Current Research in Food Science 8 (2024) 100738

Contents lists available at ScienceDirect

Current Research in Food Science

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

Explainable machine learning for predicting the geographical origin of Chinese Oysters via mineral elements analysis

Xuming Kang [a], Yanfang Zhao [a], Lin Yao [a], Zhijun Tan [a,b,c,*]

[a] Key Laboratory of Testing and Evaluation for Aquatic Product Safety and Quality, Ministry of Agriculture and Rural Affairs, Yellow Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences, Qingdao, 266071, China
[b] Pilot National Laboratory for Marine Science and Technology (Qingdao), Qingdao, 266071, China
[c] Collaborative Innovation Center of Seafood Deep Processing, Dalian Polytechnic University, Dalian, 116034, China

ARTICLE INFO

Handling Editor: Maria Corradini

Keywords:
Oyster
Elemental profile
Explainable machine learning
SHAP
Geographical origin

ABSTRACT

The traceability of geographic origin is essential for guaranteeing the quality, safety, and protection of oyster brands. However, the current outcomes of traceability lack credibility as they do not adequately explain the model's predictions. Consequently, we conducted a study to evaluate the efficacy of utilizing explainable machine learning combined with mineral elements analysis. The study findings revealed that 18 elements have the ability to determine regional orientation. Simultaneously, individuals should pay closer attention to the potential risks associated with oyster consumption due to the regional differences in essential and toxic elements they contain. Light gradient boosting machine (LightGBM) model exhibited indistinguishable performance, achieving flawless accuracy, precision, recall, F1 score and AUC, with values of 96.77%, 96.43%, 98.53%, 97.32% and 0.998, respectively. The SHapley Additive exPlanations (SHAP) method was used to evaluate the output of the LightGBM model, revealing differences in feature interactions among oysters from different provinces. Specifically, the features Na, Zn, V, Mg, and K were found to have a significant impact on the predictive process of the model. Consistent with existing research, the use of explainable machine learning techniques can provide insights into the complex connections between important product attributes and relevant geographical information.

1. Introduction

Oysters were recognized worldwide for their popularity as bivalves, primarily due to their substantial amounts of protein, amino acids, eicosapentaenoic acid (EPA, 20:5n-3), docosahexaenoic acid (DHA, 22:5n-3), and mineral elements (Liu et al., 2022; Zhao et al., 2022; Loaiza et al., 2023). In 2022, China's oyster production is projected to exceed 6.20 million tons, primarily concentrated in the provinces of Liaoning, Fujian, Shandong, Guangdong and Guangxi (MOAC, 2023). In recent years, an outbreak of dangerous algal blooms, specifically harmful microalgae, has led to the detection of various toxins, such as dinophysistoxin-2 and domoic acid, in oysters sourced from Guangdong and Guangxi provinces (Zheng et al., 2022). Moreover, oysters possess the ability to accumulate toxic elements such as cadmium in their tissues (Ng et al., 2010). Therefore, the market may be susceptible to the entry of contaminated items as a result of labeling fraud and illegal seafood trading (Sumaila et al., 2020; Matos et al., 2021). In such circumstances, human concerns regarding the quality and safety of oysters lead to a higher tendency for buyers to choose branded oysters from reputable or renowned production regions (Guo et al., 2023). Moreover, following the enforcement of the China-EU agreement on geographical indications, various oyster products with designated geographic origins (e.g., Beihai Oyster) have been incorporated into the catalog of mutually recognized items. Consequently, the significance of advancing oyster origin traceability technologies that prioritize product quality, safety, and trademark protection has been underscored.

The Fisheries and Fisheries Administration Bureau, a division of the Ministry of Agriculture and Rural Affairs in China, is primarily responsible for overseeing the quality and safety of oyster products (Zhai et al., 2020). Furthermore, to ensure quality and safety of oysters and maintain consumer trust, several approaches have been adopted in China to identify their origin. These include non-targeted lipidomics analysis (Liu

* Corresponding author. Key Laboratory of Testing and Evaluation for Aquatic Product Safety and Quality, Ministry of Agriculture and Rural Affairs, Yellow Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences, Qingdao 266071, China.
E-mail address: tanzj@ysfri.ac.cn (Z. Tan).

https://doi.org/10.1016/j.crfs.2024.100738
Received 11 January 2024; Received in revised form 6 April 2024; Accepted 12 April 2024
Available online 16 April 2024

et al., 2022), attenuated total reflectance Fourier-transform infrared spectroscopy (Guo et al., 2023), shellomics and microstructural analysis (Zhang et al., 2023). Additional methods, such as pyrolysis mass spectrometry fingerprinting (Ratel et al., 2008), mineral elemental fingerprinting (Mouchi et al., 2021), and the combination of compound-specific isotope and elemental analysis (Matos et al., 2021), have proven advantageous in identifying the origin of oysters. Among these technologies, elemental fingerprinting stands out as a fast, cost-effective, and straightforward method that relies on the understanding that the elemental composition of the maricultural environment can be accurately represented in the organisms themselves (Kang et al., 2021). Elemental fingerprinting has proven successful in determining the origin of Asian seabass, Black tiger prawns, sea cucumber and scallops (Honig et al., 2020; Han et al., 2022a; Kang et al., 2022). However, the suitability of this method for tracing the origin of Chinese oysters is questionable.

Machine learning has emerged as a valuable tool for investigating food security, food fraud and food origin traceability research, even when traditional face-to-face survey data is lacking (Jiménez-Carvelo et al., 2019; Deng et al., 2021; Kang et al., 2022; Martini et al., 2022; Huang et al., 2023). Machine learning has proved to be effective in various fields, including economic, biology, manufacturing, and health promotion research, primarily due to its strong classification performance after alleviating the influence of overfitting by data reduction, feature selection, cross-validation and other measures (Lundberg and Lee, 2017; Lundberg et al., 2020; Kim and Kim, 2022). However, understanding the inner workings of a machine learning model can be challenging due to its black-box nature, which can hinder human comprehension of the underlying reasons behind a specific prediction (Ribeiro et al., 2016; Ekanayake et al., 2022). SHapley Additive exPlanations (SHAP) has emerged as a novel framework that efficiently explains the output of a machine learning model (Lundberg and Lee, 2017). This method evaluates the average marginal contribution of a feature using SHAP values, thereby enhancing our understanding of predictions in specific scenarios. Moreover, it offers valuable insights into both individual instances and overall patterns at a global and local scale (Lundberg and Lee, 2017; 2019). The interpretation of SHAP values has played pivotal roles in establishing trust in auxiliary diagnosis, understanding spatial phenomena, enhancing strength predictions, and facilitating environmental management (Goodwin et al., 2022; Kim and Kim, 2022; Li, 2022; Ekanayake et al., 2022; Wang et al., 2022). To the best of our knowledge, however, there has been no investigation of this technique in the identification of oyster origin. By addressing this gap, we can gain users' trust, enhance our understanding of the impact of input features on predictions, and improve the accuracy of models in real-word scenarios.

This study establishes the multi-elemental profiles, comprising essential and toxic elements, in oysters sourced from the primary farming regions in China to evaluate the quality and safety of oysters across various regions. Further, the random forest (RF), traditional gradient boosting decision tree (GBDT), Light gradient boosting machine (LightGBM) and extreme gradient boosting (XGBoost) models were developed to evaluate the feasibility of utilizing mineral elements and explainable machine learning in oyster origin traceability. Subsequently, the model with the highest performance was further examined using SHAP at both global and local levels. The objective of this investigation was to identify a highly effective solution for tracing the origin of oysters by leveraging mineral elements and explainable machine learning to enhance the regulation of oyster quality and safety. Additionally, this study highlights the impact of crucial features on the model's accuracy in predicting the origin of oysters.

2. Materials and methods

2.1. Sampling

The same species of oyster (*Crassostrea gigas*) samples were collected from 94 sampling stations located in China's primary production regions, as illustrated in Fig. 1. Specifically, samples from Guangdong (n = 51) and Guangxi provinces (n = 21) were collected in 2020, while samples from Liaoning (n = 8) and Shandong provinces (n = 14) were obtained in 2021. At each sampling site, 30 individual fresh oysters with size 10–15 cm and weight 150–200 g were carefully gathered, placed in food-grade plastic bags, and refrigerated at a temperature of 4 °C for a maximum of 48 h before transportation to the laboratory (Kang et al., 2021).

2.2. Sample preparation and analysis

To eliminate any residual foreign materials, the soft tissues within the shells were excised using a plastic knife and thoroughly rinsed with distilled water. Subsequently, the soft tissues were refrigerated (−20 °C) for 12 h prior to being lyophilized at −50 °C for a duration of 4 days. The resultant dried soft tissues were then finely powdered. For the examination of mineral elements, 0.5 g oyster samples were carefully digested using a combination of concentrated nitric acid and perchloric acid on a heated plate. The resulting digest was adjusted to a known volume of 100 mL with ultrapure water.

