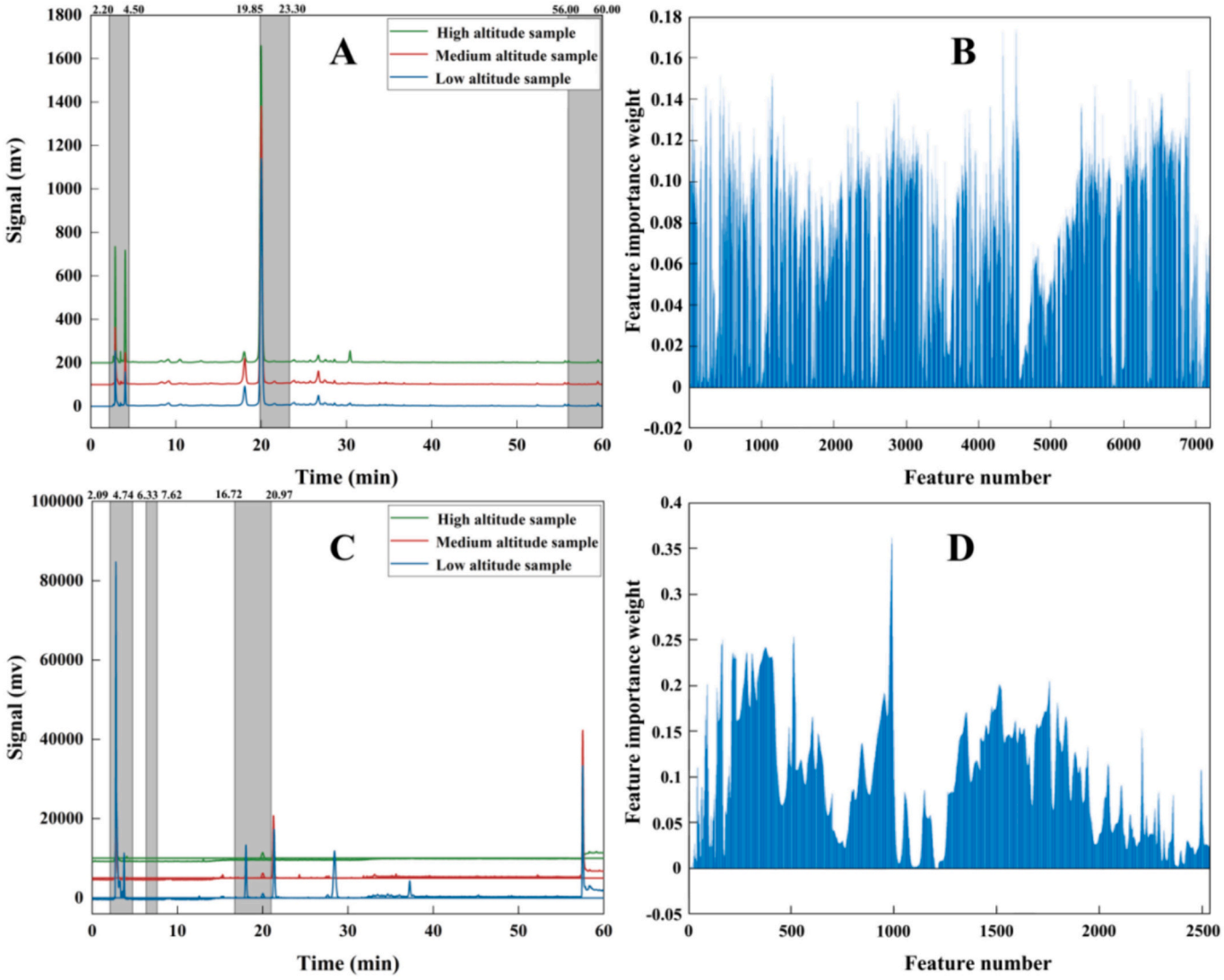


Fig. 3. The feature selection results of the raw chromatographic data from the HPLC-DAD and HPLC-ELSD.

