ORIGINAL STUDY

Development and industrial application of geographical origin identification for Taiwanese oolong tea

Tien-Lin Liu [a], Jia-Ru Dai [a], Tsung-Chen Su [a], Chun-Huo Chiu [b], Hsien-Tsung Tsai [a], Chui-Feng Chiu [a], Jin-Chih Lin [a], Chih-Yi Hu [a,*]

[a] Tea and Beverage Research Station (TBRS), No.324, Chung-Hsing RD., Yangmei, Taoyuan City 326011, Taiwan, R.O.C.
[b] Department of Agronomy, National Taiwan University, No. 1, Sec. 4, Roosevelt Rd., Taipei 10617, Taiwan, R.O.C.

Abstract

Taiwanese oolong tea is renowned for its excellent quality and enjoys a prestigious reputation both domestically and internationally. In recent years, there has been an issue with imported Taiwanese-style oolong tea being sold as genuine Taiwanese oolong tea, which has adversely affected the brand value of Taiwanese oolong tea. In this study, samples of domestic oolong tea (Taiwanese oolong tea) and Taiwanese-style oolong tea produced abroad (including China, Vietnam, Indonesia, Thailand, etc.) were collected. A multi-elements analysis method was applied to establish an elemental database of tea leaf samples. Subsequently, various widely used classification methods were employed to develop a discrimination model for identifying the origin of Taiwanese oolong tea. Utilizing the discrimination model established from a database of 727 samples to determine whether the tea leaves were Taiwanese or external, the statistical performances of classification models such as LDA, Ridge, Random Forest, Boosting, and SVM are nearly consistent. These models achieved an accuracy rate of 97.1%−97.8%, a recall rate (true positive rate) for Taiwanese origin of 98.4%−99.0%, and a precision value for predicting Taiwanese origin of 97.3%−97.8%. This identification technology has become an officially recognized and publicly recommended testing method in Taiwan (TFDAF0032.00, released on November 5, 2021) and has been effectively utilized in official administrative inspections for identification of origin, as well as providing evidence for investigative cases.

Keywords: *Camellia sinensis*, Geographical origin identification, Taiwanese oolong tea, Taiwanese-style oolong tea, Tea

1. Introduction

Taiwanese tea has been famous since ancient times, and currently, there are eight major specialty teas [1]. Among them, the two most famous Taiwanese oolong teas are High-mountain oolong tea and Dongding oolong tea. Unfortunately, unscrupulous business runners have been importing Taiwanese-style oolong tea to counterfeit authentic Taiwanese oolong tea, which not only infringes on the rights of producers and consumers but also poses a threat to the Taiwanese tea industry. Taiwanese-style oolong tea is often made from the same tea tree varieties, tea-making equipment, and tea-making methods as Taiwanese oolong tea, making it difficult to distinguish between them based on variety or appearance. To prevent the importation of Taiwanese-style oolong tea being fraudulently labeled as Taiwanese oolong tea, it is necessary to develop a geographical origin identification method to safeguard the brand value of Taiwanese oolong tea.

Since the 1980s, research reports on the geographical identification of the origin of agricultural products have been steadily increasing. Initially, the focus was on processed agricultural products such as wine, honey, tea, olive oil, and orange juice. Subsequent research focused on fresh agricultural products such as potatoes, onions, pistachios, and garlic, primarily due to the increasing

Received 1 April 2024; accepted 11 July 2024.
Available online 15 December 2024

* Corresponding author.
E-mail address: chihyi@tbrs.gov.tw (C.-Y. Hu).

https://doi.org/10.38212/2224-6614.3523

global trade in fresh agricultural products and legal requirements for labeling origin information [2]. The origin of tea will directly affect the quality, taste, and market price [3]. Tea leaves exhibit different aroma and flavor profiles from different origins [4]. Therefore, the ability to identify the origin of tea products helps ensure authenticity and traceability [5].

Techniques used for the geographical identification of the origin of tea include stable isotope fingerprinting, mineral element fingerprinting, and metabolite fingerprinting. These techniques involve the analysis of compounds with differences detected through instrumental analysis, such as characteristic metabolites, differing metabolites, mineral elements, and characteristic spectra, to establish discrimination models. These models undergo validation and blind sample confirmation to establish origin identification techniques. Commonly used chemometric methods include logistic regression method (LR) [6], linear discriminant analysis (LDA) [7,8], and partial least squares discriminant analysis (PLS-DA) [5]. Machine learning algorithms, including artificial neural networks (ANN) [9,10], k-nearest neighbor (k-NN) [11,12], random forest (RF) [13,14], and support vector machine (SVM) [11,15], have rapidly developed and are frequently used in food authentication. Compared to traditional chemometric methods, machine learning algorithms are less interpretable but do not rely on model assumptions and can handle a wide variety of data types. These approaches offer flexibility, capture complex relationships between predictors and the response variable, and achieve higher predictive accuracy, demonstrating superior sample prediction capabilities [13,16,17].