The concentrations of Mn, Fe, Cu, Zn, Sr, As, Cd, Se, Li, Al, V, Co, Ni, Ag, Ba, Pb, Bi, Ca, and Mg were measured using inductively coupled plasma mass spectrometry (ICP-MS, ELAN DRCII, PerkinElmer Instruments, USA). Additionally, potassium (K) and sodium (Na) were determined using inductively coupled plasma optical emission spectroscopy (ICP-OES, Optima 5300DV, PerkinElmer Instruments, USA). The ICP-MS and ICP-OES operation parameters proposed by Kang et al. (2018) were adopted for the analysis. The external standard approach was employed for quantification, with determination coefficients of the standard curves exceeding 0.995. Accuracy and precision of the analytical procedure were assessed using a scallop standard reference material (SRM) (GBW10024). The SRM exhibited a similar matrix composition to the analyzed samples, and the observed concentrations closely aligned with the certified values, resulting in recoveries ranging from 92.3% to 97.9% with a relative standard deviation (RSD) below 10% (n = 5).

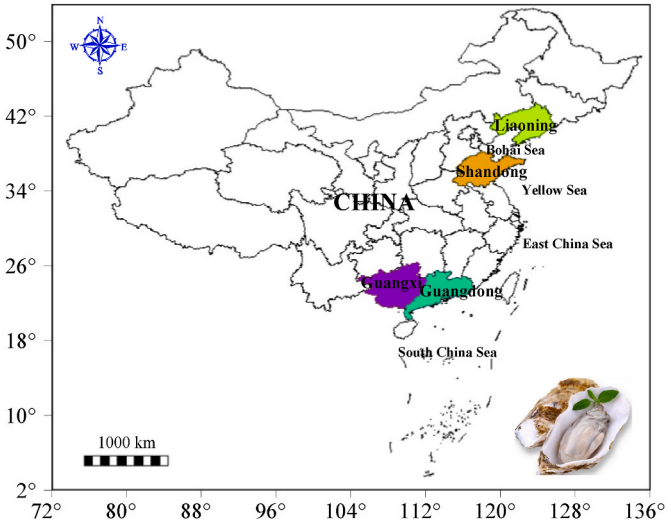

Fig. 1. Spatial distribution of oyster samples collected from four provinces in China.

2.3. Statistical analysis

Model generation and examination was implemented in Python 3.8 and Pycharm Community Edition 2021.3.1, with imbalanced-learn 0.9.0 (Lemaître et al., 2017), XGBoost 1.5.2 (Chen and Guestrin, 2016), scikit-learn 0.24.2 (Pedregosa et al., 2011), lightGBM 2.2.3 (Ke et al., 2017), and SHAP 0.41.0 libraries (Lundberg et al., 2020).

2.3.1. Machine learning models

The Random Forest model is a machine learning technique that entails the creation of numerous decision trees through random selection of features and samples. Each tree generates a prediction, and the final output is determined by aggregating these predictions (Breiman, 2001). The GBDT method is an interactive algorithm for decision tree modeling. It combines the outcomes of multiple trees that are trained independently, with each subsequent tree contributing to the correction of errors made by the preceding tree (Friedman, 2001). LightGBM is an optimized GBDT algorithm that constructs each tree in a leaf-wise manner, maximizing the fit at every step (Ke et al., 2017; Wang et al., 2022). XGBoost is a gradient boosting integration technique that includes a regularization factor in the objective function to manage the complexity of the trees. This procedure helps in mitigating the problem of overfitting in the model (Chen and Guestrin, 2016; Ekanayake et al., 2022). All the aforementioned models excel in classification tasks will be utilized as a quality control strategy by predicting the geographical origin of oysters from unknown sources.

2.3.2. Model construction and evaluation

The process of constructing and evaluating the model was illustrated in Fig. 2. Firstly, the mineral elements (Mn, Fe, Zn, Se, As, Cd, Li, Al, V, Co, Sr, Ag, Ba, Bi, Ca, K, Mg, and Na), which exhibit significant regional variations, were compiled into a data matrix consisting of 94 samples. This matrix was structured with 94 rows representing cases and 18 columns representing features. The desired dataset was a 94 × 1 matrix with a column labeled "province label" denoting the provinces Liaoning, Shandong, Guangdong, and Guangxi, assigned the labels 0, 1, 2, and 3 respectively.

Secondly, the dataset ($n = 94$) was randomly split into training and testing sets in a 0.7:0.3 ratio prior any data preprocessing to prevent data leakage during model development (Kapoor and Narayanan, 2023). Subsequently, considering the compositions of elements in oysters exhibit varying magnitudes (in g/kg or mg/kg). In addition, the results revealed an imbalanced pattern resulting from differences in actual oyster mariculture farms across provinces. The negative impact on the performance of machine learning models is inevitable (Fernández et al., 2018). To mitigate adverse effects and prevent data leakage, we selected the *Standard Scaler*, applied the *fit()* function to the training set, and subsequently used the *transform()* function on both the training and testing sets to normalize each dataset. Moreover, the synthetic minority oversampling technique (SMOTE) was exclusively applied to the training set (Kang et al., 2022).

Third, the preprocessed training set ($n = 140$) was utilized for training the models. All models underwent hyper-parameter optimization using a grid search strategy with 10-fold cross-validation on the training set. Learning curves were calculated for each hyperparameter-optimized classifier on the datasets using a 10-fold cross-validation to track the model fitting status with an increasing number of training samples.

Finally, statistical metrics were employed to assess the performance of the models using the test set ($n = 31$). This study adopted various statistical measurements including accuracy, precision, recall, and F1 score, which is defined as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{2}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{3}$$

$$\text{F1 score} = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{4}$$

where TP, FP, TN, and FN represent true positives, false positives, true negatives, and false negatives, respectively. Additionally, AUC refers to the area under the curve of the receiver operating characteristic

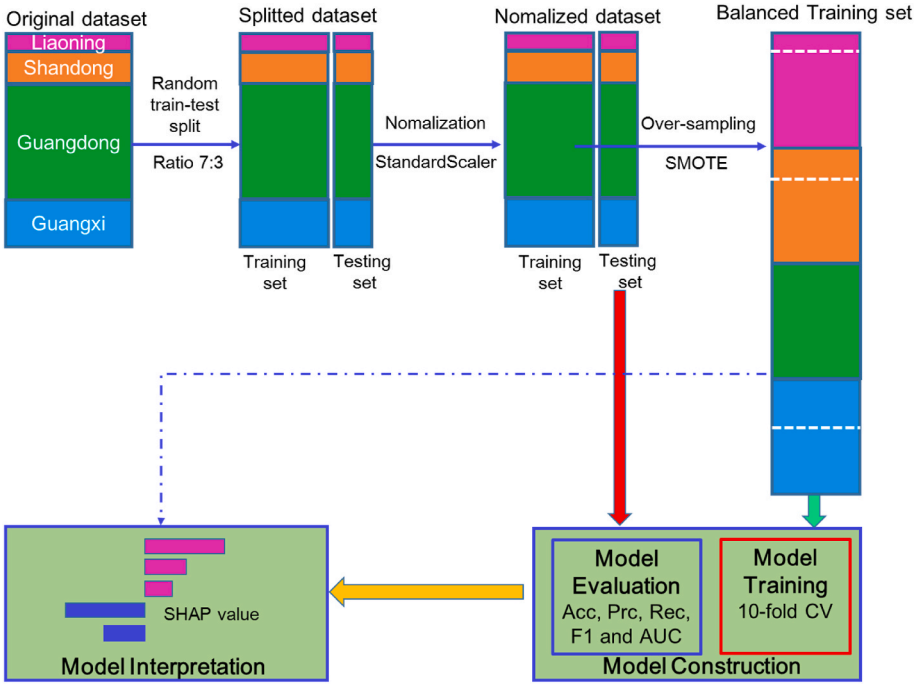

Fig. 2. Flowchart of model construction, performance evaluation and interpretation (the area of dataset rectangles represent sample size only).

(ROC), constructed using FP and TP (Wang et al., 2020).