Table 1
The results of the predictive classification model using the HPLC-DAD and HPLC-ELSD raw datasets.

DAD Chromatographic sensor (Accuracy = 64.583%)				ELSD Chromatographic sensor (Accuracy = 68.750%)			
Predicted classes	True classes			Predicted classes	True classes		
	Class 1	Class 2	Class 3		Class 1	Class 2	Class 3
Class 1	**6**	1	2	Class 1	**6**	2	0
Class 2	6	**13**	2	Class 2	5	**11**	0
Class 3	4	2	**12**	Class 3	5	3	**16**
Total	16	16	16	Total	16	16	16

Class 1: Samples at low altitude, Class 2: Samples at medium altitude, Class 3: Samples at high altitude.

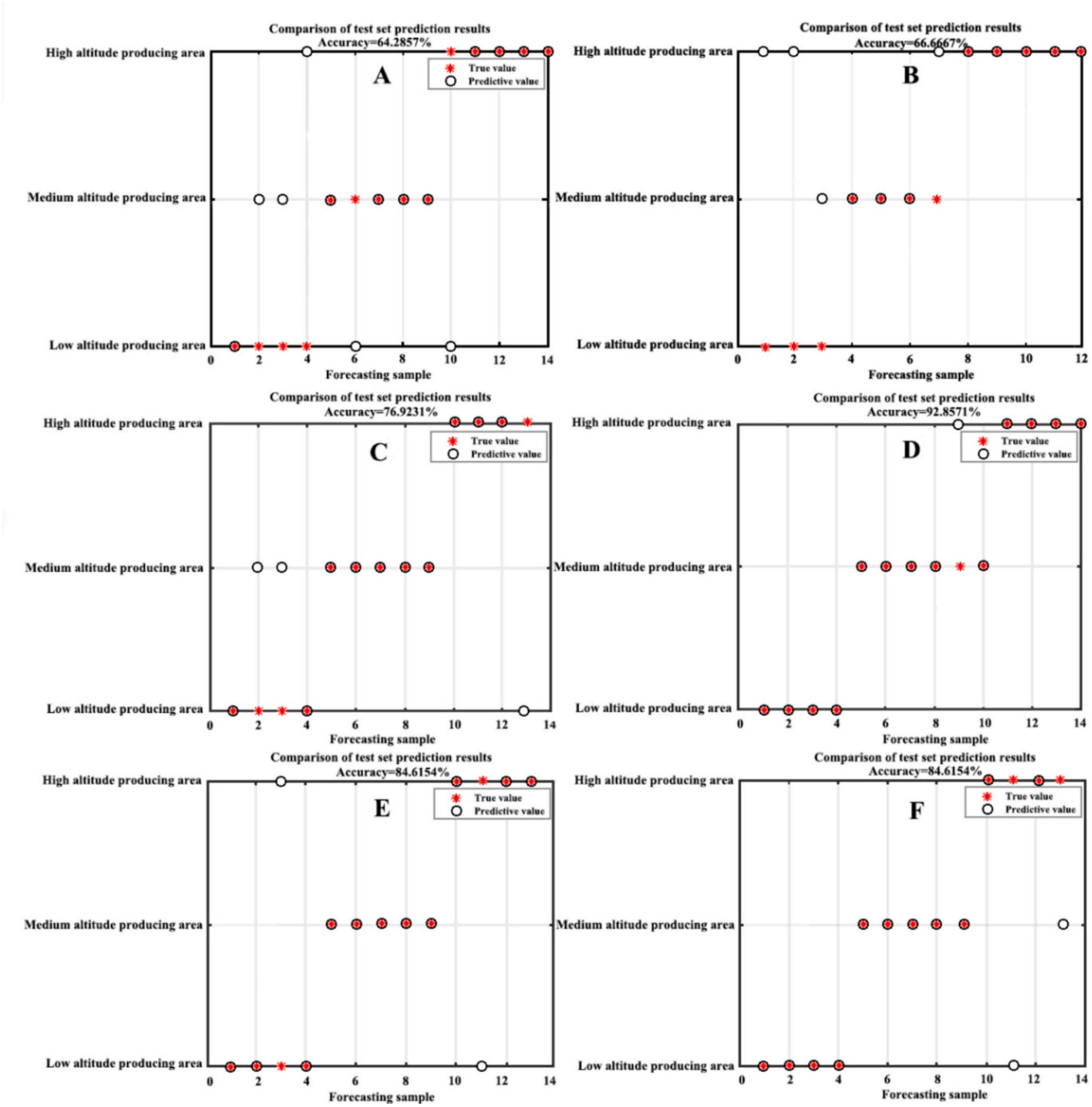


Fig. 4. The comparison of the results of the classification model with two single sensors, data fusion, and preprocessed data fusion.

detectors, which amounted to variations of 64.583% and 68.750%, respectively. The investigation also reveals that the low predictive precision is attributed to the excessive quantity of redundant attributes in the original data. As a result, irrelevant attributions diminish the quality ratio of "clustering and patterns" depicted in the data, thus leading to a subpar separation impact of the model (Dash, Ong, & Ieee., 2011). Thus, the research necessitates the exploration of a broader spectrum of algorithms to investigate the attributes from a singular detector, to further pinpoint the optimal traceability impact.

3.2. Improvement and optimization of the approach tracing dark tea origin

3.2.1. Exploration of attribute fusion

Attribute fusion can disregard redundant information represented by correlations among distinct attribute sets to the greatest extent, ultimately assisting in the progression of classification precision for subsequent classifier algorithms. Obisesan, Jimenez-Carvelo, Cuadros-Rodriguez, Ruisanchez, and Callao (2017) employed 3 data fusion schemes from the data gathered from 2 detectors, the ultraviolet detector linked with high-performance liquid chromatography and the electrospray ionization, displaying the synergistic potential of deriving information from 2 chromatograms. The strategy fusing feature-level data designated the highest precision in pinpointing palm oil from distinct sources when compared with the alternative methods. The experimentation intended to utilize the weight of attribute importance as the assessment metric to select attributions and perform chromatographic fusion on the dataset with a higher weight proportion, thus advancing the prediction accuracy of its classifier while the data dimension is also reduced.

