Current Research in Food Science 8 (2024) 100732

Contents lists available at ScienceDirect

Current Research in Food Science

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

Rapid quantitative authentication and analysis of camellia oil adulterated with edible oils by electronic nose and FTIR spectroscopy

Xiaoran Wang [a], Yu Gu [a,b,c,d,*], Weiqi Lin [e], Qian Zhang [e]

[a] College of Information Science and Technology, Beijing University of Chemical Technology, Beijing, 100029, China
[b] School of Automation, Guangdong University of Petrochemical Technology, Maoming, 525000, China
[c] School of Biomedical Engineering, Capital Medical University, Beijing, 100069, China
[d] Beijing Key Laboratory of Basic Research in Clinical Applied Biomechanics, China
[e] Xiamen Products Quality Supervision and Inspection Institute, Xiamen, 361004, China

ARTICLE INFO

Keywords:
Camellia oil
E-nose
FTIR
Adulteration detect

ABSTRACT

Camellia oil, recognized as a high-quality edible oil endorsed by the Food and Agriculture Organization, is confronted with authenticity issues arising from fraudulent adulteration practices. These practices not only pose health risks but also lead to economic losses. This study proposes a novel machine learning framework, referred to as a transformer encoder backbone with a support vector machine regressor (TES), coupled with an electronic nose (E-nose), for detecting varying adulteration levels in camellia oil. Experimental results indicate that the proposed TES model exhibits the best performance in identifying the adulterated concentration of camellia oi, compared with five other machine learning models (the support vector machine, random forest, XGBoost, K-nearest neighbors, and backpropagation neural network). The results obtained by E-nose detection are verified by complementary Fourier transform infrared (FTIR) spectroscopy analysis for identifying functional groups, ensuring accuracy and providing a comprehensive assessment of the types of adulterants. The proposed TES model combined with E-nose offers a rapid, effective, and practical tool for detecting camellia oil adulteration. This technique not only safeguards consumer health and economic interests but also promotes the application of E-nose in market supervision.

1. Introduction

Camellia oil, one of the high-quality edible oils recommended by the Food and Agriculture Organization of the United Nations, is rich in monounsaturated and polyunsaturated fatty acids, including oleic acid, linoleic acid, and antioxidants (Zhang et al., 2022a; Cao et al., 2020). These components are beneficial for the human body, improving cardiac well-being, promoting healthy cholesterol levels, and slowing down the effects of skin aging (Huang et al., 2023; Guo et al., 2020; Li et al., 2022). According to the statistics from the China Rural Development Volunteer Service Promotion Association, the current cultivation area of *Camellia oleifera* in China has reached 6.888 million hectares, yielding 192 billion yuan (approximately 28.62 billion US dollars) in production for camellia oil (http://www.moa.gov.cn/ztzl/ymksn/rmrbbd/202306/t20230612_6429902.htm). Nonetheless, the authenticity and quality of camellia oil have been compromised by fraudulent practices because of its high price and short supply (Shi et al., 2020). These practices include adulteration with cheaper oils such as soybean oil, which can pose significant health risks to consumers and undermine the economic value of authentic camellia oil. Therefore, a reliable method for identifying and authenticating of camellia oil needs to be developed to protect the consumer interests and ensure the continued development and sustainability of the camellia oil industry.

The quality of camellia oil is typically assessed by sensory evaluation and instrumental techniques. Conducted by experts using their senses of smell, sight, and taste, sensory evaluation offers a direct, intuitive assessment of camellia oil quality that can capture nuances which may be overlooked by machines (He et al., 2021). This traditional method, deeply rooted in human experience, can detect subtle differences and complexities in flavor, aroma, and appearance that are essential for quality determination. However, it is inherently subjective and influenced by various factors, including the state of the evaluator and environmental conditions, thus limiting its reproducibility and reliability (Njoman et al., 2017). Instrumental techniques, such as gas or liquid

* Corresponding author. College of Information Science and Technology, Beijing University of Chemical Technology, Beijing, 100029, China.
E-mail address: guyu@mail.buct.edu.cn (Y. Gu).

https://doi.org/10.1016/j.crfs.2024.100732
Received 15 January 2024; Received in revised form 19 March 2024; Accepted 4 April 2024
Available online 16 April 2024

chromatography and spectroscopy, offer precise and quantitative analysis, enabling the detection and identification of complex compounds within camellia oil. The techniques provide detailed insights into the chemical composition of the oil, which information is crucial for ensuring quality and authenticity. However, these techniques require complex sample preparation, costly equipment, and trained professionals. Moreover, these approaches are unsuitable for efficiently screening a large number of samples in real-world settings.

Fourier transform infrared (FTIR) spectroscopy has recently been as a new solution for detecting edible oils (Jamwal et al., 2021; Bunaciu et al., 2023). Ye and Meng (2022) employed the FTIR spectroscopy with chemometrics to authenticate edible oil samples, achieving 100% correct classification of 135 samples from 11 species and recognition rates of 100% and 92.6% for pure oil and blend samples, respectively. Jiménez-Carvelo et al. (2017) attained 100% correct classification for 67 olive oil samples and 92% for other vegetable edible oils by FTIR spectroscopy combined with partial least squares-discriminant analysis and support vector machine (SVM). Windarsih et al. (2023) developed an authentication technique using the FTIR spectroscopy and chemometrics to detect pork oil adulteration in snakehead fish oil, yielding $R^2 > 0.990$ and RMSE <5.00. Despite its precision, reliability, and non-destructiveness, FTIR's applicability is hindered by limitations such as limited portability and the need for skilled operation. These limitations do not meet the requirements of large-scale initial screening in practical market supervision and management.

The electronic nose (E-nose), an instrument designed to mimic human olfactory senses, has demonstrated its effectiveness as a powerful tool in edible oil detection (Majchrzak et al., 2018). Xu et al. (2016) used an E-nose in conjunction with cluster analysis (CA), principal component analysis (PCA), and linear discriminant analysis (LDA), to qualitatively discriminate between non-oxidized and oxidized oils, achieving accuracies of 95.8%, 98.9%, and 100%, respectively. Karami et al. (2020) employed an E-nose, combined with LDA, quadratic discriminant analysis, and SVM to determine the shelf life of edible oil, achieving classification accuracies of 96.25%, 95.8%, and 94.4%, respectively. Wei et al. (2018) used an E-nose in combination with PCA and LDA to distinguish peony seed oil from four other oils, even at considerably low (10%) levels. The utilization of E-nose technology, combined with traditional machine learning models, has demonstrated success in edible oil detection. Previous research has shown the efficacy of these combinations in accurately classifying various conditions of edible oils and distinguishing among different types of oils, with high levels of accuracy. The ability to provide precise classifications based on the complex sensory data captured by the E-nose reflects the significant potential of integrating machine learning with olfactory sensing technologies. However, the traditional approaches necessitate manual feature extraction, a process that introduces limitations when dealing with the diversity of adulteration types. Specifically, the need for manual extraction of distinctive features for each type of adulteration restricts the broader applicability of the E-nose and diminishes its utility in complex detection scenarios. Moreover, previous research has substantially concentrated on identifying adulteration from a single source, focusing narrowly on specific adulterants rather than embracing the multifaceted reality of adulteration practices encountered in the real world.

