Current Research in Food Science 8 (2024) 100654

Contents lists available at ScienceDirect

Current Research in Food Science

journal homepage: www.sciencedirect.com/journal/current-research-in-food-science

Rapid identification of the geographical origin of Baimudan tea using a Multi-AdaBoost model integrated with Raman Spectroscopy

Wei Pan [a,*], Wenjing Liu [a], Xiujuan Huang [b,**]

[a] Fujian Key Laboratory of Agro-products Quality & Safety, Institute of Agricultural Quality Standards and Testing Technology, Fujian Academy of Agricultural Sciences, Fuzhou, 350003, People's Republic of China

[b] Fujian Saifu Food Inspection Institute Co. Ltd., Fuzhou, 350003, People's Republic of China

ARTICLE INFO

Keywords:
Baimudan tea
Raman spectroscopy
Geographical origin
Multi-AdaBoost

ABSTRACT

The potential of Multi-AdaBoost in spectral analysis is substantial, particularly when combined with weak classifiers and trained to develop into a robust classifier. Given the variable quality of Baimudan tea sourced from diverse regions, the novel application of Raman spectroscopy in conjunction with the Multi-AdaBoost model to analyze the geographic origin of Baimudan tea was introduced. Initially, Raman spectra of Baimudan tea from four distinct origins in Fujian province were gathered, namely Fuan (FA), Fuding (FD), Zhenghe (ZH), and Songxi (SX). Decision Tree (DT) and Support Vector Machine (SVM) models were employed as fitting classifiers to construct the Multi-AdaBoost-DT and Multi-AdaBoost-SVM models. The results demonstrated that the Multi-AdaBoost-DT model exhibited significantly improved recognition rates for FA, FD, ZH, and SX origin compared to the DT model, with the average recognition rate increasing from 86.46% to 91.67%. In contrast, the recognition rates for FA and SX origin in the Multi-AdaBoost-SVM model remained unchanged, attributed to the model having reached a local optimum. The recognition rates of FD origin increased from 91.67% to 95.83%, a significant improvement, while those of ZH origin escalated from 83.33% to 87.50%. The average recognition rate increased from 92.71% to 94.79%. Additionally, Multi-AdaBoost-SVM and Multi-AdaBoost-DT enhanced the sensitivity and specificity of the discrimination outcomes. These results corroborated the effectiveness of the proposed Multi-AdaBoost-SVM model in identifying the geographical origin of Baimudan tea. Moreover, the Multi-AdaBoost model demonstrates potential in elevating the discrimination accuracy of weak classifiers, which bodes well for its application in food authentication and quality control.

1. Introduction

Tea has been a part of many cultures around the world for centuries, and its popularity continues to grow because of the health benefits, variety of flavors and cultural significance. It is believed that white tea was originated hundreds or even thousands of years ago in the Fujian province of China (Feng et al., 2022). Baimudan tea, also known as white Peony, constitutes a popular variety of white tea in China. The altitude, climate, and soil conditions of a region can greatly affect the growth and development of Baimudan tea plants, as well as the chemical composition of the tea leaves (Shuai et al., 2022; Wang et al., 2022). Therefore, there are obvious differences in the quality of Baimudan tea from different regions. Given that the quality of Baimudan tea varies across different regions, it is imperative to determine its specific provenance. This is particularly important for consumers who are willing to pay a premium for Baimudan tea, as they often seek an authentic and genuine experience (Qu et al., 2020). Furthermore, the presence of unscrupulous merchants who sell substandard tea to maximize profits is not only illegal but also unethical. Therefore,it is important to be able to identify the origin of Baimudan tea.

The fundamental properties of tea sourced from various regions might exhibit similarities, yet subtle distinctions can be discerned by experienced tea connoisseurs. In addition, some analytical instruments can be used, for example, liquid chromatography (Fraser et al., 2013), gas chromatography mass spectrometry (Zhu et al., 2021), mineral

* Corresponding author. Institute of Quality Standards & Testing Technology for Agro-products, Fujian Academy of Agricultural Sciences/Fujian Key Laboratory of Agro-products Quality and Safety, No. 247, Wusi Road, Gulou District, Fuzhou, Fujian, 350003, People's Republic of China.

** Corresponding author. Fujian Saifu Food Inspection Institute Co. Ltd., Building 47 and 51-52, Xihe Park, Fujian Institute of Physical Construction, CAS, 155 Yangqiao West Road, Hongshan Town, Gulou District, Fuzhou, Fujian, 350003, People's Republic of China.

E-mail addresses: 870993292@qq.com (W. Pan), 595846743@qq.com (X. Huang).

https://doi.org/10.1016/j.crfs.2023.100654
Received 4 October 2023; Received in revised form 1 December 2023; Accepted 5 December 2023
Available online 9 December 2023

Abbreviations

Words	*Abrr*
PCA	Principal Component Analysis
KNN	K-Nearest Neighbor
SVM	Support Vector Machine
MLP	Multi-Layer Perceptron
DT	Decision Tree
FA	Fuan
FD	Fuding
ZH	Zhenghe
SX	Songxi
SAMME	Stagewise Additive Modeling
TP	True Positive
FP	False Positive
TN	True Negative
FN	False Negative

element analysis (Liu et al., 2021; Mayjurek et al., 2020), infrared spectroscopy (Esteki et al., 2022; Liu et al., 2022), X-ray fluorescence spectrometry (Lim, Carey, Paul N. Williams and Koidis, 2021), inductively coupled plasma mass spectrometry (Liu et al., 2020) and the rest.