2.3.3. Shapley additive explanations (SHAP)

The RF model was selected based on its performance and its interpretation was facilitated by the SHAP framework. TreeExplainer, an innovative technique for model explanation rooted in game-theoretic Shapley values, was utilized, demonstrating its effectiveness in enhancing interpretability for tree-based models (Lundberg and Lee, 2017; 2019, 2020). The SHAP algorithm assessed the contribution of each variable to the model's output using both the model and input dataset, as elaborated below (Lundberg and Lee, 2017):

$$\varphi_i = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|!(|F| - |S| - 1)!}{|F|!} \left[f_{S \cup \{i\}}\left(x_{S \cup \{i\}}\right) - f_S(x_S) \right] \quad (5)$$

where φ_i, F and S represent the contribution of each feature, the set of all features, and all features excluding i^{th} feature, respectively. Then, two models are retrained, $f_{S \cup \{i\}}$ when i is included and f_S when it is not, and their predictions are compared to the current input $f_{S \cup \{i\}}(x_{S \cup \{i\}}) - f_S(x_S)$, where x_s indicates the values of the input features in the set S.

3. Results and discussion

3.1. Elemental compositions of oysters from different regions

Oysters are sedentary filter-feeders capable of absorbing metals present as dissolved or suspended particles in the water column. The average concentrations of Na, Mg, Ca, and K in oysters obtained from four provinces in this study were 30.26 ± 27.69, 9.94 ± 7.45, 9.49 ± 9.35, and 8.58 ± 2.32 g/kg, respectively. These values were consistent with findings from prior research studies (Lu et al., 2017). The average Al and Fe contents in oysters were 1.21 ± 0.72 and 0.74 ± 0.54 g/kg, respectively. Oysters served as hyper-accumulators for copper Cu and Zn, exhibiting mean concentrations of 0.57 ± 0.46 and 1.85 ± 1.32 g/kg, respectively, which were 1–2 orders of magnitude higher than the other trace elements. In comparison, the average concentrations of Sr, Mn, V, and As in oysters were 60.95, 58.18, 21.74, and 16.74 mg/kg, respectively, vastly surpassing the average concentrations of Cd (8.37 mg/kg), Se (5.89 mg/kg), Cr (3.41 mg/kg), Ag (3.36 mg/kg), Ba (2.93 mg/kg), Li (2.49 mg/kg), Pb (0.89 mg/kg), and Co (0.43 mg/kg). The average Bi content in oysters was only 11.50 μg/kg. Among the metals analyzed, significant positive linear correlations were observed for Cu–Zn ($R^2 = 0.856$, $p < 0.01$), Cr–Ni ($R^2 = 0.933$, $p < 0.01$), and Na–Mg ($R^2 = 0.970$, $p < 0.01$). These findings suggest that the metal stoichiometry of Cu–Zn, Cr–Ni, and Na–Mg remains relatively stable and their accumulation in oysters is closely linked.

The mean ± standard deviation of 22 elements in oysters is presented in Table 1. The results of the One-Way ANOVA test indicate significant differences in the concentrations of 18 elements (Na, Mg, Ca, K, Al, Fe, Zn, Sr, Mn, V, As, Ag, Se, Cd, Li, Ba, Co, and Bi). Similarly, significant differences in the profiling of Al, Mn, Fe, Cu, Zn, V, Cr, Co, Ni, As, Sr, Cd and Pb were detected in a research of scallop origin traceability (Kang et al., 2022). The average concentrations of As, Li, V, Sr, K, Mg, and Na in oysters from the northern regions of China (specifically Liaoning and Shandong Provinces) were significantly higher ($p < 0.05$) than those in oysters from the southern regions of China (specifically Guangdong and Guangxi Provinces). Additionally, the average Ca concentration in northern Chinese oysters was notably higher than that in oysters from Guangdong ($p < 0.05$). Moreover, the average Ag concentration in oysters from Liaoning Province was significantly greater than that in oysters from the other three provinces ($p < 0.05$). The average Fe and Se contents of oysters from Shandong Province were significantly higher than those from Guangdong ($p < 0.05$). Moreover, the average Co content of oysters from Northern China was notably lower than that of Guangxi ($p < 0.05$). Additionally, the average Al, Mn, and Ba contents of Shandong oysters were significantly higher than those of the other three provinces ($p < 0.05$). The elevated concentrations of elements in Northern China may be attributed to the lower environmental temperatures in the region, which result in reduced metabolic rates. Prior studies have established a significant negative correlation between temperature and various elements (such as Al, Ag, Co, Fe, Mn, Se, and V) in oysters (Ward and Flick, 1990). The concentration of Bi in oysters from Guangdong Province was significantly higher than in the other three provinces ($p < 0.05$). Additionally, noteworthy levels of Cd were detected in oysters from Southern China. Moreover, the average Zn concentration in oysters collected from Southern China was significantly higher than in Shandong Province ($p < 0.05$). Aside from temperature, various factors can also influence the geographical variances in mineral elements found in oysters. For instance, elevated particle loads in the water column could lead to higher concentrations of elements in oysters due to increased production of pseudo-faeces and/or diminished food and energy acquisition by the oysters (Chouvelon et al., 2022). Moreover, there were significant variations in both the content and structure of microalgae found in different sea locations (Li et al., 2021). Differences in diet, specifically phytoplankton, can influence the accumulation of K, Ca, Mg, Mn, Fe, Al, Ba and Pb in oysters (Vilhena et al., 2016). Consequently, the single factor mentioned above or the combined influence of these factors can

Table 1
Elemental concentrations (mean ± standard) of oysters from different regions.

Elements	Unit	Liaoning	Shandong	Guandong	Guangxi	*p*-value
Na	g kg^{-1}	47.89 ± 7.33^{b}	90.11 ± 13.29^{a}	18.08 ± 6.39^{c}	13.22 ± 3.75^{c}	<0.001
Mg	g kg^{-1}	14.61 ± 2.33^{b}	25.56 ± 4.01^{a}	6.71 ± 2.77^{c}	5.59 ± 0.94^{c}	<0.001
Ca	g kg^{-1}	13.81 ± 4.72^{ab}	18.49 ± 16.31^{a}	5.65 ± 4.28^{c}	11.18 ± 8.68^{bc}	<0.001
K	g kg^{-1}	12.05 ± 0.95^{a}	11.53 ± 0.91^{a}	7.50 ± 1.79^{b}	7.90 ± 1.38^{b}	<0.001
Al	g kg^{-1}	1.38 ± 0.83^{b}	1.95 ± 0.85^{a}	1.04 ± 0.55^{b}	1.09 ± 0.66^{b}	<0.001
Fe	g kg^{-1}	0.96 ± 0.45^{ab}	1.26 ± 0.49^{a}	0.61 ± 0.55^{b}	0.63 ± 0.32^{b}	<0.001
Cu	g kg^{-1}	0.38 ± 0.03^{a}	0.42 ± 0.08^{a}	0.62 ± 0.55^{a}	0.63 ± 0.41^{a}	0.260
Zn	g kg^{-1}	1.52 ± 0.14^{ab}	0.68 ± 0.16^{b}	2.05 ± 1.54^{a}	2.27 ± 0.94^{a}	0.001
Ni	mg kg^{-1}	1.93 ± 0.63^{a}	3.71 ± 1.17^{a}	3.03 ± 5.95^{a}	2.83 ± 1.16^{a}	0.841
Sr	mg kg^{-1}	87.02 ± 19.80^{b}	163.41 ± 65.76^{a}	33.81 ± 24.44^{c}	48.62 ± 30.72^{c}	<0.001
Mn	mg kg^{-1}	63.38 ± 21.78^{b}	121.02 ± 23.99^{a}	43.03 ± 21.83^{c}	51.12 ± 26.45^{bc}	<0.001
V	mg kg^{-1}	25.60 ± 1.26^{b}	28.21 ± 1.32^{a}	20.63 ± 2.71^{c}	18.64 ± 1.86^{d}	<0.001
As	mg kg^{-1}	20.00 ± 0.98^{a}	20.97 ± 1.97^{a}	15.50 ± 4.15^{b}	15.68 ± 3.29^{b}	<0.001
Ag	mg kg^{-1}	6.45 ± 0.58^{a}	2.88 ± 0.59^{bc}	2.69 ± 2.06^{c}	4.14 ± 2.33^{b}	<0.001
Se	mg kg^{-1}	6.02 ± 0.90^{ab}	6.83 ± 1.13^{a}	5.18 ± 1.17^{b}	6.96 ± 1.91^{a}	<0.001
Cd	mg kg^{-1}	5.389 ± 0.41^{b}	4.52 ± 0.81^{b}	9.57 ± 5.86^{a}	9.18 ± 2.10^{a}	0.001
Li	mg kg^{-1}	3.76 ± 1.15^{b}	5.06 ± 1.29^{a}	1.88 ± 0.72^{c}	1.78 ± 0.75^{c}	<0.001
Cr	mg kg^{-1}	3.22 ± 0.60^{a}	4.63 ± 2.55^{a}	3.45 ± 6.94^{a}	2.57 ± 0.82^{a}	0.731
Ba	mg kg^{-1}	1.90 ± 1.01^{b}	12.68 ± 8.49^{a}	1.09 ± 1.16^{b}	1.26 ± 0.82^{b}	<0.001
Co	mg kg^{-1}	0.55 ± 0.22^{a}	0.57 ± 0.17^{a}	0.41 ± 0.29^{ab}	0.36 ± 0.13^{b}	0.036
Pb	mg kg^{-1}	0.49 ± 0.17^{a}	0.55 ± 0.15^{a}	1.18 ± 2.86^{a}	0.55 ± 0.34^{a}	0.549
Bi	μg kg^{-1}	3.13 ± 2.12^{b}	2.59 ± 1.79^{b}	16.64 ± 11.12^{a}	8.18 ± 4.50^{b}	<0.001

Note: The superscript letters indicated significant difference among groups ($p < 0.05$) under Duncan test.

contribute to the observed patterns of mineral elements in oysters from different origins. The variability in mineral elements demonstrates the potential for utilizing geographical orientation to trace the origin of oysters.