This study introduces a novel machine learning framework for detecting adulteration in camellia oil, suitable for practical scenarios. The specific contributions of this study are as follows:

1. **TES Framework Innovation:** This study proposed the TES framework, a fusion of a transformer encoder and an SVM regressor, specifically engineered for the intricate analysis of E-nose data. This framework distinguishes itself by its adeptness at extracting critical, multi-dimensional features directly from E-nose signals. Its deployment marks a significant step forward, offering robustness against overfitting and exceptional efficiency in handling high-dimensional data spaces.
2. **FTIR Spectroscopy Validation:** In parallel, FTIR spectroscopy serves as an essential verification tool, corroborating the E-nose findings. This dual-analytical strategy not only validates the adulteration detection capability of the E-nose but also enriches the investigation with deeper insights into adulterant identities. It represents a novel approach to ensuring the precision and reliability of the adulteration detection process.
3. **Enhancing Consumer Protection and Industry Practices:** Fundamentally, the research empowers consumer safety and supports the camellia oil sector by introducing an accurate, reliable adulteration detection tool. It can potentially modify quality control practices, rendering the E-nose technology more accessible and practical for widespread industry adoption.

Table 1
Information on the edible oil sample used in this study.

Type	Brand	Producing Area	USD/L	Number of Samples
Camellia oil	Precious Oil	Guangxi Province, China	68	117
Camellia oil	Huang Zhong	Guangxi Province, China	71	117
Corn oil	Arawana	Fujian Province, China	2.65	108
Soybean oil	Gold Ingots	Fujian Province, China	1.6	108
Peanut oil	Lu Hua	Shandong Province, China	5.9	108

Table 2
Adulterated camellia oil samples.

Group	Adulterated with B (A + B)	Adulteration Ratio (V_B/V_{A+B})
C1C	Camellia oil 1 + corn oil	0%, 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%
C2C	Camellia oil 2 + corn oil	0%, 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%
C1S	Camellia oil 1 + soybean oil	0%, 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%
C2S	Camellia oil 2 + soybean oil	0%, 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%
C1P	Camellia oil 1 + peanut oil	0%, 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%
C2P	Camellia oil 2 + peanut oil	0%, 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%

2. Materials and methods

2.1. Edible oil samples

As shown in Table 1, two types of camellia oil mixed with corn, soybean, and peanut oil are detected. A total of 558 edible oils including camellia oil 1 (117 samples, Precious Oil), camellia oil 2 (117 samples, Huang Zhong), corn oil (108 samples, Arawana), soybean oil (108 samples, Gold Ingots), and peanut oil (108 samples, Lu Hua) were provided by the Xiamen Products Quality Supervision and Inspection Institute (http://www.xmzjy.org/), a government inspection agency in China.

Camellia oil 1 and camellia oil 2 were mixed with corn oil, soybean oil, and peanut oil, respectively. The adulterated camellia oil samples were prepared with adulteration ratios ranging from 0 to 60%, in increments of 5% (0%, 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, and 60% (v/v)). Each sample is standardized to a total volume of 5 mL. Details of the adulterated oil samples are listed in Table 2. In total, 702 samples were prepared and analyzed, with each group

Fig. 1. Experimental equipment. (a) PEN3 E-nose, (b) Thermo Nicolet 6700 FTIR.

consisting of 117 measured samples, and each adulteration ratio represented by 9 measured samples.

2.2. E-nose data acquisition

Fig. 1 (a) shows the metal oxide semiconductor (MOS)-based E-nose (PEN3 E-nose, Airsense Analytics GmbH, Germany) with 10 different MOS sensors used to detect the adulterated camellia oil samples. The experiment lasted for 6 days; each sample was measured once and each group was measured daily. Every day, for each group, 6 samples were measured in the morning as training sample and 3 samples were measured in the afternoon as testing sample. Before sample measurement, 3 mL of each sample was placed into a single hermetic vial (20 mL) and airproofed for 3 min.

All experiments in this study were conducted in a single, clean, and well-ventilated testing room of the laboratory with an area of about 45 m^2 under the same conditions (temperature: 24 ± 2 °C; relative humidity: 50 ± 2%). The oil samples were evaluated in a well-ventilated place to reduce baseline fluctuations and prevent interference with other scents. The zero gas, which was used as a baseline in this study, was produced from ambient air within the machine by using two activated charcoal filters. In the PEN3 E-nose workflow, the collection stage starts with a zero-point trim for automatic adjustment and calibration of 5 s, followed by the intake of sample gas at a constant flow velocity of 600 mL/min for a duration of 120 s. During this period, data are recorded at intervals of 1 s. Subsequently, in the flushing stage, the sensor surface is thoroughly cleaned with air filtered via activated carbon, proceeding at a flow rate of 400 mL/min for 30 s.

2.3. FTIR data acquisition

As shown in Fig. 1 (b), camellia oil samples were analyzed using the benchtop FTIR (Thermo Nicolet 6700, Dublin, Ireland) under the same conditions (temperature: 24 ± 2 °C; relative humidity: 50 ± 2%). The device was equipped with attenuated total reflectance (ATR) attachment comprising ZnSe crystal material, DTGS detector, germanium-coated KBr beam-splitter, and high-intensity air-cooled infrared light source. Initially, the background (air) spectrum of the instrument was collected, ensuring that the ATR background energy exceeded 6.3. After calibration, the oil sample was carefully introduced onto the horizontal ATR attachment, ensuring full coverage of the ATR crystal surface for spectrum measurement and the measurement lasted for 10 s. Post-analysis, the ATR crystal was thoroughly cleaned with n-hexane and anhydrous ethanol before another background spectrum was acquired. The process of adding a sample and cleaning requires 1 min. The meticulous procedure allowed for the accurate collection of spectral data, which ranged from 4000 cm^{-1}–650 cm^{-1}, with 32 scans at a resolution of 4 cm^{-1} for each oil sample.