The Relief algorithm, an attribute weighting method that assigns distinct weights to attributions based on their correlation with categories (Dash et al., 2011), pinpointed the optimal attribute subset by removing irrelevant attributes. It improved the recognition algorithms's prediction precision while decreasing data dimensionality (Xu et al., 2021). Attributes with importance weights of 0.1 (DAD) and 0.2 (ELSD), respectively, were picked. The outcomes of attribute selection corresponded to distinct chromatographic band areas (The gray areas in parts A and C in Fig. 3). In the HPLC-DAD fingerprint profile spectrum, band areas picked spanned 2.20–4.50 min, 19.85–23.30 min, and 56.00–60.00 min (Fig. 3A), yielding a selection of 1730 variables for the HPLC-ELSD data. An analogous method was adopted to integrate HPLC-CAD data. The chosen band areas in the fingerprint profile spectrum were 2.09–4.74 min, 6.33–7.62 min, 16.72–20.97 min (Fig. 3C). Finally, 556 attributes were picked for concatenation with HPLC-DAD data, leading to a 48*556 dimensional matrix.

3.2.2. Selection and optimization of chromatographic fusion methods

The research developed a PLS-DA approach to examine how the combined chromatographic data impacts the approach's capability to classify accurately. Fig. 4 depicted that PLS-DA had a higher prediction precision for the fingerprint profile spectrum of dark tea after chromatographic data fusion than the combination of the 2 single detectors, with an overall prediction precision of 76.923% (Fig. 4C). This suggested that integrated data could notably advance the discriminative performance of the classification approach for dark tea from varied altitudes, despite there was a need for further advancements in classification precision. This occurred due to the characteristics and scale of data from varied disparate origins, and the constrained effectiveness of the approach built to implement the combined data directly. Thus, it was required to further preprocess the data to enhance data quality and help advance the recognition precision of subsequent approaches.