3.2. Evaluation of oyster quality and safety in diverse regions

As highlighted in the introduction, individuals may be exposed to toxic elements when consuming oysters for essential elements. Moreover, excessive intake of essential elements can also pose risks to human health (Zoroddu et al., 2019). To assess the potential exposure risk to essential and toxic elements from consumed oysters, a comparison was made between the levels of these elements in oysters ingested by adults and the tolerable reference levels specified by the European Food Safety Authority (EFSA). This comparison utilized a semi-deterministic approach, focusing on the average intake of elements for an adult with a body weight (bw) of 70 kg, consuming a 15 g serving portion of dried edible media (equivalent to approximately 100 g of fresh oyster tissue) daily. The assumption of consuming 15 g of dried oysters daily is derived from prior studies citing the water content of oysters as 85% (Zhao et al., 2022). The average levels of essential and toxic elements in the diet obtained from consuming oysters were outlined in Table 2.

Consuming oysters from the four provinces provides a safe source of essential elements including Na, K, Ca, Fe and Mn (Table 2). In terms of Mg content, oysters from Shandong exhibit inferior quality in comparison to those from the other three provinces. However, there have been no reports of toxic effects produced by excessive magnesium ingested from food (EFSA, 2015d). Given the elevated levels of Cu and Zn present in oysters, despite these elements being essential for the human body, caution is advised regarding the quantity consumed. Excessive intake of Cu can lead to significant genetic disorders, such as Wilson's disease (EFSA, 2015b). Chronic intake of excessive Zn can lead to severe neurological diseases due to copper deficiency (EFSA, 2014c). Moreover, the potential risk associated with toxic elements such as As, Cd, V and Al from the consumption of oysters should be a primary concern (Table 2). The excessive consumption of As (particularly inorganic arsenic) by humans has been linked to various detrimental effects, including skin lesions, cancer, and developmental toxicity (EFSA, 2009b). While excessive intake of Cd can impact DNA repair, gene expression, and even lead to kidney damage (EFSA, 2009a). An excessive intake of V can lead to gastrointestinal disturbances, such as diarrhea and abdominal cramps (EFSA, 2004). However, there is still controversy surrounding the harmful effects of excessive intake of Al on the human body (EFSA, 2008). In our study, the dietary intake of toxic elements such as Pb, Se, Ni, and Cr was found to be below the safety threshold set by EFSA. Overall, oysters from Liaoning and Shandong exhibit higher levels of essential elements compared to those from Guangdong and Guangxi. Nevertheless, when evaluating the influence of multiple toxic elements in oysters, determining the superior oysters poses a challenge (Table 2).

3.3. Model comparison and selection

A learning curve is a commonly employed method for monitoring the effectiveness of model training (Kang et al., 2023). This method allows for the demonstration of an estimator's validation and training scores as the training data fluctuates. It helps determine whether the estimator is more affected by a variance error or a bias error when additional data is utilized in the training process (Gupta, 2019). The means with standard errors of the cross-validation (CV) are presented in Fig. 3. As depicted in Fig. 3a, b, and c, when a limited amount of training data was utilized, the training scores of RF, GBDT, and LightGBM significantly outperformed the CV score. As additional training data was included, the CV score gradually aligned with the training score, revealing the model's potential for generalization. Nevertheless, the training score of the XGBoost model exhibited a decreasing trend when only a small amount of

Table 2
Comparison of the dietary intake of essential and toxic elements by consuming oysters from different provinces by adults with the daily intake values recommended by EFSA.

Elements	Na mg/day	K mg/day	Ca mg/day	Cu mg/day	Fe mg/day	Mgmg/day	Mn μg/day	Zn mg/day	Se μg/day	Ni μg/day	Al mg/day	V μg/day	Cr μg/day	Cd μg/day	As μg/day	Pb μg/day
Liaoning	718.4	180.8	207.2	**5.7**	14.4	219.2	950.7	22.8	90.3	29.0	**20.7**	**384.0**	48.3	**80.8**	**300.0**	7.4
Shandong	1351.6	173.0	277.4	**6.3**	18.9	**383.4**	1815.3	10.2	102.4	55.6	**29.3**	**423.2**	69.4	**67.8**	**314.6**	8.2
Guangdong	271.2	112.5	84.8	**9.3**	9.15	100.6	645.4	**30.8**	77.7	45.4	**15.6**	**309.4**	51.8	**143.6**	**232.5**	17.7
Guangxi	198.3	118.5	167.7	**9.4**	9.45	83.8	766.8	**34.0**	104.4	42.4	**16.4**	**279.6**	38.6	**137.7**	**235.2**	8.2
Reference values of EFSA	AI (EFSA, 2019)	AI (EFSA, 2016)	TUI (EFSA, 2015a)	TUI (EFSA, 2015b)	TUI (EFSA, 2015c)	AI (EFSA, 2015d)	AI (EFSA, 2013)	TUI (EFSA, 2014c)	TUI (EFSA, 2014b)	TUI (EFSA, 2005)	TWI 1 mg/kg bw/week (EFSA, 2008)	TUI (EFSA, 2004)	TDI 300 μg/kg bw/day (EFSA, 2014a)	TWI 2.5 μg/kg bw/week (EFSA, 2009a)	TWI 15 μg/kg bw/week (EFSA, 2009b)	TDI 0.50 μg/kg bw/day (EFSA, 2010)
Daily safe intake levels for adults#	2000	3500	2500	5	45	350 (men), 300 (women)	3000	25	300	150	10	10–20	21,000	25	150	35

Note: The bold numbers indicate that the intake of essential and toxic elements in the oyster exceeds the daily intake values for adults, as calculated according to EFSA's recommendations (the unit is same with Elements); TWI, tolerable weekly intake; TDI, tolerable daily intake; TUI, Tolerable upper intake; AI, adequate intakes; # = TWI/7 × average body weight (assumed as 70 kg for adults) or TDI × average body weight or TUI or AI.