Currently, several research studies have applied multi-elements analysis to the identification of the origin of tea [13,18–23]. In Taiwan, Hu et al. (2009) utilized ICP-OES to analyze the elemental composition of Taiwanese tea and imported tea leaves, allowing for a rough differentiation of tea from specific regions. However, further development of more precise elemental analysis techniques is still needed to serve as direct evidence for the provenance authentication of Taiwanese tea [24]. Furthermore, Liu et al. (2015) utilized multi-element analysis by ICP-MS to establish a geographical origin discrimination model for Taiwanese high mountain oolong tea, achieving a high accuracy rate of 94.76% [21]. This became a prototype technology for the identification of the origin of Taiwanese oolong tea. However, international literature also points out that almost all literature on tea origin identification is only suitable for research purposes and has not been practically applied in food safety control and food industry analysis [5].

Therefore, the main objective of this study is to apply these techniques practically in the Taiwanese tea industry. To meet the scientific basis required for administrative inspections, it is necessary to further improve the accuracy of identification. The Tea and Beverage Research Station, Ministry of Agriculture, R.O.C. (Taiwan) (hereinafter referred to as TBRS) continues to enhance the origin identification technology of Taiwanese oolong tea. In addition to strengthening the collection of Taiwanese-style oolong tea samples from abroad to expand the identification database, multiple statistical models of different types are constructed for the identification of whether tea leaves are Taiwanese tea or external tea. This technology has been publicly recommended as the "Method of Test for Multi-elements in Tea" (TFDAF0032.00) by the Taiwan Food and Drug Administration of the Ministry of Health and Welfare (hereinafter referred to as Taiwan FDA), with an open date of November 5, 2021 [25].

This paper illustrates the development and practical application of the geographical origin identification for Taiwanese oolong tea as the foundation for the application in the Taiwanese tea industry (Fig. 1). Subsequently, due to the successful development of new technology, coupled with administrative inspections by government agencies, the fraudulent labeling of Taiwanese-style oolong tea as Taiwanese oolong tea during importation has been effectively prevented. This safeguards the brand value of Taiwanese oolong tea.

2. Materials and methods

2.1. Tea samples used for identification database

A total of 486 Taiwan oolong tea samples were collected from various geographical regions and representative tea-producing areas in Taiwan. Sampling areas included Chiayi County, Nantou County, Taichung City, Taitung County, Yilan County, Hualien County, Taoyuan City, and Miaoli County. Among these, Nantou County, Chiayi County, and Taichung City are the main producing areas for semi-ball-shaped and ball-shaped oolong teas, accounting for 72.1% of the total tea plantation area in Taiwan [26]. The main tea tree varieties are Chin-Shin-Oolong and TTES No. 12, along with small-leaf varieties such as Shy-Jih-Chuen and TTES No. 13. Additionally, Taiwanese-style oolong tea (non-Taiwanese tea) samples from overseas locations were collected, including China (55