2.4. Datasets

For the E-nose, six datasets (Datasets A, B, C, D, E, and F) were established using the multidimensional signals from the E-nose system. Datasets A and B were used to evaluate the proposed method for predicting the adulteration ratio of the camellia oil adulterated with corn oil. Datasets C and D were employed to assess the effectiveness of the proposed method for predicting the adulteration ratio of the camellia oil adulterated with soybean oil. Datasets E and F were used to evaluate the efficacy of the proposed method in predicting the adulteration ratio of camellia oil adulterated with peanut oil.

Dataset A: This dataset comprised 117 samples (13 adulteration ratios × 9 individual samples) from the C1C group. For each adulterated ratio, six individual samples were used as training samples and three individual samples were used as testing samples.

Dataset B: This dataset comprised 117 samples (13 adulteration ratios × 9 individual samples) from the C2C group. For each adulteration ratio, six individual samples were used as training samples and three individual samples were used as testing samples.

Dataset C: This dataset comprised 117 samples (13 adulteration ratios × 9 individual samples) from the C1S group. For each adulteration ratio, six individual samples were used as training samples and three individual samples were used as testing samples.

Dataset D: This dataset comprised 117 samples (13 adulteration ratios × 9 individual samples) from the C2S group. For each adulteration ratio, six individual samples were used as training samples and three individual samples were used as testing samples.

Dataset E: This dataset comprised 117 samples (13 adulteration ratios × 9 individual samples) from the C1P group. For each adulteration ratio, six individual samples were used as training samples and three individual samples were used as testing samples.

Dataset F: This dataset comprised 117 samples (13 adulteration ratios × 9 individual samples) from the C2P group. For each adulteration ratio, six individual samples were used as training samples and three individual samples were used as testing samples.

The analysis of the FTIR spectroscopy data focused solely on the evaluation of chemical composition. A single sample was selected for each adulteration ratio within each set to serve as a representative measurement. Thus, only chemical analysis was conducted.

3. Proposed method

3.1. Support vector machine

Support vector machines (SVM) represent a class of supervised learning methods that can be used for classification and regression tasks (Cervantes et al., 2020). The core concept is to find an optimal hyperplane that separates the data points of different classes with the maximum margin (Yao et al., 2015). SVM can address linear and nonlinear problems by using different kernel functions, such as the

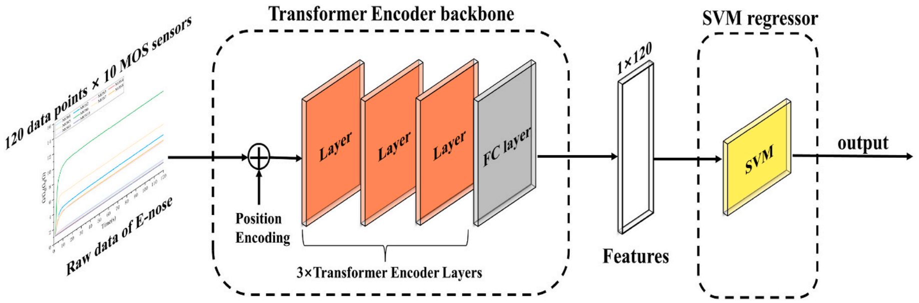


Fig. 2. Flowchart of the TES framework.

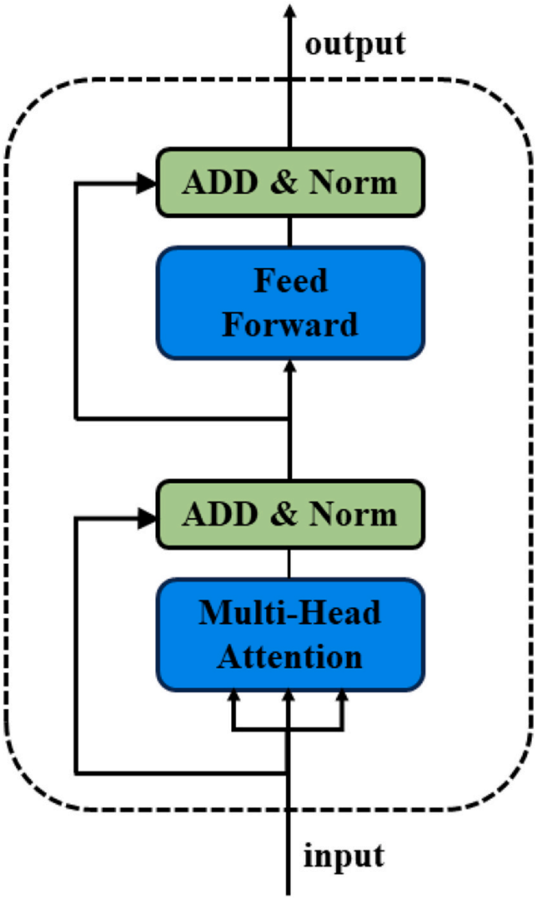


Fig. 3. Schematic diagram of the transformer encoder layer.

linear function, polynomial function, radial basis function, and sigmoid function (Deris et al., 2011). This approach has been widely applied in various fields, including computer vision, natural language processing, and food detection (Saha and Manickavasagan, 2021; Shehab et al., 2022).

3.2. Transformer

The Transformer algorithm, initially introduced by Vaswani et al. (2017), provides an alternative to sequence learning and has shown potential for application not only in natural language processing but in other fields as well (Lin et al., 2022). The model is built on an encoder-decoder architecture, each consisting of multiple layers. Notably, the Transformer eliminates the need for recurrence, which is a staple in traditional models such as recurrent neural networks, thereby facilitating enhanced parallelization (Orken et al., 2022). The architecture incorporates a multi-head self-attention mechanism and a position-wise fully connected feed-forward network in both the encoder and decoder layers. An additional sublayer in the decoder conducts multi-head attention over the encoder's output, allowing the model to use context effectively. A crucial element of the Transformer is its attention mechanism, particularly the scaled dot-product attention, which enables the model to calculate the relevance of different features in the input sequence (Rahardja et al., 2022). The Transformer lacks built-in sequence awareness; thus, positional encodings are incorporated into to convey information about the order of features. These characteristics endow the Transformer model with the ability to capture complex relationships and long-range dependencies between features (Zhang et al., 2022b).

3.3. Proposed TES framework

In this study, a novel machine learning framework was proposed, referred to as a transformer encoder with SVM regression (TES), to analyze and predict the camellia oil data. The TES flowchart is presented in Fig. 2.