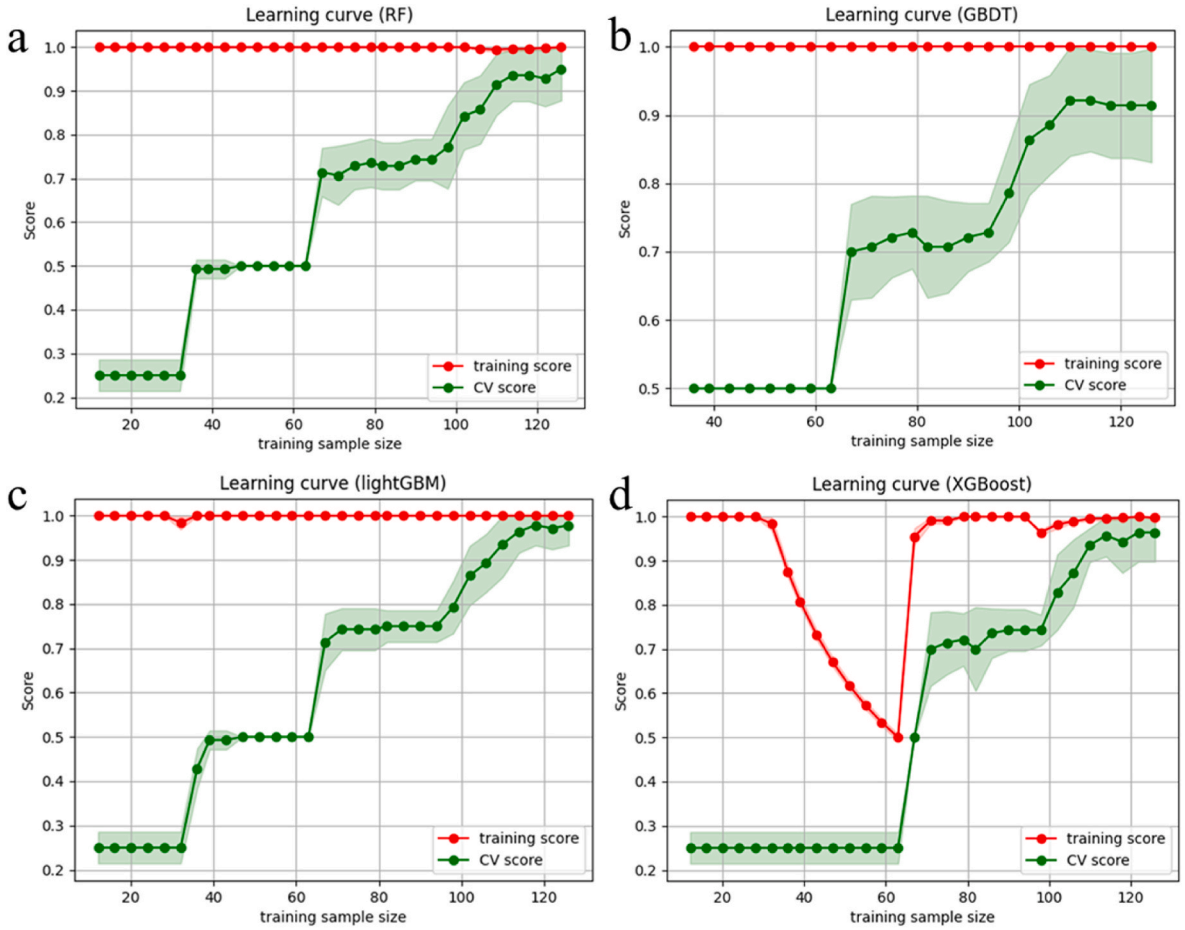

Fig. 3. Learning curves of RF, GBDT, LightGBM, and XGBoost model.

training data was used, while the CV score progressively converged with the training score as more data was incorporated (Fig. 3d). Hence, to mitigate overfitting in the XGBoost model, an increased amount of training data should be employed.

The performance of the models was evaluated utilizing a test set to ascertain their generalization capabilities. LightGBM model exhibited best performance, achieving a higher accuracy (96.77%), precision (96.43%), recall (98.53%), F1 score (97.32%), and AUC (0.998). The performance of the GBDT model was found to be comparable to that of the XGBoost model, and superior to that of the RF model (Table 3). In general, all four models exhibited satisfactory performance for oyster origin tracking. To focus on the discrimination abilities of the LightGBM model, we presented ROC curves and a confusion matrix (Fig. 4). The LightGBM model exhibited an exceptional ability to differentiate between oysters from the four provinces, as evidenced by an AUC of 0.998 (Fig. 4a). Based on the confusion matrix (Fig. 4b), all test samples from the Liaoning, Shandong, and Guangdong provinces were accurately categorized, except for one from Guangxi Province being misclassified as a neighboring province, Guangdong. The prediction accuracy of this study (96.77%) exceeded that of the pyrolysis mass spectrometry fingerprint (89.2%) and the combination of compound-specific isotope and elemental analysis (90.0%) (Ratel et al., 2008; Matos et al., 2021), and it was on par with the untargeted lipidomics approach and attenuated total reflectance Fourier-transform infrared spectroscopy in terms of determining the oyster origin (Liu et al., 2022; Guo et al., 2023). When considering the cost and accuracy of the technology, mineral element analysis combined with the LightGBM algorithm is undoubtedly suitable for identifying the origin of oysters in China.

Table 3
Overall performance of Oyster origin classification models evaluated with the test set.

Models	Accuracy (%)	Precision (%)	Recall (%)	F1 score (%)	AUC
RF	90.32	91.29	92.89	91.96	0.996
GBDT	93.55	94.87	94.87	94.87	0.995
LightGBM	96.77	96.43	98.53	97.32	0.998
XGBoost	93.55	94.87	94.87	94.87	0.977

3.4. Interpretation of LightGBM model with SHAP values

Transparency and interpretability are important criteria for assessing the viability of machine learning models. However, machine learning models with sophisticated nonlinear algorithms are often complex and may involve intricate interactions of numerous factors or features, making it challenging for users to understand the reasoning behind the model's outputs. Generally, a post-hoc explanation will be employed for a sophisticated machine learning model (Ekanayake et al., 2022). Hence, the use of SHAP explanations to uncover the inherent nature of a specific prediction, understand the interaction between each variable and the model output, and shed light on a particular example. It is essential for evaluating and enhancing model performance (Park et al., 2022).

3.4.1. Global feature interpretation

The interpretation of the global feature contribution to an origin detection model can be achieved through the utilization of a stacked bar plot (Kang et al., 2023; Huang et al., 2023). In this study, a stacked bar

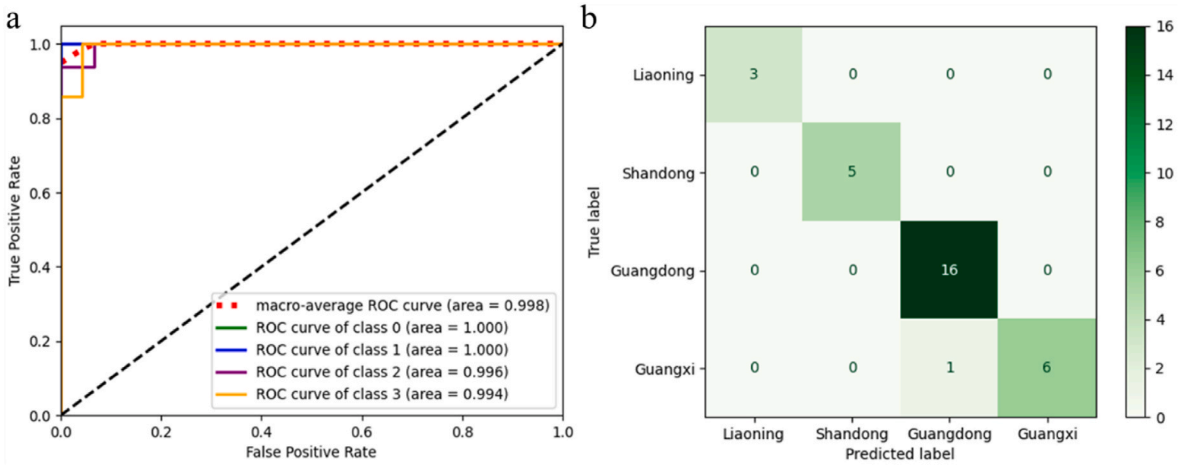

Fig. 4. LightGBM model predicting performance for samples from each geographical origin (a: ROC curves; b: confusion matrix).

plot was employed to demonstrate the impact of different features on the output of the LightGBM model. The features were arranged based on the average absolute SHAP values across the four provinces (Fig. 5a). It became apparent that among the variables, Na, Zn V, Mg, and K exerted the most significant influence on the model's output for determining the origin of the oyster. Nevertheless, owing to variations in origin, the contributions of each feature to the model output exhibited differences.

In the SHAP summary plot of Fig. 5b–e, each sample is represented as a point. The horizontal axis indicates the impact on the prediction through SHAP values, while the color gradient, ranging from blue to red, signifies the intensity of the corresponding feature value. This plot effectively combines feature importance with the directional relationship between feature values and their impact on predictions (Parsa et al., 2020; Maroni et al., 2022). For instance, characteristics such as Ag, K, V, Na and Mg exhibited the most pronounced influence on the predictions for samples from Liaoning (Fig. 5b). The presence of elevated levels of Ag, K and V increased the likelihood of classifying the sample as originating from Liaoning. These findings align with previous research demonstrating the high concentration of K in the Bohai Sea, which can impact oyster mariculture in Liaoning province (Lu et al., 2017). Another notable observation supporting these results is the significantly higher Ag content in oysters from Liaoning compared to other provinces. While the feature Fe had SHAP values close to zero, suggesting that its impact on the classification of the observations into the Liaoning group was not substantial (Fig. 5b).