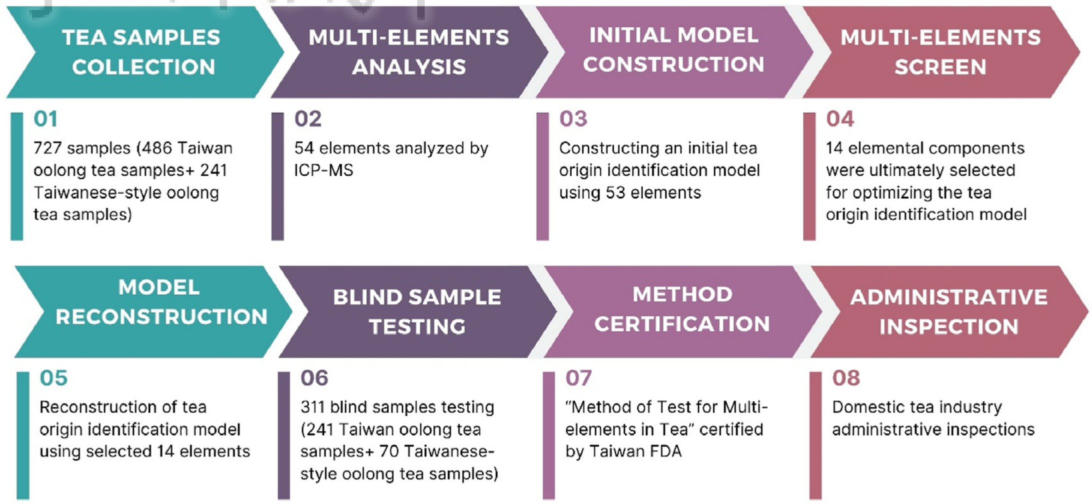

Fig. 1. The development and practical application of the geographical origin identification for Taiwanese oolong tea as the foundation for the application in the Taiwanese tea industry.

samples), Vietnam (114 samples), Indonesia (11 samples), and Thailand (61 samples), totaling 727 samples, as detailed in Table 1.

2.2. Multi-elements analysis

The tea samples were thoroughly mixed and spread evenly on plastic square trays. Approximately 6 g of tea leaves were taken using a quadrant method and ground into tea powder using a mixer mill (MM400 - Retsch, Germany). The particle size of tea powder was less than 0.5 mm. The tea powder was then transferred to a glass weighing bottle and placed in an oven at 85 °C for 15–25 h until dried, followed by cooling to room temperature inside a desiccator. Approximately 0.2 g of the dried tea powder was accurately weighed and placed in a Teflon microwave digestion vessel. Then, 6 mL of nitric acid (67–70% - Ultrapure Reagent, J T Baker) was added, and the mixture was allowed to stand for 4 h before digestion using a microwave digester (MARS 6 - CEM, USA). After cooling, the digested solution was transferred to a volumetric flask. The digestion vessel was washed with deionized water (the resistance ratio is greater than 18 MΩ·cm at 25 °C) in 5 mL portions, and the washings were combined in the volumetric flask and diluted to 50 mL with deionized water. The solution was then filtered through a membrane filter and used as the test solution for analysis by inductively coupled plasma mass spectrometry (ICP-MS) (7700x - Agilent, USA). The ICP-MS instrumental operating conditions were as follows: The flow of nebulizer gas was 1 L/min whereas the auxiliary and plasma gas flow were maintained at 0.9 L/min and 15 L/min respectively. The radio frequency power was set 1550 W. Collision gas (helium) flow rate was 4.3 mL/min. The standard solutions were prepared from multi-element calibration standard 2A (Ag, Al, As, Ba, Be, Ca, Cd, Co, Cr, Cs, Cu, Fe, Ga, K, Li, Mg, Mn, Na, Ni, Pb, Rb, Se, Sr, Tl, U, V, Zn), multi-element calibration standard 1 (Ce, Dy, Er, Eu, Gd, Ho, La, Lu, Nd, Pr, Sc, Sm, Tb, Th, Tm, Y, Yb) and multi-element calibration standard 3 (Sb, Au, Hf, Ir, Pd, Pt, Rh, Ru, Te, Sn) (Agilent Technologies, USA). Two certified reference materials (NIMJ CRM 7505-a tea leaf powder and INCT-TL-1 tea leaves) were used to validate the standard operating procedure.

Table 1. List of 727 tea samples used in this research.

Classification	Nationality	Sample size	Sample ratio
Taiwanese oolong tea (Taiwanese tea)	Taiwan (R.O.C.)	486	66.9%
Taiwanese-style oolong tea (non-Taiwanese tea)	China	55	33.1%
	Vietnam	114	
	Indonesia	11	
	Thailand	61	
Total	–	727	100%

2.3. Data analysis and model construction

The data obtained from the elemental analysis was processed using R software. Initially, the data underwent transformation analysis, followed by descriptive statistical analysis and multivariate statistical analysis. For the establishment of the geographical origin identification model, this study employed seven widely cited classification methods from the literature. These methods include traditional techniques: logistic regression (LR), linear discriminant analysis (LDA), and ridge regression, as well as machine learning algorithms: random forest (RF), boosting, support vector machine (SVM), and k-nearest neighbors (KNN). Traditional techniques provide interpretability and are computationally efficient, making them suitable for smaller datasets and simpler models. In contrast, machine learning algorithms can handle more complex and non-linear relationships in the data, potentially improving predictive accuracy and robustness. The comparison of these methods is crucial for identifying the origin of tea because each approach offers distinct advantages and addresses different aspects of classification problems. By evaluating both traditional and machine learning methods, we will gain a holistic understanding of the strengths and weaknesses of each approach, leading to more accurate prediction in determining the provenance of tea. For each classification method, the optimal solutions for the relevant adjustment parameters and the estimates of statistical evaluation indices were obtained through k-fold cross-validation.