The classification outcomes of data fusion after data preprocessing by STFT, HT, and IIR, respectively depicted that the recognition precision of STFT (Fig. 4D), HT (Fig. 4E), and IIR (Fig. 4F) were 92.857%, 86.615%, and 86.615% respectively. The STFT processing provided the best precision. In the performance assessment outcomes of the algorithm with the best preprocessing (Table 2), the sensitivity scores of the approach for distinct altitudes of dark tea changed from 79.200% to 91.700%, and the specificity scores altered from 68.400% to 79.200%, respectively. Ultimately, the Receiver Operating Characteristic (ROC) curve, drawn by graphing Sensitivity against 1-Specificity at various cutoff points as depicted in Fig. 5A depicted that the Area Under the ROC Curve (AUROC) for the approach across varied altitudes of dark tea surpassed 0.800. This suggests that the approach designated robust generalization capability.

Table 2
The results of the evaluation of the model performance based on STFT.

Categories	Number of samples	AUROC	Sensitivity	Specificity	Youden index
Samples at high altitude	16	0.850	0.881	0.708	0.589
Samples at low altitude	16	0.840	0.792	0.792	0.583
Samples at medium altitude	16	0.840	0.917	0.684	0.601

3.3. Model validation

To further validate the efficacy of the PLS-DA approach based on a Fast Fourier transform, a permutation test with external cross-validation was implemented. Permutation test, as a 'random algorithm' based on the probabilistic notion, aimed to assess the importance of the approach's outcomes by randomly permuting the response matrix (Y) (de Andrade, de Gois, Xavier, & Luna, 2020) and then the approach was rebuilt to utilize the same modeling settings to compute the probability outcomes that occur by chance (Lopez, Etxebarria-Elezgarai, Amigo, & Seifert, 2023). It serves as a powerful tool to evaluate the validity of regression methodologies (Ballabio & Consonni, 2013), where the Q^2 score is used to gauge the predictive capability, while R^2 is employed to measure the explanatory power. In the research, the permutation test was repeated 200 times ($n = 200$). Although a higher number of repetitions leads to a more stable background distribution, excessive repetitions can put a heavy burden on computational resources and reduce the learning efficacy of machine learning approaches.

Fig. 5B depicts that all the R^2 and Q2 scores to the left are less than those on the extreme right, implying a lack of overfitting in the altitude-specific dark tea classification approach (Bi et al., 2021). The R^2X, R^2Y, and Q^2 scores recorded as 0.860, 1, and 0.915, respectively, suggest that the variation among dark teas from varied origins after the post-intermediate data fusion was significant, and the resulting approach designated strong predictive performance. Furthermore, an unsupervised PCA approach was implemented to cross-validate the potential natural clustering in the data fusion matrix by implementing STTF and detecting possible outliers. The first 2 principal components of the PCA accounted for 95.2% of the variability in the dataset (PC1 = 73.0%, PC2 = 22.2%). In the score plot (Fig. 5C), dark tea samples from distinct altitudes could be clustered according to low, medium, and high altitude dark teas, respectively, and were discriminated without any outliers, thus depicting that the intergroup differences of dark tea from distinct origins after data fusion were significant. Based on all these performance metrics, it could be considered that the classification approach based on data fusion unified with STTF could be employed to identify dark tea from distinct altitudes.