Precise determination of the origin of tea can be achieved through the use of analytical instruments, which offer greater accuracy compared to relying solely on visual or sensory characteristics. However, this method is relatively expensive and time-consuming, making it more suitable for researchers and specialty tea companies rather than individual consumers. Consequently, the development of a rapid and straightforward method for individual consumers to identify the origin of Baimudan tea is of significant importance.

Raman Spectra are capable of detecting and identifying molecular vibration information, boasting the advantages of rapidity, non-destructiveness, and sensitivity. Chemometrics can be employed to analyze and interpret the complex data generated by Raman spectroscopy, while Raman spectroscopy itself can provide valuable information concerning the molecular structure and chemical composition of the sample. This combination of techniques can be used to identify and quantify compounds in complex mixtures, for example, rice origin (Sha et al., 2021; Wang et al., 2021), pesticide and veterinary drug residue (Girmatsion et al., 2021; Lin et al., 2021), egg freshness (Liu et al., 2020; Sasikan Katemala, Molee, Thumanu and Yongsawatdigul, 2021), meat preservation (Yang et al., 2020; Robert et al., 2020), illegal additives in food (Karunathilaka et al., 2018) and other fields. So far, the identification of origin of Baimudan tea based on Raman spectroscopy has not been reported.

In this study, we developed an approach to realize rapid, nondestructive, precise identification of the origin of Baimudan tea by Raman spectroscopy. At the same time, deep learning method was introduced to identify the origin of Baimudan tea more effectively and accurately. Deep learning, as a tool of chemometrics, has the potential to significantly improve the generalization ability of analytical models. And AdaBoost algorithm is one of the typical ensemble learning algorithms. As a chemometric tool, deep learning has the potential to significantly enhance the generalization capability of analytical models. The AdaBoost algorithm, a prominent ensemble learning algorithm, was employed in this study. While traditional AdaBoost is commonly used for binary classification applications, Multi-AdaBoost was extended to address complex multi-classification problems, rendering it more suitable for such scenarios. Furthermore, Multi-AdaBoost holds great potential for application in spectral analysis (Li and Rong, 2022).

In other words, Raman spectroscopy combined with Multi-AdaBoost algorithm was used to identify Baimudan tea from different geographical origin. Initially, Raman spectra were collected from four areas in Fujian province, namely Fuding, Fuan, Zhenghe, and Songxi. Principal Component Analysis (PCA) was utilized to extract relevant features. Subsequently, classification models, including K-Nearest Neighbor (KNN), Support Vector Machine (SVM), Multi-Layer Perceptron (MLP), and Decision Tree (DT), were constructed. Ultimately, employing the Multi-AdaBoost ensemble learning algorithm, the classifier model with superior discriminatory capabilities was further optimized. By using Raman spectroscopy combined with Multi-AdaBoost, it would be possible to accurately identify the geographical origin of Baimudan tea. This could be particularly useful in fields such as agriculture and food science, where accurate identification of crops and ingredients is crucial.

2. Materials and methods

2.1. Samples

The samples of Baimudan tea (white tea) were collected from four producing areas in Fujian province, namely Fuan (FA), Fuding (FD), Zhenghe (ZH) and Songxi (SX). The picking standard for Baimudan tea dictated that the top bud and two adjacent leaves were to be harvested from the tender shoot of the tea plant during the initial harvest in the spring of 2022, with the length of the top bud and two leaves being approximately equal. Furthermore, the "three white" criteria had to be met, which required that the buds and two leaves be covered with white fluff. A total of twenty samples were collected from each region, all of which were locally sourced tea greens from various towns. Table 1 of the supplementary material detailed the sampling locations of the various samples. The authentication of all samples was conducted by Professor Su Feng (a senior tea evaluator at the Fujian Provincial Planting Technology Extension Station). Fig. 1(a) shows the comparison of Baimudan tea from these four different producing areas. Subsequently, the samples were dried in an oven at 60 °C to eliminate moisture before being stored in a dryer. After cooling, the samples were mechanically ground into a powder, filtered through a 60-mesh sieve, and finally stored in the dryer.