2.4. Blind samples testing

Using a new sample set consisting of 311 tea samples (comprising 241 Taiwanese oolong tea samples and 70 Taiwanese-style oolong tea samples), known for their places of origin, applied to a statistical model constructed using 14 elements, for blind sample testing purposes.

3. Results

3.1. Multi-elements analysis and an initial model construction

All tea samples underwent analysis via ICP-MS, resulting in data for 54 different elements. The element “Ruthenium (Ru)" was excluded from the analysis as all measurements were zero. Additionally, considering the significant right skewness observed in the detection values of various elements, a natural logarithmic transformation was applied to the elemental data. For elements with original detection values of zero, a small positive value (one-tenth of the smallest non-zero value for that element) was added to these values before the data transformation. This transformation substantially corrected the right skewness in the original data and resulted in a more symmetric distribution of detection values for each element. Furthermore, pairwise correlation coefficients were calculated for all analyzed elements, revealing that the majority of elements showed no significant correlation.

The transformed analysis data of the 53 elements were utilized in seven classification methods, including traditional methods like logistic regression (LR), linear discriminant analysis (LDA), and ridge regression, as well as machine learning algorithms such as random forest (RF), boosting, support vector machine (SVM), and k-nearest neighbors (KNN). The parameters for each classification method were adjusted, and their performance was evaluated using 5-fold cross-validation to assess the effectiveness of each method.

Based on the classification analysis distinguishing between Taiwanese and non-Taiwanese (external) tea origins, seven models were compared regarding their performance in distinguishing Taiwanese and non-Taiwanese (external) teas, and the results are summarized in Table 2. In terms of accuracy rate, LDA, Ridge, Random Forest, Boosting, and SVM were almost consistent, achieving discrimination accuracy rates of 97.5%–98.8%, while LR and KNN performed less favorably, with rates of 92.9% and 84.3%, respectively. Regarding the recall rates for Taiwanese tea origins, LDA, Ridge, Random Forest, Boosting, and SVM were almost consistent, achieving recall rates of nearly 98.0%–99.0%, while LR and KNN showed poorer performance, with rates of 95.1% and 93.8%, respectively. Precision values for predicting Taiwanese origin was almost consistent among LDA, Ridge, Random Forest, Boosting, and SVM, reaching approximately 98.0%–98.4%, while LR and KNN showed significantly poorer performance, with rates of 94.1% and 83.8%, respectively. Based on the performance of each classifier in various evaluation indices, LDA, Ridge, Random Forest, Boosting, and SVM were preliminarily selected as suitable statistical models for Taiwanese oolong tea origin identification.

Table 2. The statistical metrics of seven classification models using 53 elements.

Method	LDA	LR	Ridge	Random Forest	Boosting	SVM	KNN
Accuracy rate	0.9800	0.9292	0.9800	0.9800	0.9800	0.9746	0.8425
Recall Rate	0.9877	0.9507	0.9897	0.9877	0.9856	0.9795	0.9384
Precision Value	0.9816	0.9411	0.9797	0.9816	0.9836	0.9815	0.8385

3.2. Multi-elements screen and models reconstruction

To mitigate the potential interference of numerous elements and fertilization effects on the elemental analysis, and considering detection limits and the practical cost of analysis in the industry application, 14 elemental components were ultimately selected using the Lasso regression method for optimizing the tea origin identification model. These elements include lithium (Li), vanadium (V), chromium (Cr), nickel (Ni), copper (Cu), zinc (Zn), rubidium (Rb), strontium (Sr), cadmium (Cd), cesium (Cs), barium (Ba), lanthanum (La), cerium (Ce), and lead (Pb). Among these 14 elements, rare earth elements such as La and Ce can be used to identify the geographical origin of agricultural products [27], while other elements have also been mentioned in previous studies [22,23].