Zn emerged as the primary contributor in predicting the samples from Shandong, with the impact of high feature value samples to model output was negative (Fig. 5c). In addition, V, Na, Ba, Mg, Mn, Li and Sr were recognized as notable features influencing the model's result. Samples with high feature values of these elements had a positive effect

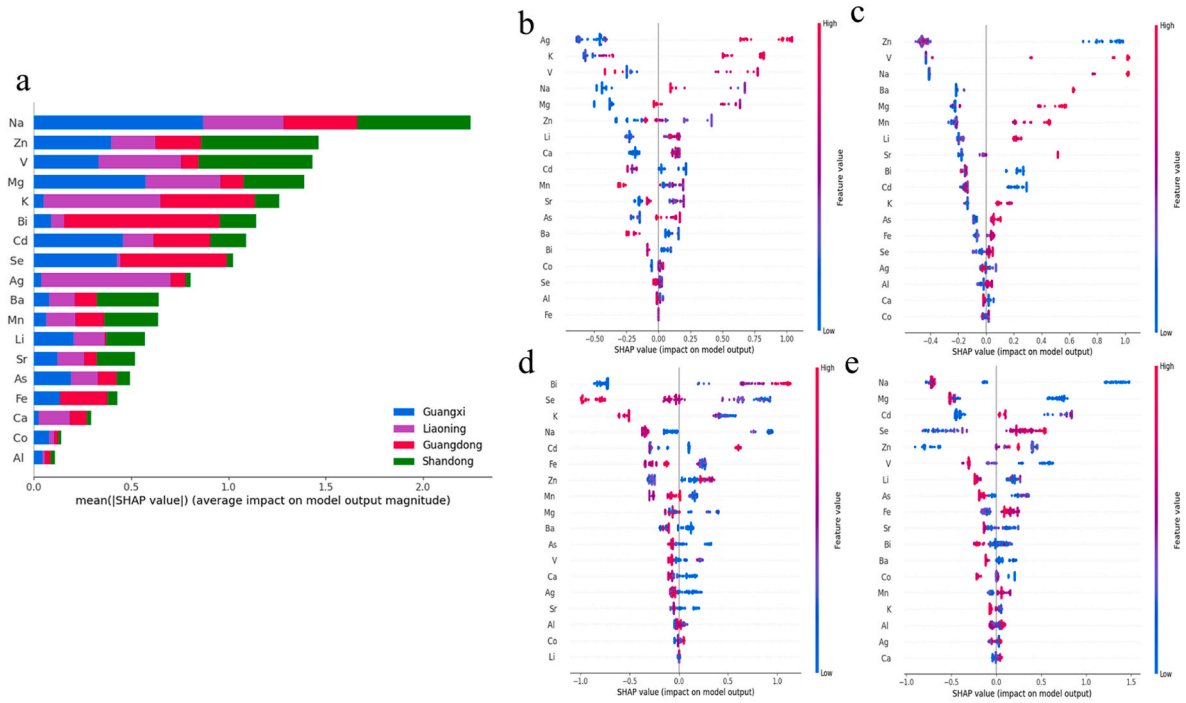

Fig. 5. Global interpretation of the LightGBM model based on SHAP values: (a) the feature importance for the LightGBM model; while (b), (c), (d), and (e) represent the feature contributions for predicting samples from Liaoning, Shandong, Guangdong, and Guangxi, respectively. The color bar, ranging from blue to red, indicates the magnitude of feature values from low to high. Additionally, the position of the points on the horizontal axis denotes the positive or negative association between the features and target variables. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

on the model's output. In contrast, additional features like Se, Ag, Al, and Ca had minimal influence on the model's output (Fig. 5c). Oysters exhibit salinity conformity, with the Na levels in their tissues being rapidly influenced by the salinity levels in their environment (Lu et al., 2017). Additionally, variations in Mg and Sr may reflect changes in ambient temperature, as the Mg/Ca and Sr/Ca ratios have been utilized as proxies for temperature reconstruction (Poulain et al., 2015).

Bi, Se, K, and Na were identified as the primary influencers on the model output for Guangdong samples, where samples with low feature values of K, Na, and Se exhibited positive SHAP values (Fig. 5d). This prediction can be explained by the generally low levels of K and Na found in South China Sea oysters (Lu et al., 2017).

Na, Mg, Cd, and Se were determined as the key features influencing the predictions of Guangxi samples, where low feature values of Na and Mg exhibited a positive SHAP value (Fig. 5e). Nevertheless, the feature value of Se showed an opposite effect on the model output in predictions for Guangdong and Guangxi samples (Fig. 5d and e). In addition, high feature values of V, Li and As were discovered to have a negative impact (Fig. 5e). These findings indicate that the LightGBM model shared contributions from high and low feature value samples across the Liaoning, Shandong, Guangdong, and Guangxi identifications. Moreover, variations in environmental characteristics, such as temperature, salinity, elemental contents, and other factors, enabled the determination of the oyster's origin.

3.4.2. *Local feature interpretation*

The force plot of SHAP illustrates the contribution of local features based on SHAP values for randomly selected cases from each province, along with the predicted geographical origin (Fig. 6). A pink band represents a feature that increases the possibility of predicted provinces, while a blue band suggests that the feature reduces the possibility of predicted provinces. The length of the band signifies the magnitude of the effect (Nordin et al., 2023; Pradhan et al., 2023). The assigned sample is classified as belonging to the projected province if the model output is to the right of the base value; otherwise, it is not (Han et al., 2022b; Huang et al., 2023). It can be seen that the model output is clearly visible to the right of the base value, indicating that the four cases all belong to the selected provinces. However, the feature contributions vary for each instance. For the Liaoning situations, the base value is −3.06. The selected samples demonstrate a relatively higher projection towards the Liaoning group at 1.95, with all the features increases the predicted Liaoning province for this sample (Fig. 6a). In the Shandong cases, the 10 features (e.g. Na, V, Zn, and Ba) are associated with positive SHAP values, while the remaining 8 features have minimal impact on determining the origin attribution (Fig. 6b). Conversely, in the Guangdong cases, Se has little detrimental influence, with the selected samples showing a significantly higher projected Guangdong value at 1.74. In the Guangxi cases, the features Na, Mg, Cd and V contribute to pushing the prediction above the base value (Fig. 6d). In general, the effect of features on geographical origin in the case study is consistent with the mariculture environment knowledge as discussed in section 3.1. Further, the local feature interpretation aligns with the trend observed in the global feature interpretation, thereby demonstrating the reliable explanatory capability of SHAP for individual predictions made by our models.

4. Conclusion

This study emphasizes the regional discrepancies in essential and toxic elements found in oysters across various production areas, while underscoring the potential hazards of consuming oysters excessively. Moreover, the efficacy of explainable machine learning in identifying the provenance of oysters through mineral elements was demonstrated. The performance of the LightGBM model excelled in classifying oysters from the four provinces compared to the other three models. Subsequently, the LightGBM model was interpreted using SHAP to provide insights on the importance of Na, Zn, V, Mg, and K, in the estimation process. These findings establish causal relationships among features, enhancing user confidence and ensuring reliable practical applications of the models. This innovative contribution extends beyond the development and utilization of machine learning, advancing the concept of interpretability within the field of seafood origin traceability. Ultimately, this research will aid in the prevention and control of food fraud and ensure food safety.

Although successful in achieving its objectives, this study had several limitations. Specifically, it did not address the impact of oyster species and sampling time on the predictive model's reliability. Therefore, future studies should consider a more comprehensive research plan, including a broader range of influencing factors, incorporating additional sampling locations, and integrating alternative discrimination

Fig. 6. Local interpretation based on SHAP values for instances randomly selected from (a) Liaoning, (b) Shandong, (c) Guangdong, and (d) Guangxi.

techniques to enhance result accuracy. Moreover, the risk of potential data leakage should be considered in the field of food origin traceability.

Ethical approval

This research does not contain any studies with human participants or animals carried out by all of the authors.

CRediT authorship contribution statement

Xuming Kang: Conceptualization, Methodology, Software, Data curation, Writing – original draft, Funding acquisition. **Yanfang Zhao:** Writing – review & editing. **Lin Yao:** Review. **Zhijun Tan:** Writing – review & editing, Supervision, Funding acquisition.

Declaration of competing interest

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

Data availability

Data will be made available on request.

Acknowledgments

This research was supported by the National Natural Science Foundation of China [grant number 32202156], the Shandong Provincial Natural Science Foundation [grant number ZR2022QC067], the Central Public-interest Scientific Institution Basal Research Fund, YSFRI, CAFS [grant number 20603022022017], and the Central Public-interest Scientific Institution Basal Research Fund, CAFS [grant numbers 2023TD76 and 2023TD28]. Additionally, the earmarked fund for CARS-49 contributed to this study.