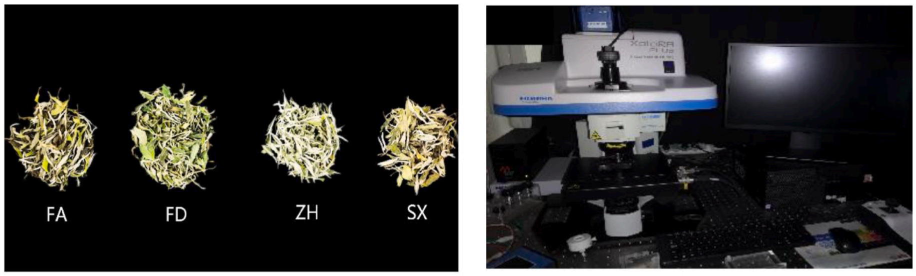
2.2. Collection of Raman Spectra

The appropriate quantity of tea powder was meticulously placed in the center of a spotlessly clean slide, and compressed using a cover slip. Raman spectra were subsequently acquired using an Xplora PLUS Raman spectrometer equipped with a CCD detector (HORIBA, France), as illustrated in Fig. 1 (b). The spectrometer parameters were set as follows: the numerical aperture of the 50× objective was 0.55NA, the laser wavelength was 785 nm (10% laser power, 3.6 mW), the objective lens was 50 times, the grating was 1200 gr/nm, the wavelength range spanned 100-3700 cm^{-1}, the spectral resolution was 1.3 cm^{-1}, and each Raman spectrum was scanned for a duration of 8 s. The environmental conditions during the collection of spectra were maintained at 23 °C with a relative humidity of 45%. Each sample was tested four times, with the sampling locations for each test being randomly selected.

Table 1
Tentative assignment of the Raman bands in Spectra of Baimudan tea.

Bands of different origins	Assignments	Compound in Baimudan tea(Yang and Wu, 2008)
732s, 1110v	δ(CH_2), γ(benzene-ring), γ(C–H),γ(OH)	Catechin, tea polyphenols
606w, 805m, 1176sh	ν(C–N), γ(Amide I^-), γs (C–O–C), ν(NH_3^+)	protein
1648br	γas (COO^-), δas (NH_3^+)	Theanine
1960s	ν(CH), ν(C=O)	Insaturated fatty acid, organic acid

Assignments: γ: Stretching; δ: Deformation; s: Symmetric; as: asymmetric.
Intensities: w: Weak; m: Medium; s: Strong; v: Very; sh: Shoulder; br: Broad.



(A) (B)

Fig. 1. Materials and environmental instruments: (a) Comparison of Baimudan tea from different producing areas; (b) the Xplora PLUS Raman spectrometer used in the experiment.

2.3. Data processing

2.3.1. Division of spectral data

A total of 320 spectra were collected from four origins, with each origin contributing 80 spectra. Subsequently, the spectral data were apportioned into a calibration set and a prediction set in a ratio of 7:3. The calibration set encompassed 56 samples per species, while the prediction set included 24 samples per species. Firstly, the every spectral was labeled with a serial number. Then, the program could generate a serial of random number representing calibration set within [0,320] according to random seed. Later, calibration set could be selected according to random number and the remaining data was marked as prediction set. In other words, 224 spectra of the calibration set were used to train the parameters of the model, and 96 spectra of the prediction set were used to verify the predicted effect of the model. The spectral data were processed by Scikit-learn machine learning library in python and the pictures were plotted by Origin 9.0.

2.3.2. Construction of traditional model

The construction of classifier models, including KNN, MLP, DT, and SVM, was based on the calibration set and prediction set. The MLP model, specifically, is a type of feed-forward neural network consisting of an input layer, a hidden layer, and an output layer. It processes input characteristics and output results through weighting and nonlinear operations, utilizing the back-propagation method for network training. This enables the network to effectively handle linear non-fractional data and achieve accurate classification (Mrugalski et al., 2008). DT model is a predictive scheme based on a tree structure, which may be binary or non-binary in nature. It establishes a mapping relationship between object values and attribute values, typically consisting of a root node, multiple internal nodes, and several leaf nodes, among others. The middle node represents the decision outcome in DT, while each internal node signifies a judgment on an attribute. Each branch corresponds to the output of a judgment result, and the sample set within each node is subsequently partitioned into child nodes (Leo. Breiman, Friedman, Olshen and Stone, 1984).The KNN model segments the feature vector space based on the training dataset, determines the distance between the new sample and each category by comparing their feature sets, and predicts the category through majority voting and other methods (Workman and Weyer, 2007). The SVM model, a supervised binary classification approach, fundamentally involves identifying a separation hyperplane that accurately categorizes diverse sample data, while maximizing the distance from the point nearest to the hyperplane to differentiate between categories (Zhu and Dai, 2005).

The parameter optimization during operation of KNN and SVM were discussed below. Numerical test results show that exhaustive search has good performance on small data set. Grid search algorithm is used to process SVM and KNN hyper parameter. K-fold Cross validation is also taken into account in hyper parameter optimization, with k = 5. Actually, grid search is greedy algorithm to adjust every hyper parameter to best performance and it is a perfect fit on small data set scenarios.

2.3.3. Construction of multi-AdaBoost-DT and multi-AdaBoost-SVM model

A weak classifier refers to a model that performs slightly better than random chance. In contrast to strong or powerful classifiers, weak classifiers have limited predictive capacity based on the available data. In order to further improve the discrimination and generalization ability of the classifiers, the Multi-AdaBoost multi-classifier model proposed by Zhu in 2009 was used for analysis (Zhu et al., 2009). The concept is to selectively discard a portion of the examples from the original dataset and solve the optimization problem using a reduced training set. This approach remains effective as long as the overall training error rate stays below 50%. By integrating weakened classifiers directly into Multi-Adaboost, through augmenting weights assigned to misclassified samples while diminishing weights assigned to correctly classified samples, we are able to improve the misclassification rate. Finally, a series of weak classifiers are trained and combined into a powerful classifier in a serial manner. Differently, Multi-AdaBoost introduced the Stagewise Additive Modeling (SAMME) algorithm, which uses multi-class exponential loss functions to rapidly reduce errors to a low value. So, the model can adaptively implement multi-class Bayesian rules by fitting the additive model of multi-class problems. Fig. 2 shows the flow chart of Multi-AdaBoost classifier ensemble learning model.