Based on the selected 14 elements, we will reconstruct the tea origin identification model using the following five statistical methods with overall superior performance: traditional linear discriminant analysis (LDA) and ridge regression, as well as machine learning algorithms including random forest (RF), boosting, and support vector machine (SVM). Each discrimination model will be evaluated using 5-fold cross-validation under a discrimination threshold (defaulted to 0.5, where any discrimination result greater than 0.5 is classified as originating from Taiwan). Various metrics, including accuracy rate, true-positive rate (recall rate for Taiwan as the origin) and precision (positive predicted value for predicting Taiwan as the origin), would be assessed for each model, as detailed in Table 3. For the geographical identification analysis of Taiwanese oolong tea, when distinguishing between Taiwanese tea and tea from overseas (non-Taiwanese), the performances of LDA, Ridge, Random Forest, Boosting, and SVM are nearly identical. They achieve an accuracy rate ranging from 97.1% to 97.8%, a recall rate (true positive rate) for Taiwan as the origin ranging from 98.4% to 99.0%, and a precision value for predicting Taiwan as the origin ranging from 97.3% to 97.8%.

Table 3. The statistical metrics of five classification models using 14 elements.

Method	LDA	Ridge	Random Forest	Boosting	SVM
Accuracy rate	0.978	0.977	0.976	0.971	0.977
Recall Rate	0.989	0.99	0.988	0.984	0.988
Precision Value	0.977	0.975	0.976	0.973	0.978

3.3. Blind sample testing

The tea origin identification model constructed using the 14 known elements was validated with blind samples of known origin (Table 4). There were 241 blind samples known to be from Taiwan, out of which the model predicted 237 correctly, resulting in an accuracy rate of 98.3%. There were 70 blind samples known to be from overseas, out of which the model predicted 69 correctly, yielding an accuracy rate of 98.6%. In total, there were 311 samples, resulting in an overall accuracy rate of 98.4%.

4. Discussion

The compositional characteristics of agricultural products, such as proteins, carbohydrates, fats, mineral elements, etc., are influenced by various factors including variety, human cultivation practices, manufacturing processes, seasonal variations, and geographical factors such as soil composition and altitude. Apart from anthropogenic influences, differences arising from various natural geographical environments serve as important evidence of geographical differentiation, thus becoming a crucial basis for the geographical identification of the origin of agricultural products. Furthermore, compared to other components in tea leaves, the concentration of mineral elements in tea leaves tends to be more stable and less susceptible to changes due to processing steps or storage time [28].

In the application of ICP-MS analysis for fingerprinting mineral elements in tea leaves, several studies have demonstrated the ability to distinguish the geographical origin from different countries as well as within different regions of the same country. Pilgrim et al. (2010) utilized ICP-MS to analyze the isotopic or elemental content of tea leaves to differentiate between Sri Lanka, China, India, and Taiwan. Different discrimination functions were

Table 4. Results of blind sample testing.

Geographical origin	Blind samples	Model prediction		Prediction correction	Accuracy rate
		Taiwanese	Non-Taiwanese		
Taiwanese	241	237	4	237	98.3
Non-Taiwanese	70	1	69	69	98.3
Total	311			306	98.4

used, achieving an accuracy rate of 97.6% [22]. Zhang et al. (2018) utilized ICP-MS to analyze Fenghuangdancong tea (a kind of oolong tea) and other Chinese teas. They achieved a similarity classification rate of 93.3% and a cross-validated prediction ability of 96.6% through fingerprinting analysis of 45 elements and linear discriminant analysis (LDA). This technology can be used to authenticate Fenghuangdancong tea [20]. Zhang et al. (2020) employed ICP-MS to analyze 71 green tea samples from five adjacent regions in Guizhou Province, China. They found significant differences in 39 elements among different regions. Through stepwise linear discriminant analysis (S-LDA) with cross-validation, the predictive ability reached 88.3% [18]. Liu et al. (2020) analyzed Yongchuanxiuya tea (a kind of green tea) from six different producing areas using ICP-MS and ICP-OES. They employed HCA, PCA, PLS-DA, BP-ANN, and LDA models for origin identification, with blind sample verification. The results showed that the prediction accuracy of both BP-ANN and LDA methods exceeded 90%, with LDA demonstrating the best performance, achieving an accuracy rate of 100% [19]. Deng et al. (2020) utilized a random forest (RF) model to establish origin identification technology for Chinese green tea producing areas, achieving a prediction accuracy rate exceeding 91%. They considered RF to have higher accuracy compared to other statistical models [13].