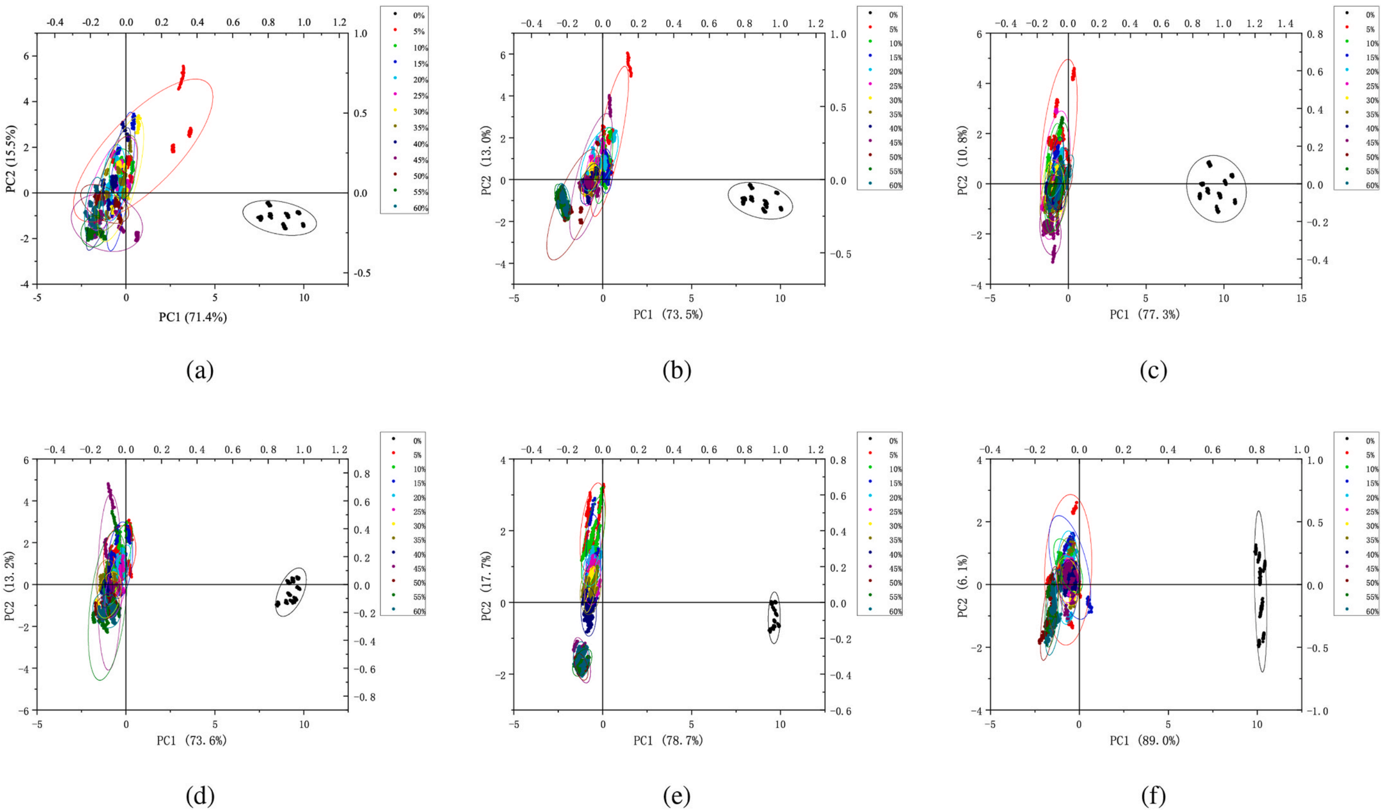
The raw data from a sample consists of a 120 (measurement time of 120 s) × 10 (number of MOS sensors) matrix. As the backbone of the framework, the transformer encoder is primarily responsible for extracting significant features from the input data. The raw E-nose data are directly fed into the transformer encoder backbone without preproccessing. The data, augmented with positional encoding, is processed by a three-layer transformer encoder, which amplifies the capacity of the model to capture expressive features. Each layer within the transformer encoder, detailed in Fig. 3, collaborates to enhance the feature extraction capabilities of the model.

Multi-Head Attention: The transformer encoder uses multi-head attention mechanisms to simultaneously focus on different parts of the input signal. It enables the model to capture intricate patterns and relationships within the data, which are essential for accurate analysis.

Add & Norm: After attention computation, the output undergoes an "Add & Norm" layer. This step incorporates residual connections and normalization, ensuring the stability of activations and hastening convergence during training.

Feed Forward: Each attention output is subsequently conveyed via a feed-forward neural network. The feed-forward network further processes the data and priming it for feature extraction.

After dimensionality reduction via a fully connected layer, the transformer encoder yields 120 features, which are then fed into the SVM regressor for final prediction with an RBF kernel. The transformer encoder can discern intricate structures within the data and capture long-range dependencies. This ability is essential when complex multivariate data are involved. SVM excels in handling small-to-medium-sized, high-dimensional data. It exhibits resilience to overfitting and delivers reliable predictive performance. The application of the SVM regressor on the features obtained from the transformer encoder allows for a comprehensive and nuanced understanding of the data, yielding highly accurate predictions.



(a) (b) (c)

(d) (e) (f)

Fig. 4. PCA score plots depicting the different samples of adulterated camellia oil with PC1 and PC2. (a) camellia oil 1 adulterated with corn oil, (b) camellia oil 1 adulterated with soybean oil, (c) camellia oil 1 adulterated with peanut oil, (d) camellia oil 2 adulterated with corn oil, (e) camellia oil 2 adulterated with soybean oil, and (f) camellia oil 2 adulterated with peanut oil.

4. Result and discussion

4.1. *E-nose analysis*

4.1.1. *Principal component analysis*

Principal component analysis (PCA) is a technique designed to maximize variance through the generation of uncorrelated variables, thereby reducing the dimensionality of datasets while enhancing interpretability and minimizing information loss (Lever et al., 2017). It offers several advantages, such as simplifying the data structure, enhancing data visualization, extracting the most relevant features, as well as removing noise and redundancy (Jolliffe and Cadima, 2016; Abid et al., 2018). Nonetheless, PCA has limitations. It exhibits sensitivity to outliers and scaling issues, poses the risk of information loss during dimensionality reduction, and operates under assumptions of linearity and normality in the dataset (Parsons et al., 2009; Chao et al., 2018). Thus, careful use of PCA is recommended. When appropriate, complementing PCA with alternative analytical techniques may be advantageous (Zou et al., 2006).