According to the principle of Multi-AdaBoost classifier ensemble learning model, the fitting classifiers can be DT, SVM, logistic regression, and the rest. In this experiment, DT and SVM models are selected as fitting classifiers. Therefore, the algorithm flow of the constructed Multi-AdaBoost-DT and Multi-AdaBoost-SVM models were shown in Fig. 3. As can be seen from the figure, DT or SVM classifier model was trained using the divided calibration set, and the distribution of sample weight was adjusted according to the discriminant results of four categories by the classifier. The adjustment way was to improve the weight of wrongly discriminated samples and reduce the weight of correctly discriminated samples.

2.3.4. Performance parameters of model

The resulting models were evaluated by the recognition rates, which was defined as:

$$\text{recognition rate} = \frac{\text{the number of samples correctly identified by model}}{\text{the number of samples}} \times 100\% \quad (1)$$

In order to compare the effect of Multi-AdaBoost model in detail, sensitivity and specificity were also used as supplementary parameters. And sensitivity and specificity were defined as:

$$\text{Sensitivity} = \frac{TP}{TP + FN} \times 100\% \quad (2)$$

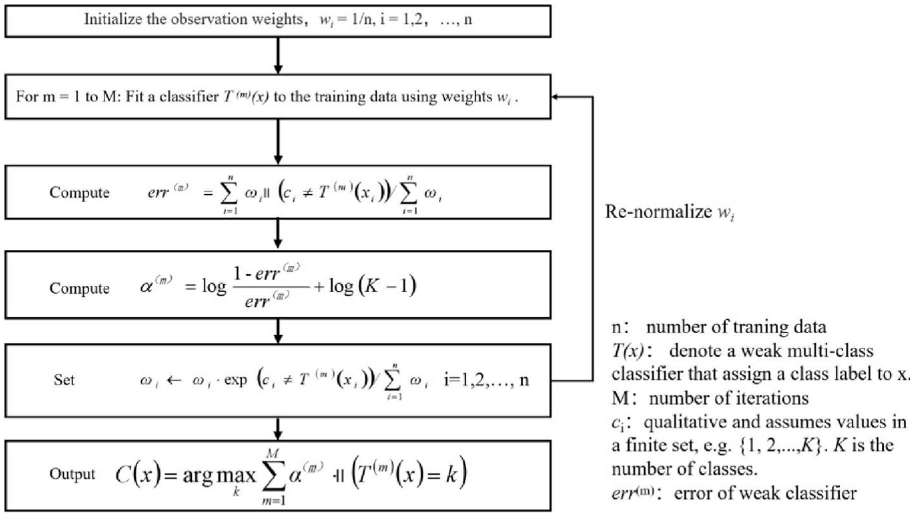


Fig. 2. Flow chart of Multi-AdaBoost classifier ensemble learning model.

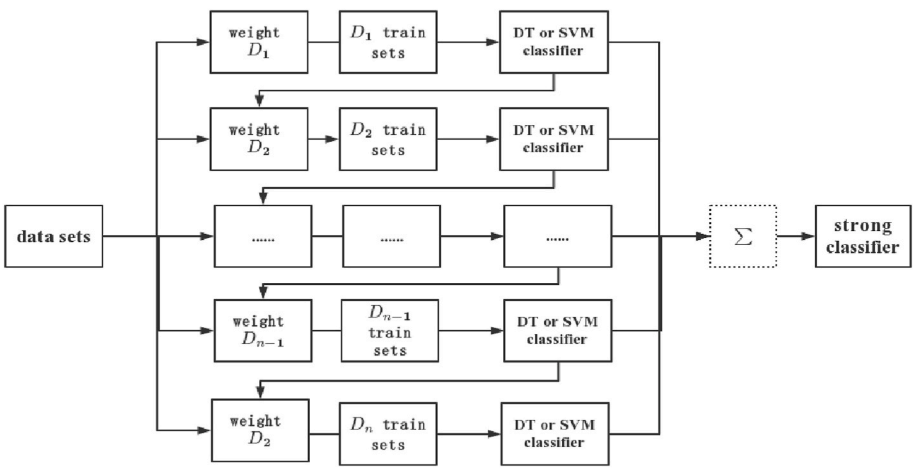


Fig. 3. The algorithm flow diagram of the Multi-AdaBoost-DT and Multi-AdaBoost-SVM model.

$$\text{specificity} = \frac{FP}{FP + TN} \times 100\% \tag{3}$$

where True Positive (TP) is the number of actually positive and predicted to be positive, False Positive (FP) is the number of actually negative but predicted to be positive, True Negative (TN) is the number of actually negative and is predicted to be negative, False Negative (FN) is the number of actually positive but predicted to be negative.