References

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.
Chen, T., Guestrin, C., 2016. XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining, vol. 16. KDD', pp. 785–794. https://doi.org/10.48550/arXiv.1603.02754. Association for Computer Machinery.
Chouvelon, T., Auby, I., Mornet, L., Bruzac, S., Charlier, K., Araújo, D.F., Gonzalez, J.L., Gonzalez, P., Gourves, P.Y., Méteigner, C., Perriere-Rumèbe, M., Rigouin, L., Rozuel, E., Savoye, N., Sireau, T., Akcha, F., 2022. Role of suspended particulate material on growth and metal bioaccumulation in oysters (*Crassostrea gigas*) from a French coastal semi-enclosed production area, Arcachon Bay. J. Mar. Syst. 234, 103778.
Deng, X.Y., Cao, S.H., Horn, A.L., 2021. Emerging applications of machine learning in food Safety. Annu. Rev. Food Sci. Technol. 12, 513–538.
EFSA, 2009a. Scientific Opinion of the Panel on Contaminants in the food Chain on a request from the European Commission on cadmium in food. EFSA J. 7 (3), 980. https://doi.org/10.2903/j.efsa.2009.980.
EFSA, 2009b. Scientific Opinion of the Panel on Contaminants in the food Chain on a request from the European Commission on arsenic in food. EFSA J. 7 (10), 1351. https://doi.org/10.2903/j.efsa.2009.1351.
EFSA, 2010. Scientific Opinion of the Panel on Contaminants in the food Chain on a request from the European Commission on lead in food. EFSA J. 8 (4), 1570. https://doi.org/10.2903/j.efsa.2010.1570.
EFSA AFC Panel (EFSA Panel on Food Additives, Flavourings, Processing Aids and Food Contact Materials), 2008. Scientific Opinion on Safety of aluminium from dietary intake. The EFSA J 754, 1–34. https://doi.org/10.2903/j.efsa.2008.754.
EFSA NDA Panel (EFSA Panel on Dietetic Products, Nutrition and Allergies), 2004. Scientific Opinion on tolerable upper intake level of vanadium. EFSA J. 33, 1–22. https://doi.org/10.2903/j.efsa.2004.33.
EFSA NDA Panel (EFSA Panel on Dietetic Products, Nutrition and Allergies), 2005. Scientific Opinion on tolerable upper intake level of nickel. EFSA J. 146, 1–21. https://doi.org/10.2903/j.efsa.2005.146.
EFSA NDA Panel (EFSA Panel on Dietetic Products, Nutrition and Allergies), 2013. Scientific Opinion on dietary reference values for manganese. EFSA J. 11 (11), 3419. https://doi.org/10.2903/j.efsa.2013.3419.
EFSA NDA Panel (EFSA Panel on Dietetic Products, Nutrition and Allergies), 2014a. Scientific Opinion on dietary reference values for chromium. EFSA J. 12 (10), 3845. https://doi.org/10.2903/j.efsa.2014.3845.
EFSA NDA Panel (EFSA Panel on Dietetic Products, Nutrition and Allergies), 2014b. Scientific Opinion on dietary reference values for selenium. EFSA J. 12 (10), 3846. https://doi.org/10.2903/j.efsa.2014.3846.
EFSA NDA Panel (EFSA Panel on Dietetic Products, Nutrition and Allergies), 2014c. Scientific Opinion on dietary reference values for zinc. EFSA J. 12 (10), 3844. https://doi.org/10.2903/j.efsa.2014.3844.
EFSA NDA Panel (EFSA Panel on Dietetic Products, Nutrition and Allergies), 2015a. Scientific Opinion on dietary reference values for calcium. EFSA J. 13 (5), 4101. https://doi.org/10.2903/j.efsa.2015.4101.
EFSA NDA Panel (EFSA Panel on Dietetic Products, Nutrition and Allergies), 2015b. Scientific Opinion on dietary reference values for copper. EFSA J. 13 (10), 4253. https://doi.org/10.2903/j.efsa.2015.4253.
EFSA NDA Panel (EFSA Panel on Dietetic Products, Nutrition and Allergies), 2015c. Scientific Opinion on dietary reference values for iron. EFSA J. 13 (10), 4254. https://doi.org/10.2903/j.efsa.2015.4254.
EFSA NDA Panel (EFSA Panel on Dietetic Products, Nutrition and Allergies), 2015d. Scientific Opinion on dietary reference values for magnesium. EFSA J. 13 (7), 4186. https://doi.org/10.2903/j.efsa.2015.4186.
EFSA NDA Panel (EFSA Panel on Dietetic Products, Nutrition and Allergies), 2016. Scientific Opinion on dietary reference values for potassium. EFSA J. 14 (10), 4592. https://doi.org/10.2903/j.efsa.2016.4592.
EFSA NDA Panel (EFSA Panel on Dietetic Products, Nutrition and Allergies), 2019. Scientific Opinion on dietary reference values for sodium. EFSA J. 17 (9), 5778. https://doi.org/10.2903/j.efsa.2019.5778.
Ekanayake, I.U., Meddage, D.P.P., Rathnayake, U., 2022. A novel approach to explain the black-box nature of machine learning in compressive strength predictions of concrete using Shapley additive explanations (SHAP). Case Stud. Constr. Mater. 16, e01059.
Fernández, A., García, S., Herrera, F., Chawla, N., 2018. SMOTE for learning from imbalanced data: progress and challenges, marking the 15-year anniversary. J. Artif. Intell. Res. 61 (1), 863–905.
Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. Ann. Stat. 29 (5), 1189–1232.
Goodwin, N.L., Nilsson, S.R.O., Choong, J.J., Golden, S.A., 2022. Toward the explainability, transparency, and universality of machine learning for behavioral classification in neuroscience. Curr. Opin. Neurobiol. 73, 102544.
Guo, B.J., Zou, Z.W., Huang, Z., Wang, Q.Y., Qin, J.H., Guo, Y., Pan, S.H., Wei, J.B., Guo, H.W., Zhu, D., Su, Z.H., 2023. A simple and green method for simultaneously determining the geographical origin and glycogen content of oysters using ATR–FTIR and chemometrics. J. Food Compos. Anal. 119, 105229.
Gupta, 2019. Evaluating the accuracy of valuation multiples on Indian firms using regularization techniques of penalized regression. Theor. Econ. Lett. 9, 180–209.
Han, C., Li, L., Dong, X., Gao, Q.F., Dong, S.L., 2022a. Current progress in the authentication of fishery and aquatic products using multi-element and stable isotope analyses combined with chemometrics. Rev. Aquacult. 12686.
Han, L., Yang, G.J., Yang, X.D., Song, X.Y., Xu, B., Li, Z.H., Wu, J.T., Yang, H., Wu, J.W., 2022b. An explainable XGBoost model improved by SMOTE-ENN technique for maize lodging detection based on multi-source unmanned aerial vehicle images. Comput. Electron. Agric. 194, 106804.
Honig, A., Etter, R., Pepperman, K., Morello, S., Hannigan, R., 2020. Site and age discrimination using trace element fingerprints in the blue mussel, *Mytilus edulis*. J. Exp. Mar. Biol. Ecol. 522, 151249.
Huang, J.X., Li, Z., Zhang, W., Lv, Z.Y., Dong, S.Y., Feng, Y., Liu, R.X., Zhao, Y., 2023. Explainable machine learning-assisted origin identification: chemical profiling of five lotus (*Nelumbo nucifera Gaertn.*) parts. Food Chem. 404, 134517.
Jiménez-Carvelo, A.M., González-Casado, A., Bagur-González, M.G., Cuadros-Rodríguez, L., 2019. Alternative data mining/machine learning methods for the analytical evaluation of food quality and authenticity-a review. Food Res. Int. 122, 25–39.
Kang, X.M., Zhao, Y.F., Liu, W., Ding, H.Y., Zhai, Y.X., Ning, J.S., Sheng, X.F., 2021. Geographical traceability of sea cucumbers in China via chemometric analysis of stable isotopes and multi-elements. J. Food Compos. Anal. 99, 103852.
Kang, X.M., Zhao, Y.F., Peng, J.X., Ding, H.Y., Tan, Z.J., Han, C., Sheng, X.F., Liu, X.Y., Zhai, Y.S., 2022. Authentication of the geographical origin of Shandong scallop *Chlamys farreri* using mineral elements combined with Multivariate data analysis and machine learning algorithm. Food Anal. Methods 15, 2984–2993.
Kang, X.M., Zhao, Y.F., Shang, D.R., Zhai, Y.X., Ning, J.S., Sheng, X.F., 2018. Elemental analysis of sea cucumber from five major production sites in China: a chemometric approach. Food Control 94, 361–367.
Kang, X.M., Zhao, Y.F., Tan, Z.J., 2023. An explainable learning for geographical origin traceability of mussels *Mytilus edulis* based on stable isotope ratio and compositions of C, N, O and H. J. Food Compos. Anal. 123, 105508.
Kapoor, S., Narayanan, A., 2023. Leakage and the reproducibility crisis in machinelearning-based science. Patterns 4, 100804.
Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q.W., Liu, T.Y., 2017. LightGBM: a highly efficient gradient boosting decision tree. Adv. Neural Inf. Process. Syst. 30, 3149–3157.
Kim, Y., Kim, Y., 2022. Explainable heat-related mortality with random forest and SHapley Additive exPlanations (SHAP) models. Sustain. Cities Soc. 79, 103677.
Lemaître, G., Nogueira, F., Aridas, C.K., 2017. Imbalanced-learn: a python toolbox to tackle the curse of imbalanced datasets in machine learning. J. Mach. Learn. Res. 18, 1–5.