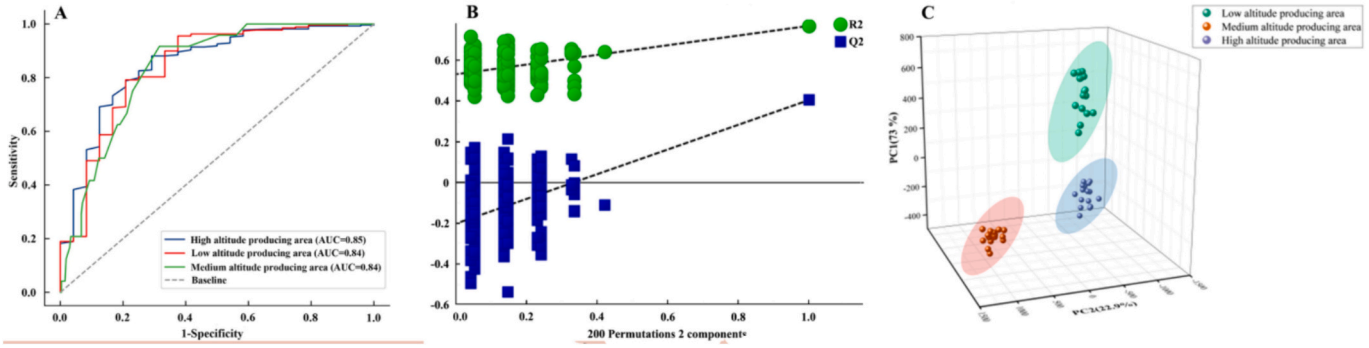


Fig. 5. The model validation results of the STFT-based classification model.

4. Conclusion

The research implemented a strategy fusing feature-level data, unifying fingerprint profiles attained from HPLC-DAD and HPLC-ELSD analyses, to successfully suggest an efficient approach to detect dark teas originating from distinct altitudes. Dark teas presented distinct chromatographic fingerprints when investigated with DAD and ELSD detectors. Note that the classification approach unifying ELSD and DAD sensors designated a higher recognition ratio, highlighting the importance of integrating multi-sensor data in dark tea investigation. Following attribute derivation and the construction of the HPLC-DAD-ELSD fusion approach, the classification precision and model performance of dark teas from distinct altitudes were substantially enhanced. The research outcomes indicate that the fusion approach employing the STFT preprocessing algorithm achieved optimal performance with a classification precision of 92.85%. In conclusion, the combined utilization of the non-targeted HPLC-DAD-ELSD approach and the scheme that fuses feature-level data leads to an economically efficient method to pinpoint the geographical origin of dark teas based on altitude.

CRediT authorship contribution statement

Zhenhong Wang: Writing – original draft, Methodology, Funding acquisition, Conceptualization. **Yuanxi Han:** Writing – original draft, Methodology, Investigation. **Liyou Zhang:** Visualization. **Yongxiang Ye:** Investigation. **Liping Wei:** Conceptualization. **Liang Li:** Supervision, Methodology.

Declaration of competing interest

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

Data availability

Data will be made available on request.

Acknowledgment

The research was financially supported by the National Natural Science Foundation of China (U21A20232) and Graduate Teaching Reform and Construction Project of University-Xizang Agriculture and Animal Husbandry College (YJSJG2023-015).

References

de Andrade, B. M., de Gois, J. S., Xavier, V. L., & Luna, A. S. (2020). Comparison of the performance of multiclass classifiers in chemical data: Addressing the problem of overfitting with the permutation test. *Chemometrics and Intelligent Laboratory Systems, 201*, 7. https://doi.org/10.1016/j.chemolab.2020.104013

Ballabio, D., & Consonni, V. (2013). Classification tools in chemistry. Part 1: Linear models. PLS-DA. *Analytical Methods, 5*(16), 3790–3798. https://doi.org/10.1039/c3ay40582f

Bi, Y. S., Wu, J. W., Zhai, X. R., Shen, S. H., Tang, L. B., Huang, K., & Zhang, D. W. (2021). Application of partial least squares-discriminate analysis model based on water chemical compositions in identifying water inrush sources from multiple aquifers in mines. *Geofluids, 2021*, 17. https://doi.org/10.1155/2021/6663827

Borras, E., Ferre, J., Boque, R., Mestres, M., Acena, L., & Busto, O. (2015). Data fusion methodologies for food and beverage authentication and quality assessment - A review. *Analytica Chimica Acta, 891*, 1–14. https://doi.org/10.1016/j.aca.2015.04.042