In this study, PCA was applied to the six datasets (Datasets A-F), which were processed using the OriginPro 2023 software. The measurement phase lasted 120 s, and the response value of each sensor was stabilized after 80 s, consequently, the final 20 response points were selected as the input features for PCA. As shown in Fig. 4 (a)-(f), the x-axis and y-axis represent principal component 1 (PC1) and principal component 2 (PC2), respectively. The two-dimensional (2D) PCA projections reveal that the variance explained by the first principal component (PC1) ranges from 71.4% to 88.9%, whereas that by the second principal component (PC2) ranges from 6.1% to 17.7%. The cumulative variance of PC1 and PC2 exceeds 85%, indicating that the two principal components contained sufficient sample information. The score plots exhibit a distinct yet limited separation between pure and adulterated camellia oil samples, with a considerable degree of overlap among samples with varying adulteration ratios. Despite the qualitative separation achieved by PCA, the overlapping clusters for varying adulteration levels indicate that PCA is not a suitable method for quantitative analysis. This shortcoming is critical, given the importance of quantitative assessment for both regulatory compliance and consumer safety.

Table 3
Comparative performances of the SVM, Transformer Encoder, and TES Models in detecting adulteration ratios in camellia oil across multiple datasets.

Model	Evaluation Metrics	Dataset A	Dataset B	Dataset C	Dataset D	Dataset E	Dataset F
SVM	RMSE	0.1495	0.0678	0.0787	0.0800	0.1443	0.1491
	MAE	0.1368	0.0561	0.0683	0.0674	0.1131	0.1126
	R^2	0.3611	0.6877	0.8231	0.8171	0.4048	0.3645
Transformer Encoder	RMSE	0.1126	0.1051	0.0863	0.0792	0.1074	0.1215
	MAE	0.0906	0.0780	0.0680	0.0602	0.0909	0.0906
	R^2	0.6377	0.6845	0.7873	0.8480	0.6702	0.5528
TES	RMSE	**0.0387**	**0.0458**	**0.0458**	**0.0458**	**0.0520**	**0.0583**
	MAE	**0.0286**	**0.0374**	**0.0393**	**0.0374**	**0.0374**	**0.0507**
	R^2	**0.9580**	**0.9394**	**0.9402**	**0.9400**	**0.9239**	**0.9019**

PCA can qualitatively differentiate between pure and adulterated camellia oil only limitedly. With this restriction considered, more advanced analytical methods are needed. Machine learning algorithms can provide a nuanced understanding that PCA fails to offer. These algorithms can distinguish between various levels of adulteration; in addition, they can also potentially provide quantitative measures of adulteration ratios, presenting a more comprehensive solution to the complex issue of food adulteration.

4.1.2. Experiment I: ablation experiment

The comparative results of the ablation experiments on six datasets are listed in Table 3. The standalone SVM model, characterized by its use of a kernel-based learning algorithm, exhibited higher RMSE and MAE values in Datasets A, E, and F. These discrepancies suggesting deviation from the actual values, could be attributed to the inherent limitation of the SVM in capturing complex, non-linear relationships within the high-dimensional data derived from E-nose measurements. With its self-attention mechanism, the Transformer Encoder yielded improved results over SVM, particularly in its adept handling of the sequential and temporal aspects of the dataset. Notably, the TES model integrates the feature extraction capabilities of the Transformer Encoder with the regression prowess of the SVM, demonstrating a significant enhancement in performance. This hybrid model achieves substantial reductions in MAE and RMSE values across all datasets, indicating more precise detection of adulteration ratios. The ablation study clearly illustrates how the fusion of these two components—the sequential pattern recognition by the Transformer Encoder and the precise regression modeling by the SVM—contributes to the overall effectiveness of the TES model. The superior performance of TES can be attributed to the specific structural synergy between the transformer encoder and the SVM.

Essentially, the sequential data from the E-nose encapsulates a temporal narrative of the volatile organic compound (VOC) profile of the sample, which can strongly indicate adulteration ratios. The attention mechanisms of the Transformer Encoder are particularly well-suited to interpret this narrative, allowing the model to focus on the most salient features indicative of signal adulteration. Its ability to assign different weights to different parts of the sequence ensures that even subtle changes in the VOCs of the oil, potentially correlating with adulteration, are effectively captured and emphasized.

After feature extraction, the SVM component assumes responsibility for performing regression on these high-dimensional, nuanced features extracted by the Transformer Encoder. The strength of the SVM in this hybrid model lies in its ability to efficiently handle the high-dimensional space, employing a kernel trick to transform the data and find an optimal boundary between classes. The robustness of the SVM to overfitting and its efficacy in high-dimensional spaces complement the feature extraction of the Transformer Encoder. These characteristics offer a double-layered defense against both underfitting and overfitting, ensuring that the model remains generalizable and accurate.

Table 4
Performance metrics for detecting corn oil adulteration in camellia oil using various machine learning models.

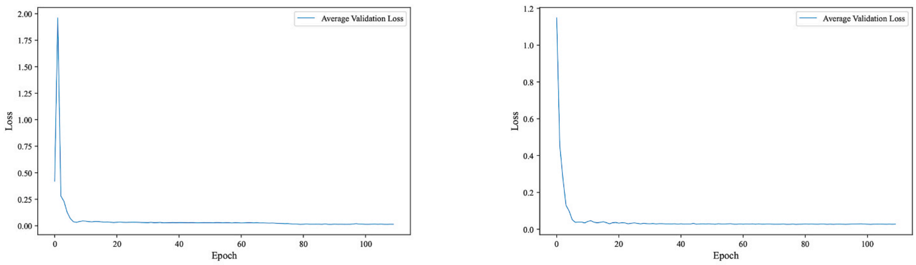
Model	Evaluation Metrics	Dataset A	Runtime on Dataset A (s)	Dataset B	Runtime on Dataset B (s)
RF	RMSE	0.1643	0.39	0.1080	0.41
	MAE	0.1294		0.0822	
	R^2	0.2284		0.6669	
XGBoost	RMSE	0.1759	0.15	0.0981	0.20
	MAE	0.1386		0.0793	
	R^2	0.1158		0.7248	
KNN	RMSE	0.1502	0.06	0.0934	0.07
	MAE	0.1037		0.0642	
	R^2	0.3556		0.7509	
BPNN	RMSE	0.1429	31.99	0.0888	30.27
	MAE	0.1171		0.0756	
	R^2	0.4162		0.7747	
TES	RMSE	0.0387	4.24	0.0458	4.15
	MAE	0.0286		0.0374	
	R^2	0.9580		0.9394	

The comparative results of the ablation experiments on six datasets are listed in Table 3. A thorough evaluation of all models was conducted via 3-fold cross-validation to minimize overfitting and enhance the reliability of the results. The final 20 response points were selected as the input features for the SVM. The SVM model, equipped with a nonlinear kernel, is adept at creating nonlinear hyperplanes for feature separation. However, when confronted with manually extracted features that exhibit a high degree of confusion, SVM alone may have difficulty representing and discriminating these features effectively because of their complexity. By contrast, the Transformer Encoder excels at modeling the relational information inherent in sequential data, enabling the model to capture the nuanced characteristics of VOC encountered during the measurement process. Nonetheless, employing a Transformer Encoder followed solely by a fully connected (FC) layer falls short of ideal, as the linear nature of FC layers does not fully exploit the complex patterns identified by the Transformer Encoder.