Li, X.Y., Yu, R.C., Gen, H.X., Li, Y.F., 2021. Increasing dominance of dinofagellate red tides in the coastal waters of Yellow Sea, China. Mar. Pollut. Bull. 168, 112439.
Li, Z.Q., 2022. Extracting spatial effects from machine learning model using local interpretation method: an example of SHAP and XGBoost. Comput. Environ. Urban Syst. 96, 101845.
Liu, Z.Q., Zhao, M.T., Wang, X.W., Li, C., Liu, Z.Y., Shen, X.R., Zhou, D.Y., 2022. Investigation of oyster *Crassostrea gigas* lipid profile from three sea areas of China based on non-targeted lipidomics for their geographic region traceability. Food Chem. 386, 132748.
Loaiza, I., Wong, C., Thiyagarajan, V., 2023. Comparative analysis of nutritional quality of edible oysters cultivated in Hong Kong. J. Food Compos. Anal. 118, 105159.
Lu, G.Y., Ke, C.H., Zhu, A.J., Wang, W.X., 2017. Oyster-based national mapping of trace metals pollution in the Chinese coastal waters. Environ. Pollut. 224, 658–669.
Lundberg, S.M., Erion, G., Chen, H., Degrave, A., Prutkin, J.M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., Lee, S.I., 2019. Explainable AI for Trees: from Local Explanations to Global Understanding. https://doi.org/10.48550/arXiv.1905.04610.
Lundberg, S.M., Erion, G., Chen, H., DeGrave, A., Prutkin, J.M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., Lee, S.I., 2020. From local explanations to global understanding with explainable AI for trees. Nat. Mach. Intell. 2, 56–67.
Lundberg, S.M., Lee, S.L., 2017. A unified approach to interpreting model predictions. In: Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (Eds.), Advances in Neural Information Processing Systems, vol. 30. Curran Associates, Inc.
Maroni, G., Pallante, L., Benedetto, G.D., Deriu, M.A., Piga, D., Grasso, G., 2022. Informed classification of sweeteners/bitterants compounds via explainable machine learning. Curr. Res. Food Sci. 5, 2270–2280.
Martini, G., Bracci, A., Riches, L., Jaiswal, S., Corea, M., Rivers, J., Husain, A., Omodei, E., 2022. Machine learning can guide food security efforts when primary data are not available. Nat. Food 3, 716–728.
Matos, M.P.V., Engel, M.E., Mangrum, J.B., Jackson, G.P., 2021. Origin determination of the Eastern oyster (*Crassostrea virginica*) using a combination of whole body compound-specific isotope analysis and heavy metal analysis. Anal. Methods 13 (31), 3493–3503.
MOAC (Ministry of Agriculture, China), 2023. China Fisheries Yearbook. China Agriculture Publisher, Beijing, China.
Mouchi, V., Godbillot, C., Dupont, C., Vella, M.A., Forest, V., Ulianov, A., Lartaud, F., Rafélis, M.D., Emmanuel, L., Verrecchia, E.P., 2021. Provenance study of oyster shells by LA-ICP-MS. J. Archaeol. Sci. 132, 105418.
Ng, T.Y.T., Chuang, C.Y., Stupakoff, I., Christy, A.E., Cheney, D.P., Wang, W.X., 2010. Cadmium accumulation and loss in the Pacific oyster *Crassostrea gigas* along the west coast of the USA. Mar. Ecol. Prog. Ser. 401, 147–160.
Nordin, N., Zaino, Z., Noor, M.H.M., Chan, L.F., 2023. An explainable predictive model for suicide attempt risk using an ensemble learning and Shapley Additive Explanations (SHAP) approach. Asian J. Psychiat. 79, 103316.
Park, J., Lee, W.H., Kim, K.T., Park, C.Y., Lee, S., Heo, T.Y., 2022. Interpretation of ensemble learning to predict water quality using explainable artificial intelligence. Sci. Total Environ. 832, 155070.
Parsa, A.B., Shabanpour, R., Mohammadian, A.K., Auld, J., Stephens, T., 2020. A data-driven approach to characterize the impact of connected and autonomous vehicles on traffic flow. Transp. Lett. 687–695.
Pedregosa, F., Varoquaux, G., Gramfort, A., Michl, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: machine learning in python. J. Mach. Learn. Res. 12, 2825–2830.
Poulain, C., Gillikin, D.P., Thébault, J., Munaron, J.M., Bohn, M., Robert, R., Paulet, Y. M., Lorrain, A., 2015. An evaluation of Mg/Ca, Sr/Ca, and Ba/Ca ratios as environmental proxies in aragonite bivalve shells. Chem. Geol. 396, 42–50.
Pradhan, B., Dikshit, A., Lee, S., Kim, H., 2023. An explainable AI (XAI) model for landslide susceptibility modeling. Appl. Soft Comput. 142, 110324.
Ratel, J., Berge, P., Berdague, J.L., Cardinal, M., Engel, E., 2008. Mass spectrometry based sensor strategies for the authentication of oysters according to geographical origin. J. Agric. Food Chem. 56, 321–327.
Ribeiro, M.T., Singh, S., Guestrin, C., 2016. "Why should I trust you?" explaining the predictions of any classifier. In: The 22nd ACM SIGKDD International Conference ACM. https://doi.org/10.1145/2939672.2939778.
Sumaila, U.R., Zeller, D., Hood, L., Palomares, M.L.D., Li, Y., Pauly, D., 2020. Illicit trade in marine fish catch and its effects on ecosystems and people worldwide. Sci. Adv. 6, eaaz3801.
Vilhena, M.P.S.P., Costa, M.L., Berrêdo, J.F., Paiva, R.S., Souza, C.C.S., 2016. Chemical elements in pearl oysters (*Paxyodon ponderosus*), phytoplankton and estuarine sediments from eastern Amazon (Northern Brazil): bioaccumulation factors and trophic transfer factors. J. South Am. Earth Sci. 67, 1–10.
Wang, D., Thunéll, S., Lindberg, U., Jiang, L.L., Trygg, J., Tysklind, M., 2022. Towards better process management in wastewater treatment plants: process analytics based on SHAP values for tree-based machine learning methods. J. Environ. Manag. 301, 113941.
Wang, Y.M., Feng, L.W., Li, S.J., Ren, F., Du, Q.Y., 2020. A hybrid model considering spatial heterogeneity for landslide susceptibility mapping in Zhejiang Province, China. Catena 188, 104425.
Ward, D.R., Flick, G.J., 1990. The effect of salinity and temperature on selected elements in oysters (*Crassostrea virginica*). J. Food Compos. Anal. 3, 96–98.
Zhai, Y.X., Guo, M.M., Jiang, Y.H., Yao, L., Zhao, Y.F., Wu, H.Y., Li, F.L., Tan, Z.J., 2020. Analysis on the quality and safety risks of shellfish products. Chin. Fish. Qual. Stand. 10 (4), 1–25 (in Chinese with English abstract).
Zhang, A.H., Wei, Z.Y., Huang, J.L., Liu, C., 2023. Shell proteins and microstructural analysis identify the origin of shell arts with species resolution in pearl oysters. J. Archaeol. Sci. 151, 105729.
Zhao, X.N., Peng, J.X., Wu, H.Y., Zheng, G.C., Guo, M.M., Kang, X.M., Tan, Z.J., 2022. Biomonthly variation in nutrient composition and taste components of Crassostrea gigas cultured in Rushan, Southern yellow sea. Aquacult. Res. 53 (18), 6711–6720.
Zheng, G.C., Wu, H.Y., Che, H.Y., Li, X.K., Zhang, Z.H., Peng, J.X., Guo, M.M., Tan, Z.J., 2022. Residue analysis and assessment of the risk of dietary exposure to domoic acid in shellfish from the coastal areas of China. Toxins 14, 862.
Zoroddu, M.A., Aaseth, J., Crisponi, G., Medici, S., Peana, M., Nurchi, V.M., 2019. The essential metals for humans: a brief overview. J. Inorg. Biochem. 195, 120–129.