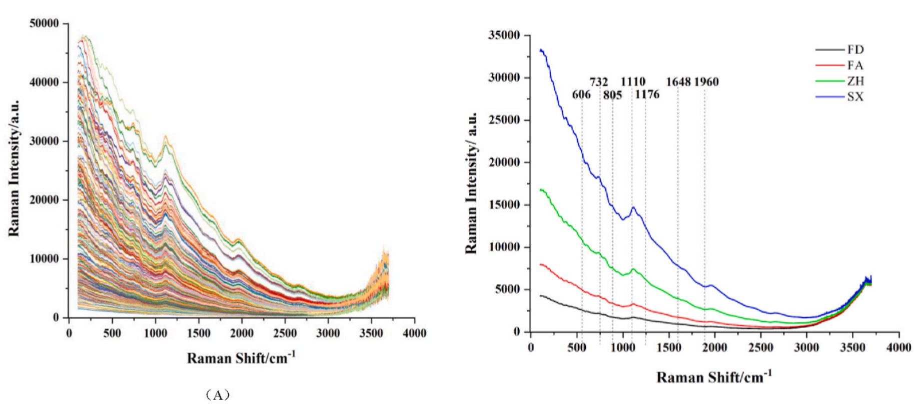


Fig. 4. Raman Spectra of Baimudan tea: (a) all samples from four producing areas; (b) the average spectra of each producing areas.

3. Results and discussion

3.1. Spectral analysis

Fig. 4 shows the Raman spectra of Baimudan tea samples from different origins in the range of 100~2700 cm^{-1}. Table 1 shows the characteristic peaks of Raman spectral of Baimudan tea, and the different characteristic peaks respectively correspond to the substances in the Baimudan tea (Wang et al., 2021; Yang and Wu, 2008). Fig. 4(a) displays the average spectra of each producing areas. Notable differences are observed among the average spectra, with variations in peak intensity in the Raman spectrum corresponding to the content of catechins, tea polyphenols, amino acids, proteins, unsaturated fatty acids, and organic acids. The results also showed that the nutrient content of Baimudan tea from different producing areas was different. However, as can be seen from Fig. 4(b), the Raman spectra of all samples from different origin exhibited a high degree of similarity, with similar peak positions and peak intensity. Beside, the overlapping and stacking problems between the Raman spectra of different samples were evident, thereby rendering it challenging to directly analyze the spectra for the purpose of distinguishing Baimudan tea from diverse origins.

Therefore, PCA was employed to analyze the Raman spectra of Baimudan tea and two principal components were obtained, and the cumulative contribution rate of the principal components reached 99.37%. Fig. 5 shows the two-dimensional principal component analysis score map of Baimudan tea from different origins. It is distinct that some samples are overlapped and deviated, but on the whole, the four categories are distributed in different areas, indicating that it is feasible to identify Baimudan tea from the four producing areas by Raman spectroscopy combined with chemometrics.

3.2. Analysis and comparison of traditional models

Table 2 shows the identification results by KNN, SVM, MLP and DT models. As can be seen from the table, the recognition rates of FA by of four models was 95.83%, 100%, 95.83% and 91.67%, respectively, and the recognition rates of SVM model was the highest. The recognition rates of FD were 100%, 91.67%, 79.17% and 83.33% respectively, and the recognition rates of KNN model was the highest. The identification accuracy of ZH were 79.17%, 83.33%, 75.00% and 83.33%, respectively, and the accuracy of SVM model was the highest. The recognition rates of SX were 95.83%, 95.83%, 87.50%, 87.50%, respectively.

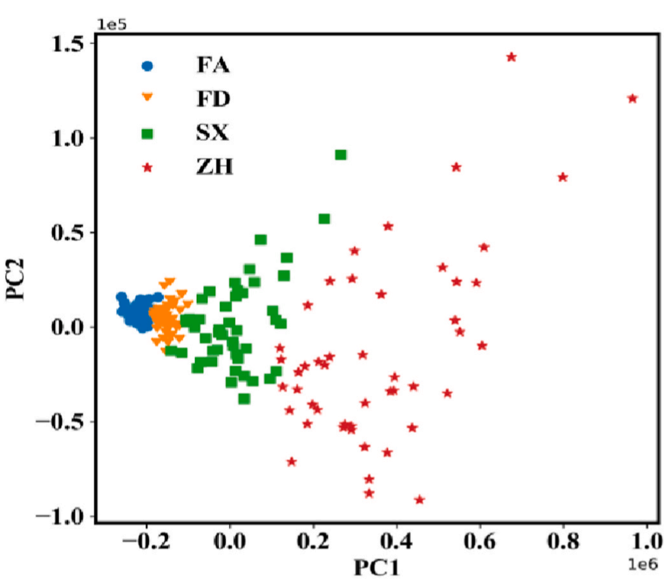


Fig. 5. The score plot from the principal component analysis of Baimudan tea.

Table 2
The results of Baimudan tea from different origins by different discriminant analysis models.