Cui, C. J., Xu, Y. F., Jin, G., Zong, J. F., Peng, C. Y., Cai, H. M., & Hou, R. Y. (2023). Machine learning applications for identify the geographical origin, variety, and processing of black tea using 1H NMR chemical fingerprinting. *Food Control, 148*, 12. https://doi.org/10.1016/j.foodcont.2023.109686

Dash, M., Ong, Y. S., & Ieee.. (2011). Relief-C: Efficient feature selection for clustering over noisy data. In *23rd IEEE international conference on tools with artificial intelligence (ICTAI)* (pp. 869–872). Boca Raton, FL: Ieee Computer Soc.

Deng, Z. W., Chen, Z., Fu, J. S., & Yun, Y. H. (2023). Application progress of data fusion strategy in food origin traceability. *Chinese Journal of Analytical Chemistry, 51*(1), 11. ://WOS:000931577800001.

Fang, S. M., Huang, W. J., Wei, Y. M., Tao, M., Hu, X., Li, T. H., & Ning, J. M. (2019). Geographical origin traceability of Keemun black tea based on its non-volatile composition combined with chemometrics. *Journal of the Science of Food and Agriculture, 99*(15), 6937–6943. https://doi.org/10.1002/jsfa.9982

Fu, J., Liu, R. R., Chen, Y. F., & Xing, J. G. (2023). Discrimination of geographical indication of Chinese green teas using an electronic nose combined with quantum neural networks: A portable strategy. *Sensors and Actuators B: Chemical, 375*, 9. https://doi.org/10.1016/j.snb.2022.132946

García-Seval, V., Martínez-Alfaro, C., Saurina, J., Núñez, O., & Sentellas, S. (2022). Characterization, classification, and authentication of Spanish blossom and honeydew honey by non-targeted HPLC-UV and offline SPE HPLC-UV polyphenolic fingerprinting strategies. *Foods, 11*(15), 20. https://doi.org/10.3390/foods11152345

Gu, H. W., Yin, X. L., Peng, T. Q., Pan, Y., Cui, H. N., Li, Z. Q., & Liu, Z. (2022). Geographical origin identification and chemical markers screening of Chinese green tea using two-dimensional fingerprints technique coupled with multivariate chemometric methods. *Food Control, 135*, 9. https://doi.org/10.1016/j.foodcont.2021.108795

Jiang, Q. H., Mei, S., Zhan, C. X., Ren, C. H., Song, Z. Y., & Wang, S. P. (2022). Fast and nondestructive discrimination of fresh tea leaves at different altitudes based on near-infrared spectroscopy and various chemometrics methods. *Food Science and Technology, 43*, 7. https://doi.org/10.1590/fst.98922

Jiang, Y. H., Boorboori, M. R., Xu, Y. A., & Lin, W. X. (2021). The appearance of volatile aromas in Tieguanyin tea with different elevations. *Journal of Food Science, 86*(10), 4405–4416. https://doi.org/10.1111/1750-3841.15898

Jiménez-Carvelo, A. M., Martín-Torres, S., Ortega-Gavilán, F., & Camacho, J. (2021). PLS-DA vs *sparse* PLS-DA in food traceability. A case study: Authentication of avocado samples. *Talanta, 224*. https://doi.org/10.1016/j.talanta.2020.121904

Jin, J. Y., Zhao, M. Y., Zhang, N., Jing, T. T., Liu, H. T., & Song, C. K. (2020). Stable isotope signatures versus gas chromatography-ion mobility spectrometry to determine the geographical origin of Fujian Oolong tea *(Camellia sinensis)* samples. *European Food Research and Technology, 246*(5), 955–964. https://doi.org/10.1007/s00217-020-03469-0