Compared to other research reports on tea origin identification, our study collected 727 samples of Taiwanese oolong tea from different geographical regions and samples of Taiwanese-style oolong tea from overseas (486 from Taiwan and 241 from other countries). We utilized ICP-MS to analyze 54 elements in the tea samples and selected 14 key elemental components, including lithium (Li), vanadium (V), chromium (Cr), nickel (Ni), copper (Cu), zinc (Zn), rubidium (Rb), strontium (Sr), cadmium (Cd), cesium (Cs), barium (Ba), lanthanum (La), cerium (Ce), and lead (Pb), to optimize the origin identification model. Previous studies have mentioned rare earth elements (REEs), which include 15 lanthanides (La) elements, as well as Yttrium (Y) and Scandium (Sc), totaling 17 elements. These elements have been considered suitable for application in tea origin identification [27]. Among the 14 discriminatory elements selected in this study, only 2 (La and Ce) belong to rare earth elements (REEs). This might be attributed to the fact that the oolong tea samples collected for this study may not have been exclusively from a single tea season, as the content of most rare earth elements (REEs) in tea leaves is mainly influenced by the season of harvest [27]. Furthermore, another study has indicated that 8 elements including Mg, Ni, Rb, Sr, Cd, Cs, Ba, and Pb show the highest correlation with tea origins [28]. Except for Mg, which is recognized as being influenced by fertilization, the other 7 elements are within the scope of the 14 discriminatory elements selected in this study.

Through five common classification statistical methods, including two traditional statistical classification methods and three machine learning algorithms, each method had its strengths and weaknesses. Generally, traditional classification methods perform better when there is fewer data and different categories can be linearly separated in the feature space. However, when there is a sufficient amount of data and different categories cannot be linearly separated in the feature space, machine learning algorithms have the advantage in classification methods. In our study, considering the predictive accuracy of the model methods, traditional classification methods generally performed better when distinguishing between Taiwanese and non-Taiwanese origins. LDA, Ridge, Random Forest, Boosting, and SVM achieved almost identical accuracy rates ranging from 97.1% to 97.8%. The recall rates for Taiwanese origin ranged from 98.4% to 99.0%, and the precision values for predicting Taiwanese origin ranged from 97.3% to 97.8%. The database established in this study comprises over seven hundred samples, which is rare. Increasing the number of samples indeed improves the accuracy of identification. Moreover, the tea samples identified belong to the same category of oolong tea, with similar tea tree varieties, tea-making equipment, and tea-making techniques. They differ only in the producing country. Through elemental analysis combined with traditional chemometric methods and machine learning algorithms, our study has demonstrated good effectiveness in distinguishing between Taiwanese and overseas teas. Furthermore, it has been applied practically in the tea industry to uphold the rights and interests of tea farmers and consumers.

Comparing the performance of the models originally using 53 elements to the simplified version using 14 elements, there is little difference in the three metrics of accuracy rate, recall rate for Taiwan as the origin, and precision value for predicting Taiwan as the origin (Table 5). Therefore, considering the cost-saving aspect, the simplified analysis model using 14 elements can be adopted as the geographical origin identification model for Taiwanese oolong tea.

Table 5. The statistical metrics of five classification models between 53 and 14 elements.

Method	LDA	Ridge	Random Forest	Boosting	SVM
Accuracy rate	0.980#/0.978*	0.980#/0.977*	0.980#/0.976*	0.980#/0.971*	0.975#/0.977*
Recall Rate	0.988#/0.989*	0.990#/0.990*	0.988#/0.988*	0.986#/0.984*	0.980#/0.988*
Precision Value	0.982#/0.977*	0.980#/0.975*	0.982#/0.976*	0.984#/0.973*	0.982#/0.978*

Note: # and * respectively represent the statistical values of 53 elements and 14 elements in each model.

5. Conclusion

Integrated analysis shows that trace elements in tea leaves do indeed achieve significant effectiveness in geographical origin identification applications. This technology has been publicly endorsed by the Taiwan FDA as the recommended inspection method 'Method of Test for Multi-elements in Tea' (TFDAF0032.00), released on November 5, 2021 [25]. Furthermore, it has been actively utilized in the geographical origin identification inspections conducted by other official agencies in Taiwan and serves as supporting evidence for investigation cases. These cases include projects such as geographical origin inspections of tea products in the market, which involve participation from the Office and Food Safety, Executive Yuan, the Taiwan FDA, and the Agriculture and Food Agency, Ministry of Agriculture (hereinafter referred to as AFA) [29], as well as geographical origin identification projects of local specialty tea evaluation competitions involving participation from the AFA [30], and extended investigation projects led by the Investigation Bureau, Ministry of Justice [31]. In addition, the policy-making authority for Taiwan's domestic tea industry, the AFA, has enacted regulations according to the 'Agricultural Production and Certification Act'. Starting from January 1, 2023, domestically produced tea must provide one of the following: agricultural and food products traceability barcode (QR Code), Traceable Agricultural Products (TAP), or CAS Organic agricultural products (CAS Organic), aiming to protect the rights and interests of consumers and tea farmers [32]. Finally, the impact of blending tea from different origins on overall origin identification will also be further explored in subsequent research. Additionally, the Taiwanese tea industry includes other important specialty teas such as Oriental Beauty tea, Wenshan Paochong tea, and Red oolong tea. In the future, we will develop other geographical origin identification techniques for these specialty teas and continuously improve inspection methods to uphold the brand value of Taiwanese tea.