The ablation experiment evaluates the performance of the SVM used independently, the Transformer Encoder coupled with an FC layer, and the integrated TES model that merges the dynamic feature extraction of the Transformer Encoder with the nonlinear regression capabilities of the SVM. This comprehensive approach harnesses ability of the Transformer Encoder to decipher complex, time-sensitive VOC patterns alongside the precision of the SVM in high-dimensional regression. By integrating these two components, the TES model not only captures the complex data patterns revealed by the Transformer Encoder but also leverages the robustness and accuracy of the SVM in regression. This synergy enhances the precision of the model in identifying adulteration ratios in camellia oil, as evidenced by the substantial enhancements in RMSE, MAE, and R^2 metrics relative to those of models used independently.

Table 3 provides a quantitative comparison of the performance of the SVM, Transformer Encoder, and TES models in identifying adulteration ratios in camellia oil across multiple datasets. The SVM model shows RMSE values varying from 0.0678 in Dataset B to 0.1495 in Dataset A, as well as MAE values ranging from 0.0561 in Dataset B to 0.1368 in Dataset A. The variation in performance, together with R^2 values ranging from 0.3611 to 0.8231, indicates the presence of challenges in accurately predicting adulteration ratios in certain datasets. By contrast, the Transformer Encoder shows enhancements in RMSE, MAE, and R^2 values, demonstrating its improved handling of the complexity of sequential data. For instance, in Dataset D, the RMSE and MAE values for the Transformer Encoder are 0.0792 and 0.0602, and those for the SVM are 0.0800 and 0.0674. The R^2 values of 0.6377 in Dataset A and 0.8480 in Dataset D suggest a more reliable model fit. The TES model, integrating the Transformer Encoder and SVM, exhibits significantly superior performance. It consistently achieves the lowest RMSE and MAE values, not exceeding 0.0583 and 0.0507 respectively, while maintaining R^2 values above 0.9019 across all datasets. These findings indicate a considerably strong predictive capability and model fit. The TES model exploits the ability of the Transformer Encoder to capture complex data patterns and the robustness and accuracy of the SVM in regression to deliver superior performance in detecting adulteration ratios in camellia oil. The results further corroborate the impact of model structure on performance, with the architecture of the TES significantly enhancing its predictive accuracy.

In summary, the strength of the TES model lies in its customized approach, effectively combining the advanced feature extraction capabilities of the Transformer Encoder with the stable performance of the SVM in regression. This combination results in a powerful tool for the accurate detection of adulteration ratios, demonstrating superior performance across various datasets and underscoring the robustness and generalizability of the model.



(a) (b)

Fig. 5. The average validation loss of the TES during training of 3-fold cross-validation. (a) Dataset A, (b) Dataset B.

4.1.3. *Experiment II: Camellia oil adulterated with corn oil*

The study assessed the performance of five machine learning models on Datasets A and B, specifically designed to detect varying levels of adulteration in camellia oil with corn oil. The models included Random Forests (RF), XGBoost, K-Nearest Neighbor (KNN), Backpropagation Neural Network (BPNN), and the proposed TES framework. The final 20 response points were selected as the input features for RF, XGBoost, KNN, and BPNN.

The BPNN model comprised two hidden layers and one output layer, utilizing the ReLU activation function. The training and validation process coupled with the Early Stopping callback function for halting the training process when no significant improvement was detected. In addition, the RF, XGBoost, and KNN used default parameter settings. A thorough evaluation of all models was conducted via 3-fold cross-validation to minimize overfitting and enhance the reliability of the results.

In Table 4, Dataset A reveals that the proposed TES model outperforms its counterparts, exhibiting an RMSE of 0.0387, an MAE of 0.0286, and an R^2 of 0.9580, with a runtime of 4.24 s. By contrast, RF, XGBoost, KNN, and BPNN fall short with R^2 values of 0.2284, 0.1158, 0.3556, and 0.4162 and runtimes of 0.39, 0.15, 0.06, and 31.99 s, respectively. Fig. 5 (a) exhibits a rapid decrease in average validation loss and a stable plateau as epochs increase, signifying the TES model's swift convergence and consistent performance.

A similar trend is observed in Dataset B. The TES model demonstrates superior performance, showing an RMSE of 0.0458, MAE of 0.0374, and an R^2 of 0.9394, with a runtime of 4.15 s, respectively. Fig. 5 (b) shows a quick initial drop in loss, followed by a steady low-loss state, which provides evidence of the model's effective learning and generalization capabilities across different datasets. RF shows improvement in R^2 with 0.6669 in 0.41 s but remains inferior in performance to the TES model. XGBoost exhibits a slight advantage, with an R^2 of 0.7248 in 0.20 s. Both KNN and BPNN remain competitive, with R^2 values of 0.7509 and 0.7747 in 0.07 and 30.27 s, respectively.

In both datasets, TES exhibits superior performance, registering not only the lowest RMSE and MAE but the highest R^2 as well, indicating an exceptional fit to the data. The TES model further achieves a commendable balance between accuracy and runtime, rendering it efficient and effective. The limited efficacy of RF, XGBoost, KNN and BPNN, particularly in Dataset A, suggests a potential constraint in their applicability for this specific task.

The exceptional performance of the TES model indicates its high potential for detecting adulteration in camellia oil. This finding carries significant implications for leveraging E-nose technology in food quality and safety applications. Therefore, the TES model emerges as a promising candidate for further research and practical implementation in food quality monitoring.

4.1.4. *Experiment III: Camellia oil adulterated with soybean oil*

The experiment meticulously evaluated the performance of five distinct machine learning algorithms in the context of detecting adulteration in camellia oil. The algorithms were subjected to Datasets C and D. Each dataset represented a distinct adulteration scenario where camellia oil 1 and camellia oil 2 were mixed with soybean oil at varying ratios. A thorough evaluation of all models was conducted via 3-fold cross-validation to minimize overfitting and enhance the reliability of the results.

Table 5
Performance metrics for soybean oil adulteration in camellia oil using various machine learning models.