Kfoury, N., Morimoto, J., Kern, A., Scott, E. R., Orians, C. M., Ahmed, S., & Robbat, A. (2018). Striking changes in tea metabolites due to elevational effects. *Food Chemistry, 264*, 334–341. https://doi.org/10.1016/j.foodchem.2018.05.040

Klikarova, J., & Ceslova, L. (2022). Targeted and non-targeted HPLC analysis of coffee-based products as effective tools for evaluating coffee authenticity. *Molecules, 27*(21), 24. https://doi.org/10.3390/molecules27217419

Li, Y. J., Lei, J. C., Yang, J. N., & Liu, R. H. (2014). Classification of Tieguanyin tea with an electronic tongue and pattern recognition. *Analytical Letters, 47*(14), 2361–2369. https://doi.org/10.1080/00032719.2014.908381

Liu, H. L., Meng, Q., Zhao, X., Ye, Y. L., & Tong, H. R. (2021). Inductively coupled plasma mass spectrometry (ICP-MS) and inductively coupled plasma optical emission

spectrometer (ICP-OES)-based discrimination for the authentication of tea. *Food Control, 123*, 8. https://doi.org/10.1016/j.foodcont.2020.107735

Lopez, E., Etxebarria-Elezgarai, J., Amigo, J. M., & Seifert, A. (2023). The importance of choosing a proper validation strategy in predictive models. A tutorial with real examples. *Analytica Chimica Acta, 1275*, 15. https://doi.org/10.1016/j.aca.2023.341532

Ma, W. J., Shi, Y. L., Yang, G. Z., Shi, J., Ji, J. P., Zhang, Y., & Lv, H. P. (2022). Hypolipidaemic and antioxidant effects of various Chinese dark tea extracts obtained from the same raw material and their main chemical components. *Food Chemistry, 375*, 10. https://doi.org/10.1016/j.foodchem.2021.131877

Mao, Y., Wei, B. Y., Teng, J. W., Xia, N., Zhao, M. M., Huang, L., & Ye, Y. (2018). Polysaccharides from Chinese Liupao dark tea and their protective effect against hyperlipidemia. *International Journal of Food Science and Technology, 53*(3), 599–607. https://doi.org/10.1111/ijfs.13633

Núñez, N., Martínez, C., Saurina, J., & Núñez, O. (2021). High-performance liquid chromatography with fluorescence detection fingerprints as chemical descriptors to authenticate the origin, variety, and roasting degree of coffee by multivariate chemometric methods. *Journal of the Science of Food and Agriculture, 101*(1), 65–73. https://doi.org/10.1002/jsfa.10615

Obisesan, K. A., Jimenez-Carvelo, A. M., Cuadros-Rodriguez, L., Ruisanchez, I., & Callao, M. P. (2017). HPLC-UV and HPLC-CAD chromatographic data fusion for the authentication of the geographical origin of palm oil. *Talanta, 170*, 413–418. https://doi.org/10.1016/j.talanta.2017.04.035

Pan, H. J., Le, M. M., He, C. N., Yang, C. S., & Ling, T. J. (2023). Dark tea: A popular beverage with possible medicinal application. *Chinese Herbal Medicines, 15*(1), 33–36. https://doi.org/10.1016/j.chmed.2022.08.005

Pinu, F. R. (2018). Grape and wine metabolomics to develop new insights using untargeted and targeted approaches. *Fermentation-Basel, 4*(4), 23. https://doi.org/10.3390/fermentation4040092

Qu, J. Y., Ye, M. K., Wen, C., Cheng, X. Y., Zou, L. R., Li, M. Y., & Wang, J. (2023). Compound dark tea ameliorates obesity and hepatic steatosis and modulates the gut microbiota in mice. *Frontiers in Nutrition, 10*, 14. https://doi.org/10.3389/fnut.2023.1082250

Su, H., Wu, W. Q., Wan, X. C., & Ning, J. M. (2019). Discriminating geographical origins of green tea based on amino acid, polyphenol, and caffeine content through high-performance liquid chromatography: Taking Lu'an guapian tea as an example. *Food Science & Nutrition, 7*(6), 2167–2175. https://doi.org/10.1002/fsn3.1062