Model Origins	FA	FD	ZH	SX	Average Recognition Rate
KNN	95.83%	100.00%	79.17%	95.83%	92.71%
SVM	100.00%	91.67%	83.33%	95.83%	92.71%
MLP	95.83%	79.17%	75.00%	87.50%	84.38%
DT	91.67%	83.33%	83.33%	87.50%	86.46%

Based on the identification results, the FA samples exhibited the best identification outcomes, while the ZH samples displayed the worst performance. It was observed that the likelihood of misidentifying ZH as FD and SX samples was increased due to the minimal differences between the spectra of ZH, FD, and SX origin samples. On the whole, the average of recognition rate of Baimudan tea from the four origins by KNN, SVM, MLP and DT models reached 92.71%, 92.71%, 84.37% and 86.46%, respectively. The SVM model demonstrates the highest recognition rate, potentially due to its ability to select the optimal SVM base kernel function, which in turn enables the identification of the optimal penalty factor and kernel parameters to maximize the performance of the SVM model. Consequently, the SVM model exhibits an enhanced performance in the discrimination of Baimudan tea, even with a limited number of samples.

3.3. Analysis and comparison of Multi-AdaBoost models

When the dataset comprises a larger number of features, the KNN model necessitates significant computational resources, suffers from poor interpretability, and exhibits high dependence. In order to further optimize the discrimination effect of the model, DT model and SVM model with strong interpretation were used as fitting classifiers to construct Multi-AdaBoost-DT and Multi-AdaBoost-SVM models. And the results are shown in Table 3. Compared with DT model, the recognition rates of Multi-AdaBoost-DT model for FA, FD, ZH and SX was increased from 91.67% to 95.83%, 83.33%–87.5%, 83.33%–91.67%, 87.50%–91.67%, respectively. The average of recognition rate increased from 86.46% to 91.67%. Compared with the SVM model, the recognition rates of the Multi-AdaBoost-SVM model for FA and SX remained unchanged, possibly because the recognition rates of Multi-Adaboost-SVM model had reached the best. However, the recognition rates of FD origin increased from 91.67% to 95.83%, the recognition rates of ZH origin increased from 83.33% to 87.50%. Consequently, the average of recognition rate increased from 92.71% to 94.79%.

Fig. 6 displays the distribution of the predicted values generated by SVM, DT, Multi-AdaBoost-DT and Multi-AdaBoost-SVM. Samples labeled 1–24 was FA origin, samples labeled 25–48 was FD origin, samples labeled 49–72 was ZH origin, samples labeled 73–96 was SX origin. The true value of FA, FD, ZH and SX were assigned 0, 1, 2 and 3, respectively. Based on Fig. 6(b), the following results were obtained: (1) for FA origin, no samples was misjudged; (2) for FD origin, one sample was misjudged, incorrectly identified as FA; (3) for ZH origin, three samples were misjudged: two were misclassified as FD origin, and one as

Table 3
The recognition rate of Multi-AdaBoost model.

Model Origins	FA	FD	ZH	SX	Average Recognition Rate
SVM	100.00%	91.67%	83.33%	95.83%	92.71%
DT	91.67%	83.33%	83.33%	87.50%	86.46%
Multi-AdaBoost-DT	95.83%	87.50%	91.67%	91.67%	91.67%
Multi-AdaBoost-SVM	100.00%	95.83%	87.50%	95.83%	94.79%

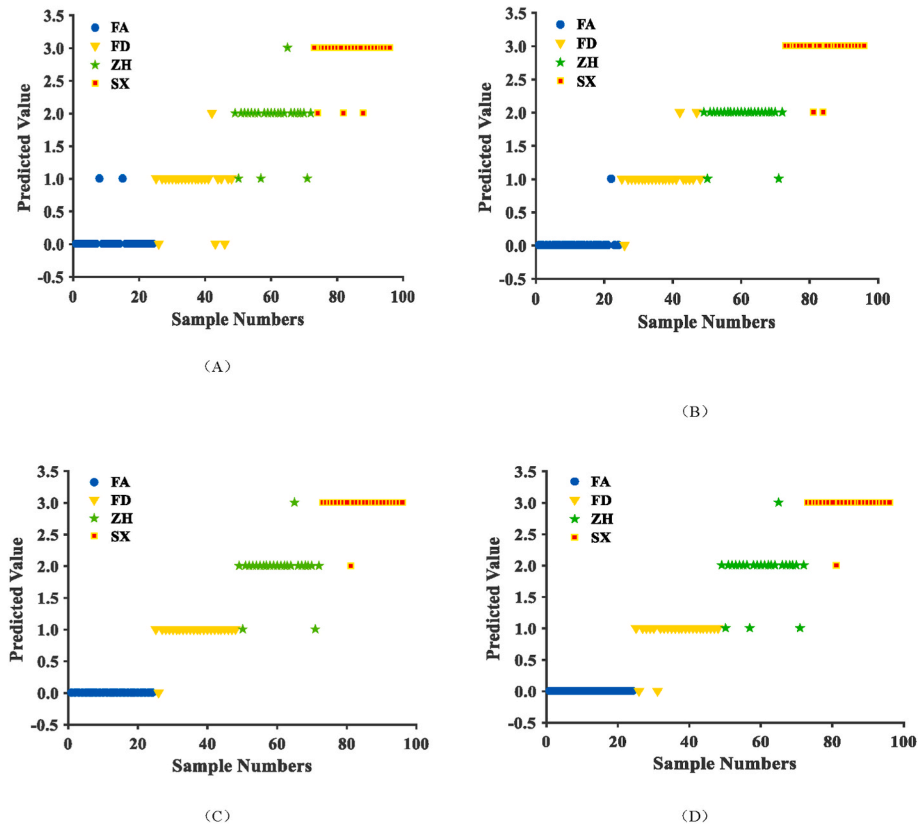


Fig. 6. The distribution of predicted values of different models: (a)SVM model; (b)DT model; (c) Multi-AdaBoost-SVM model; (d) Multi-AdaBoost-DT model.