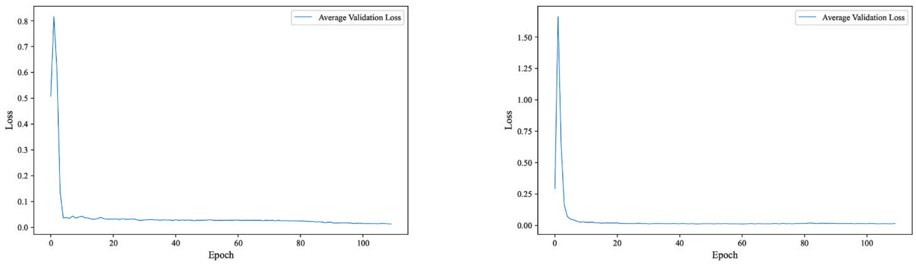
Model	Evaluation Metrics	Dataset C	Runtime on Dataset C (s)	Dataset D	Runtime on Dataset D (s)
RF	RMSE	0.0696		0.0756	
	MAE	0.0456	0.37	0.0530	0.46
	R^2	0.8614		0.8367	
XGBoost	RMSE	0.0791		0.0773	
	MAE	0.0656	0.32	0.0615	0.41
	R^2	0.8214		0.8293	
KNN	RMSE	0.0838		0.0847	
	MAE	0.0718	0.06	0.0595	0.06
	R^2	0.7995		0.7952	
BPNN	RMSE	0.0757		0.0787	
	MAE	0.0603	30.95	0.0586	24.26
	R^2	0.8363		0.8232	
TES	RMSE	0.0458		0.0458	
	MAE	0.0393	4.12	0.0374	4.08
	R^2	0.9402		0.9400	

As shown in Table 5, in Dataset C, the TES model showed unmatched performance, with an RMSE of 0.0458, MAE of 0.0393, and R^2 of 0.9402; runtime was 4.12 s. Fig. 6 (a) displays a pronounced drop in loss during the initial epochs, suggesting that the TES model achieves rapid convergence. Subsequently, the loss stabilizes, implying that the model's parameters have reached an optimal state. The RF, XGBoost, and KNN model achieved R^2 of 0.8614, 0.8214, and 0.7995, respectively, with execution times of 0.37, 0.32, and 0.06 s. BPNN showed an R^2 of 0.8363, but with a notably higher runtime of 30.95 s, which further highlights the efficiency of the TES model.

For Dataset D, the TES model maintained its superior performance, achieving an RMSE of 0.0458, MAE of 0.0374, and R^2 of 0.9400; runtime was 4.08 s. Fig. 6 (b) also demonstrates a rapid initial reduction for the validation loss curve, reaching a low, stable level quickly, which suggests that the TES model shows consistent convergence on both the training and validation sets. The RF and XGBoost model had R^2 of 0.8367 and 0.8293 with runtimes of 0.46 and 0.41 s; meanwhile, KNN achieved an R^2 of 0.7952, finishing in 0.06 s. BPNN yielded R^2 of 0.8232, with considerably longer runtimes of 24.26 s. The consistently high performance of the TES model across both datasets, together with its speed, underscores its robustness and efficiency.

4.1.5. *Experiment IV: Camellia oil adulterated with peanut oil*

Datasets E and F represent adulteration scenarios where different camellia oil samples were blended with peanut oil at various ratios. A thorough evaluation of all models was conducted via 3-fold cross-validation to minimize overfitting and enhance the reliability of the



(a)

(b)

Fig. 6. The average validation loss of the TES during training of 3-fold cross-validation. (a) Dataset C, (b) Dataset D.

Table 6
Performance metrics for peanut oil adulteration in camellia oil using various machine learning models.

Model	Evaluation Metrics	Dataset E	Runtime on Dataset E (s)	Dataset F	Runtime on Dataset F (s)
RF	RMSE	0.0957		0.1203	
	MAE	0.0776	0.77	0.0705	0.83
	R^2	0.7385		0.5867	
XGBoost	RMSE	0.1060		0.1129	
	MAE	0.0783	0.43	0.0664	0.40
	R^2	0.6787		0.6361	
KNN	RMSE	0.1595		0.1742	
	MAE	0.1389	0.06	0.1446	0.07
	R^2	0.2735		0.1332	
BPNN	RMSE	0.1195		0.1227	
	MAE	0.0901	23.5	0.0817	27.72
	R^2	0.5921		0.5699	
TES	RMSE	0.0520		0.0583	
	MAE	0.0374	4.22	0.0507	4.09
	R^2	0.9239		0.9019	

results. As shown in Table 6, Dataset E indicates that the TES model demonstrates superior performance, boasting an RMSE of 0.0520, an MAE of 0.0374, and an R^2 value of 0.9239, with a runtime of 4.22 s. The performance surpasses that of RF, XGBoost, and BPNN with R^2 values of 0.7385, 0.6787, and 0.5921, respectively; runtimes are 0.77, 0.43, and 23.5 s. KNN shows the least favorable results in terms of RMSE, MAE, and R^2.

Meanwhile, Dataset F shows that TES maintains its superior performance, achieving an RMSE of 0.0583, an MAE of 0.0507, and an R^2 of 0.9019, with a runtime of 4.09 s. RF, XGBoost, and BPNN demonstrate inferior performance, with R^2 values of 0.5867, 0.6361, and 0.5699, respectively; runtimes were 0.83, 0.40, and 27.72 s. KNN exhibits inferior performance with an R^2 of 0.1332.

Across both datasets, the consistently superior performance of the TES model verifies to its reliability for this specific task. The model delivered accurate results and performed efficiently. The average validation loss curves for Dataset E (shown in Fig. 7 (a)) and Dataset F (shown in Fig. 7 (b)) validate the reliability of the TES model's performance, reflecting its consistent and effective training process across varying datasets. RF, XGBoost, and BPNN exhibited inconsistent results, and their performances were consistently overshadowed by the TES model. KNN showed the poorest performance in both datasets, suggesting that it is not appropriate for this task scenario.

The consistent success of the TES model across diverse datasets and adulteration scenarios underscores its potential for broad applications in food safety. Thus, this study reaffirms the pivotal role of machine learning, particularly the effectiveness of the TES model, in harnessing E-nose technology for detecting camellia oil adulteration. The inclusion of runtime metrics further emphasizes the practicality and efficiency of these models in real-world scenarios.