Sun, X. D., Zhang, M., Zhang, S., Chen, Y. X., Chen, J. H., Wang, P. J., & Gao, X. L. (2024). Classification of *Rosa roxburghii* Tratt from different geographical origins using non-targeted HPLC-UV-FLD fingerprints and chemometrics. *Food Control, 155*, 8. https://doi.org/10.1016/j.foodcont.2023.110087

Sun, X. D., Zhang, M., Zhang, S., Wang, P. J., Chen, J. H., & Gao, X. L. (2023). Non-targeted HPLC-FLD fingerprinting for the classification, authentication, and fraud quantitation of Guizhou paprika by chemometrics. *Journal of Food Composition and Analysis, 120*, 8. https://doi.org/10.1016/j.jfca.2023.105346

Wang, C. M., Nie, C. N., Du, X., Xu, W., Zhang, X., Tan, X. Q., & Li, P. W. (2022). Evaluation of sensory and safety quality characteristics of "high mountain tea". *Food Science & Nutrition, 10*(10), 3338–3354. https://doi.org/10.1002/fsn3.2923

Wang, M., Li, J. L., Liu, X. H., Liu, C. S., Qian, J. J., Yang, J., & Zeng, L. T. (2022). Characterization of key odorants in Lingtou Dancong oolong tea and their differences induced by environmental conditions from different altitudes. *Metabolites, 12*(11), 19. https://doi.org/10.3390/metabo12111063

Wang, T. S., Liang, A. R. D., Ko, C. C., & Lin, J. H. (2022). The importance of the region of origin and geographical labeling for tea consumers: The moderating effect of traditional tea processing method and tea prices. *Asia Pacific Journal of Marketing and Logistics, 34*(6), 1158–1177. https://doi.org/10.1108/apjml-02-2021-0121

Wu, X., Liu, Y., Guo, J. Q., Wang, J. X., Li, M. Z., Tan, Y. Z., & Feng, Y. F. (2021). Differentiating Pu-erh raw tea from different geographical origins by ^{1}H-NMR and U-HPLC/Q-TOF-MS combined with chemometrics. *Journal of Food Science, 86*(3), 779–791. https://doi.org/10.1111/1750-3841.15624

Xu, F., Kong, F. Z., Peng, H., Dong, S. F., Gao, W. Y., & Zhang, G. T. (2021). Combing machine learning and elemental profiling for geographical authentication of Chinese geographical indication (GI) rice. *npj Science of Food, 5*(1), 6. https://doi.org/10.1038/s41538-021-00100-8

Yun, J., Cui, C. J., Zhang, S. H., Zhu, J. J., Peng, C. Y., Cai, H. M., & Hou, R. Y. (2021). Use of headspace GC/MS combined with chemometric analysis to identify the geographic origins of black tea. *Food Chemistry, 360*, 9. https://doi.org/10.1016/j.foodchem.2021.130033

Zhang, L. Z., Dai, H. M., Zhang, J. L., Zheng, Z. Q., Song, B., Chen, J. Y., & Huang, Y. (2023). A study on the origin traceability of white tea (white peony) based on near-infrared spectroscopy and machine learning algorithms. *Foods, 12*(3), 24. https://doi.org/10.3390/foods12030499

Zheng, K., Zhao, Q., Chen, Q., Xiao, W. Q., Jiang, Y. D., & Jiang, Y. H. (2019). The synergic inhibitory effects of dark tea (Camellia sinensis) extract and p38 inhibition on the growth of pancreatic cancer cells. *Journal of Cancer, 10*(26), 6557–6569. https://doi.org/10.7150/jca.34637

Zhu, M. Z., Li, N., Zhou, F., Ouyang, J., Lu, D. M., Xu, W., & Wu, J. L. (2020). Microbial bioconversion of the chemical components in dark tea. *Food Chemistry, 312*, 18. https://doi.org/10.1016/j.foodchem.2019.126043