Authors' contributions

The prototype technique and tea sample collection were carried out by T.-L. Liu, C.-F. Chiu and J.-C. Lin. C.-Y. Hu, H.-T. Tsai and T.-C. Su participated in the research design. Multi-elements analysis was conducted by J.-R. Dai. The statistical models were constructed by C.-H. Chiu. And C.-Y. Hu, C.-H. Chiu. and J.-R. Dai. wrote the manuscript.

Declaration of competing interest

The authors declare that they have no conflict of interest.

Acknowledgements

We would like to thank the Ministry of Agriculture for funding this study and the Taiwan FDA for certifying this method (TFDAF0032.00).

References

[1] Su TC, Yang MJ, Huang HH, Kuo CC, Chen LY. Using sensory wheels to characterize consumers' perception for authentication of Taiwan specialty teas. Foods 2021;10:836.
[2] Luykx D, Van Ruth SM. An overview of analytical methods for determining the geographical origin of food products. Food Chem 2008;107:897−911.
[3] Li YJ, Li C, Li JH, Shi W, Li Z, Yao YF, et al. Identification of Pu'er tea origin based on near infrared spectroscopy. Hubei Agric Sci 2020;59:138−41 [In Chinese, English abstract].
[4] Seow WJ, Low DY, Pan WC, Gunther SH, Sim X, Torta F, et al. Coffee, black tea, and green tea consumption in relation to plasma metabolites in an asian population. Mol Nutr Food Res 2020;67:2000527.
[5] Shuai MY, Peng CY, Niu HL, Shao DL, Hou RY, Cai HM. Recent techniques for the authentication of the geographical origin of tea leaves from *Camellia sinensis*: a review. Food Chem 2022;374:131713.
[6] Gumus O, Yasar E, Gumus ZP, Ertas H. Comparison of different classification algorithms to identify geographic origins of olive oils. J Food Sci Technol 2020;57:1535−43.
[7] Ma G, Zhang Y, Zhang J, Wang G, Chen L, Zhang M, et al. Determining the geographical origin of Chinese green tea by linear discriminant analysis of trace metals and rare earth elements: taking Dongting Biluochun as an example. Food Control 2016;59:714−20.
[8] D'Archivio AA, Giannitto A, Maggi MA, Ruggieri F. Geographical classification of Italian saffron (*Crocus sativus* L.) based on chemical constituents determined by high-performance liquid-chromatography and by using linear discriminant analysis. Food Chem 2016;212:110−6.
[9] Liu Y, Yao LY, Xia ZZ, Gao YG, Gong ZY. Geographical discrimination and adulteration analysis for edible oils using two-dimensional correlation spectroscopy and convolutional neural networks (CNNs). Spectrochim Acta Mol Biomol Spectrosc 2021;246:118973.
[10] Sun LX, Danzer K, Thiel G. Classification of wine samples by means of artificial neural networks and discrimination analytical methods. Fresenius' J Anal Chem 1997;359:143−9.