SX, (4) for SX origin, one sample was misjudged, identified as ZH. Similarly, the discriminant results of other models can be obtained. Compared with the other four models, the Multi-AdaBoost-SVM model demonstrated the least number of misjudged samples.

Table 4 shows the sensitivity and specificity of Multi-AdaBoost model. The details of the methodology were provided in the supplementary materials. It was evident that both sensitivity and specificity were increased by Multi-AdaBoost approach. Notably, the Multi-AdaBoost-SVM model demonstrates the highest sensitivity and specificity. In summary, Tables 3 and 4 demonstrated the effectiveness of the Multi-AdaBoost-SVM model in identifying Baimudan tea. Furthermore, the constructed Multi-AdaBoost-DT and Multi-AdaBoost-SVM models are robust classifier systems that can significantly enhance the classification accuracy of samples.

Table 4
The sensitivity and specificity of Multi-AdaBoost model.

Model Origins		FA	FD	ZH	SX
SVM	Sensitivity	100.00%	91.67%	79.17%	95.83%
	Specificity	100.00%	95.83%	98.61%	97.22%
Multi-AdaBoost-SVM	Sensitivity	100.00%	95.83%	90.91%	95.83%
	Specificity	100.00%	97.22%	98.61%	97.22%
DT	Sensitivity	91.67%	83.33%	79.17%	87.50%
	Specificity	95.83%	93.06%	94.44%	97.22%
Multi-AdaBoost-DT	Sensitivity	95.83%	87.50%	87.50%	91.67%
	Specificity	98.61%	95.83%	94.52%	98.61%

4. Conclusion

In this study, we successfully developed a rapid, nondestructive, and precise method for identifying the origin of Baimudan tea. The proposed Multi-AdaBoost-SVM model has demonstrated exceptional accuracy and outperformed other learning algorithms. The average recognition rate achieved 94.79%. The ZH samples exhibited a sensitivity of 90.91% and specificity of 98.61%, while the other three samples all displayed sensitivities and specificities above 95%. Furthermore, when compared to the DT model and SVM model, both the recognition rate and sensitivity/specificity of the Multi-Adaboost-SVM and Multi-Adaboost-DT models were significantly improved. The results confirmed that the Multi-AdaBoost classifier model can effectively train weak classifiers into a strong classifier, showcasing great potential in the field of food authentication and quality control.

CRediT authorship contribution statement

Wei Pan: conceived and designed the study, conducted the research, wrote and revised the manuscript. **Wenjing Liu:** analyzed the data, review and editing, final draft Supervision, and monitoring. **Xiujuan Huang:** provided technical support for the research, helped revise the manuscript.

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

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

Acknowledgment

The authors gratefully acknowledge the financial support of the

Public Welfare Project Program of Fujian Province (2021R1022008).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.crfs.2023.100654.

References

Breiman, Leo, Friedman, J., Olshen, R., Stone, C., 1984. Classification and Regression Trees. Wadsworth and Brooks, Monterey, CA.

Esteki, M., Memarbashi, N., Simal-Gandara, Jesus, 2022. Classification and authentication of tea according to their geographical origin based on FT-IR fingerprinting using pattern recognition methods - ScienceDirect. J. Food Compos. Anal. 106, 104321.

Feng, Z., Li, M., Li, Y., Yin, J., Wan, X., Yang, X., 2022. Characterization of the key aroma compounds in infusions of four white teas by the sensomics approach. Eur. Food Res. Technol. 248 (5), 1299–1309.

Fraser, K., Lane, G.A., Otter, D.E., Hemar, Y., Quek, S.Y., Harrison, S.J., Rasmussen, S., 2013. Analysis of metabolic markers of tea origin by UHPLC and high resolution mass spectrometry. Food Res. Int. 53 (2), 827–835.

Girmatsion, M., Mahmud, A., Abraha, B., Xie, Y.-f., Cheng, Yu-liang, Yu, Hang, Yao, W.-r., Guo, Y.-h., He, Qian, 2021. Rapid detection of antibiotic residues in animal products using surface-enhanced Raman Spectroscopy: a review. Food Control 126, 108019.

Karunathilaka, S.R., Yakes, B.J., He, K., Brückner, L., Mossoba, M.M., 2018. First use of handheld Raman spectroscopic devices and on-board chemometric analysis for the detection of milk powder adulteration. Food Control 92, 137–146.

Katemala, Sasikan, Molee, A., Thumanu, K., Yongsawatdigul, J., 2021. Meat quality and Raman spectroscopic characterization of Korat hybrid chicken obtained from various rearing periods. Poult. Sci. 100 (2), 1248–1261.