[11] Kabir MH, Guindo ML, Chen R, Liu F. Geographic origin discrimination of millet using vis-NIR spectroscopy combined with machine learning techniques. Foods 2021;10:2767.
[12] Wang F, Zhao HY, Yu CD, Tang J, Wu W, Yang QL. Determination of the geographical origin of maize (*Zea mays* L.) using mineral element fingerprints. J Sci Food Agric 2020; 100:1294–300.
[13] Deng XF, Liu Z, Zhan Y, Ni K, Zhang YZ, Ma WZ, et al. Predictive geographical authentication of green tea with protected designation of origin using a random forest model. Food Control 2020;107:106807.
[14] Maione C, Batista BL, Campiglia AD, Barbosa F, Barbosa RM. Classification of geographic origin of rice by data mining and inductively coupled plasma mass spectrometry. Comput Electron Agric 2016;121:101–7.
[15] Bona E, Marquetti I, Link JV, Makimori GYF, da Costa Arca V, Guimarães Lemes AL, et al. Support vector machines in tandem with infrared spectroscopy for geographical classification of green arabica coffee. LWT–Food Sci Technol 2017;76:330–6.
[16] Peng CY, Zhang YL, Song W, Cai HM, Wang Y, Granato D. Characterization of Brazilian coffee based on isotope ratio mass spectrometry (δ13C, δ18O, δ2H, and δ15N) and supervised chemometrics. Food Chem 2019;297:124963.
[17] Gromski PS, Correa E, Vaughan AA, Wedge DC, Turner ML, Goodacre R. A comparison of different chemometrics approaches for the robust classification of electronic nose data. Anal Bioanal Chem 2014;406:7581–90.
[18] Zhang M, Huang C, Zhang J, Qin H, Ma G, Liu X, et al. Accurate discrimination of tea from multiple geographical regions by combining multi-elements with multivariate statistical analysis. J Food Meas Char 2020;14:3361–70.
[19] Liu HL, Zeng YT, Zhao X, Tong HR. Improved geographical origin discrimination for tea using ICP-MS and ICP-OES techniques in combination with chemometric approach. J Sci Food Agric 2020;100:3507–16.
[20] Zhang X, Wu H, Huang X, Zhang C. Establishment of element fingerprints and application to geographical origin identification of Chinese Fenghuangdancong tea by ICP-MS. Food Sci Technol Res 2018;24:599–608.
[21] Liu TL, Chiou CF, Lin JC, Chen KR, Huang CT, Lin RH, et al. Determination the geographic origin of tea. 2015 international forum on tea culture, creativity & science. Taiwan; Nantou 2015:97–9 [In Chinese, English abstract].
[22] Pilgrim TS, Watling RJ, Grice K. Application of trace element and stable isotope signatures to determine the provenance of tea (*Camellia sinensis*) samples. Food Chem 2010;118:921–6.
[23] Marcos A, Fisher A, Rea G, Hill SJ. Preliminary study using trace element concentrations and a chemometrics approach to determine the geographical origin of tea. J Anal At Spectrom 1998;13:521–5.
[24] Hu CY, Kuo KL, Tsai YJ, Tsai JS. The feasibility of element analysis technology for certificating the production origin of Taiwan Tea. Taiwan Tea Res Bull 2009;28:63–74 [In Chinese, English abstract].
[25] Taiwan Food and Drug Administration. Method of test for multi-elements in tea (TFDAF0032.00) (November 5, 2021). Available at: https://www.fda.gov.tw/TC/siteList.aspx?sid=1574&scid=722. [Accessed 8 April 2024].
[26] Agriculture and Food Agency. Special crops productions in 2022. In: agriculture and food agency. In: Nantou: agriculture and food agency; 2023. p. 40.
[27] Zhao H, Yang Q. The suitability of rare earth elements for geographical traceability of tea leaves. J Sci Food Agric 2019; 99:6509–14.
[28] Zhao H, Yu C, Li M. Effects of geographical origin, variety, season and their interactions on minerals in tea for traceability. J Food Compos Anal 2017;63:15–20.
[29] Office and Food Safety, Executive Yuan. A new milestone for Taiwanese tea - scientifically distinguishing Taiwanese tea from overseas tea. Available at: https://www.ey.gov.tw/ofs/E2E3786536626ADC. [Accessed 8 April 2024].
[30] Agriculture and Food Agency. Enhancing the authentication of origin for premium tea competitions to uphold public rights (October 7, 2022). Available at: https://www.afa.gov.tw/cht/index.php?code=list&flag=detail&ids=307&article_id=25124. [Accessed 8 April 2024].
[31] Investigation Bureau, Ministry of Justice. The manager of an online e-commerce platform, accused of substituting Vietnamese tea for Taiwanese high-mountain tea, has been detained and denied visitation rights (October 7, 2022). Available at: https://www.mjib.gov.tw/news/Details/1/793. [Accessed 8 April 2024].
[32] Agriculture and Food Agency. Domestic tea produced on New Year's Day next year should provide traceability labels (October 7, 2022). Available at: https://www.afa.gov.tw/cht/index.php?code=list&flag=detail&ids=307&article_id=25138. [Accessed 8 April 2024].