Li, H., Rong, J.Y., 2022. Modeling and analysis of near infrared spectroscopy based on Ada Boost method. Jilin Normal University Journal (Natural Science Edition) 43 (3), 83–89.

Lim, C.M., Carey, M., Williams, Paul N., Koidis, A., 2021. Rapid classification of commercial teas according to their origin and type using elemental content with X-ray fluorescence (XRF) spectroscopy. Curr. Res. Food Sci. 4, 45–52.

Lin, M.H., Sun, L., Kong, F., Lin, M., 2021. Rapid detection of paraquat residues in green tea using surface-enhanced Raman spectroscopy (SERS) coupled with gold nanostars. Food Control 130, 108280.

Liu, H.L., Zeng, Y.T., Zhao, X., Ton, H.R., 2020. Improved geographical origin discrimination for tea by using ICP-MS and ICP-OES techniques combined chemometric approach. J. Sci. Food Agric. 100 (8), 3507–3516.

Liu, Y.L., Ren, X.N., Yu, H., Cheng, Y.L., Guo, Y.H., Yao, W.R., Xie, Y.F., 2020. Non-destructive and online egg freshness assessment from the egg shell based on Raman spectroscopy - ScienceDirect. Food Control 118, 107426.

Liu, W.W., Chen, Y., Liao, R.X., Zhao, J., Yang, H., Wang, F.H., 2021. Authentication of the geographical origin of Guizhou green tea using stable isotope and mineral element signatures combined with chemometric analysis. Food Control 125, 107954.

Liu, Y., Huang, J., Li, M., Chen, Y., Cui, Q., Lu, C., Wang, Y., Li, L., Xu, Z., Zhong, Y., Ning, J., 2022. Rapid identification of the green tea geographical origin and processing month based on near-infrared hyperspectral imaging combined with chemometrics. Spectrochim. Acta, Part A 267, 120537.

Mayjurek, Z., Zawisza, B., Beer, D.D., Joubert, E., Walczak, B., 2020. Authentication of honeybush and rooibos herbal teas based on their elemental composition. Food Control 123, 107757.

Mrugalski, M., Witczak, M., Korbicz, J., 2008. Confidence estimation of the multi-layer perceptron and its application in fault detection systems. Eng. Appl. Artif. Intell. 21 (6), 895–906.

Qu, W.L., Song, X.M., Qu, J.Y., 2020. Detection of diabetes mellitus based on deep belief network and gradient boosting decision tree. Trans. Chin. Soc. Agric. Eng. 41 (3), 112–120.

Robert, C., Fraser-Miller, S.J., Jessep, W.T., Bain, W.E., Hicks, T.M., Ward, J.F., Craigie, C.R., Loeffen, M., Gordon, K.C., 2020. Rapid discrimination of intact beef, venison and lamb meat using Raman spectroscopy. Food Chem. 343, 128441.

Sha, M., Li, L.C., Huang, J.L., Yu, Q., Liu, C., Fang, X.T., 2021. Effect of data processing method on identification of rice from different geographical origins by Raman spectroscopy. J. Chin. Inst. Food Sci. Technol. 21 (5), 369–376.

Shuai, M.Y., Peng, C.Y., Niu, H.L., Shao, D.L., Hou, R.Y., Cai, H.M., 2022. Recent techniques for the authentication of the geographical origin of tea leaves from camellia sinensis: a review. Food Chem. 374, 131713.

Wang, Y.X., Tan, F., Xin, Y.M., Li, H., Zhao, X.Y., Lu, B.X., 2021. Identification of rice from similar areas with different pretreatment methods of Raman spectrum. Spectrosc. Spectr. Anal. 41 (2), 565–571.

Wang, R.X., Huang, X.X., Li, Q., Li, J., Tan, J., 2022. Comparison of catechins, aroma components and sensory quality of different types of white tea. Science and Technology of Food Industry 43 (5), 315–321.

Workman, J., Weyer, L., 2007. Practical Guide to Interpretive Near-Infrared Spectroscopy.

Yang, H.B., David, L.H., Zhang, Y.M., Zhu, L.X., Dong, P.C., Wang, X.Y., Mao, Y.W, Luo, X., Stephanie, M.F., 2020. Preliminary investigation of the use of Raman spectroscopy to predict beef spoilage in different types of packaging. Meat Sci. 165, 108136.

Yang, X.G., Wu, Q.L., 2008. Raman Spectroscopy Analysis and Application. National Defense Industry Press.

Zhu, Y.P., Dai, N.W., 2005. Text classifier based on SVM decision tree. Pattern Recognit. Artif. Intell. 18 (4), 412–416.

Zhu, J., Zou, H., Rosset, S., Hastie, T., 2009. Multi-class AdaBoost. Stat. Interface 2, 349–369.

Zhu, J.C., Niu, Y.W., Xiao, Z.B., 2021. Characterization of the key aroma compounds in Laoshan green teas by application of odour activity value (OAV), gas chromatography-mass spectrometry-olfactometry (GC-MS-O) and comprehensive two-dimensional gas chromatography mass spectrometry (GC × GC-qMS). Food Chem. 339, 128136.