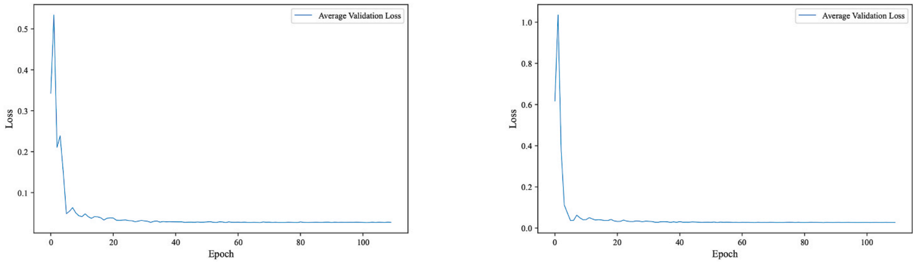
4.2. FTIR analysis

4.2.1. FTIR spectra of adulterated oils

The E-nose, coupled with the TES framework, demonstrates remarkable effectiveness in predicting the adulteration ratios of camellia oil. The synergy between the sensor array and computational modeling yields high accuracy and precision, supporting its applicability in various real-world scenarios. The use of FTIR as a complementary analytical measure to validate the accuracy of the E-nose aligns with existing literature, where some studies (He and Lei, 2020; Han et al., 2020; Meng et al., 2023) have successfully employed FTIR spectroscopy for the detection of adulteration in edible oils, indicating its established role in authenticity assessment.

Fig. 8 (a)-(c) present the FTIR spectra of camellia oil 1 adulterated with corn oil, soybean oil, and peanut oil. Fig. 8 (d)–(f) show the FTIR spectra of camellia oil 2 adulterated with corn oil, soybean oil, and peanut oil. The FTIR spectra reveal consistent patterns across different adulteration levels for both types of camellia oil. This finding suggests the sufficient sensitivity of FTIR spectroscopy for detecting even low levels of adulteration, as discerned from the differences in absorbance values at different wavenumbers.

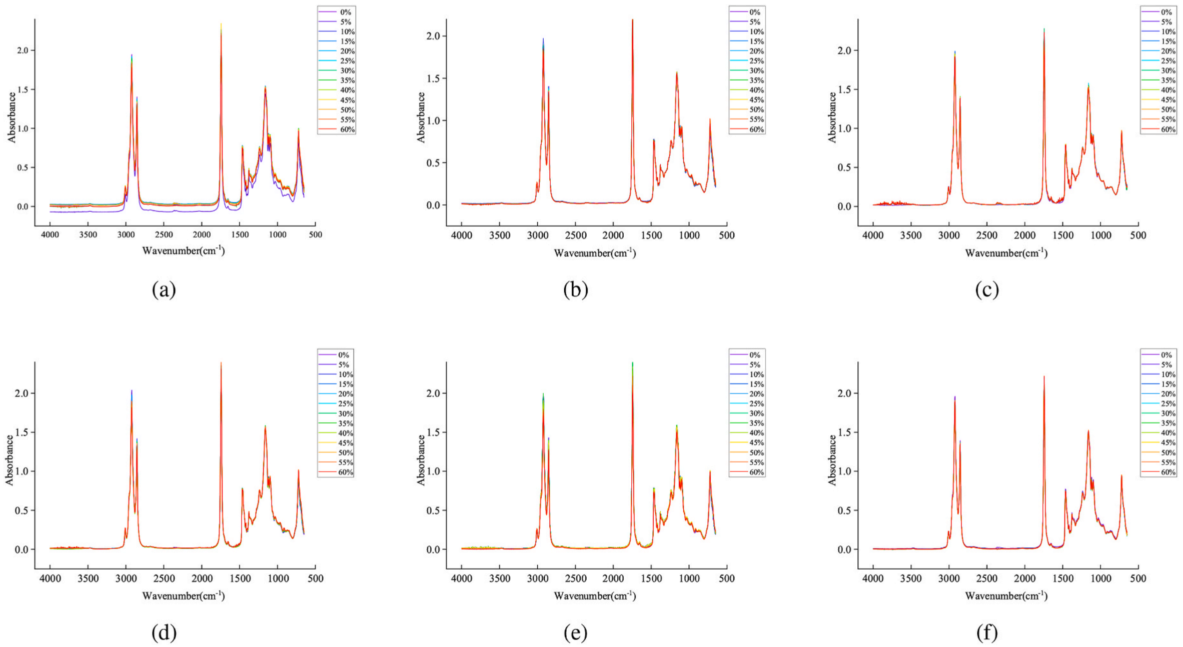
Distinct differences in spectra were observed when camellia oil samples were adulterated with different types of edible oils. For instance, the spectral lines for camellia oil adulterated with corn oil exhibited absorbance peaks different from those shown by camellia oil



(a)

(b)

Fig. 7. The average validation loss of the TES during training of 3-fold cross-validation. (a) Dataset E, (b) Dataset F.



(a)

(b)

(c)

(d)

(e)

(f)

Fig. 8. FTIR spectra of adulterated oils. (a) camellia oil 1 adulterated with corn oil, (b) camellia oil 1 adulterated with soybean oil, (c) camellia oil 1 adulterated with peanut oil, (d) camellia oil 2 adulterated with corn oil, (e) camellia oil 2 adulterated with soybean oil, and (f) camellia oil 2 adulterated with peanut oil.

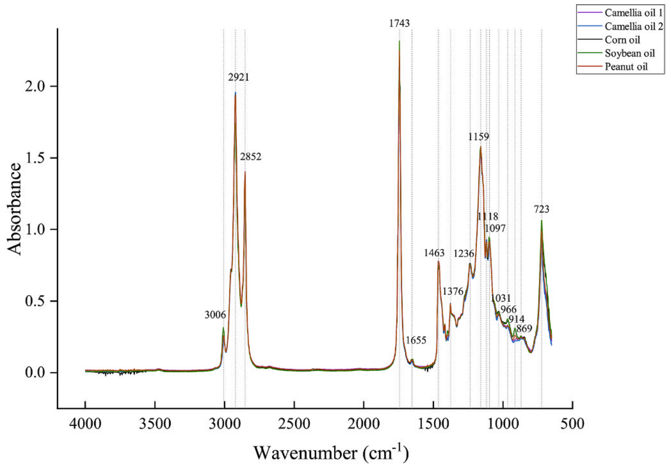


Fig. 9. FTIR spectra for the pure edible oils spanning the entire 4000 to 650 cm^{-1} region.

adulterated with soybean or peanut oils. This difference suggests the suitability of the FTIR method could be used not only for detecting the presence of adulterants but also for identifying the type of adulterant.

The results indicate the viability of using FTIR spectroscopy as a rapid, non-destructive method for quality control in the camellia oil industry. While the FTIR method demonstrates considerable accuracy in detecting adulteration in camellia oils, it is not without limitations. Compared with E-nose technology, FTIR spectroscopy requires sample preprocessing, and its operation necessitates the expertise of professionally trained personnel. In addition, the equipment is relatively expensive. Thus, FTIR spectroscopy may not be suitable for large-scale rapid preliminary screening in routine market surveillance and management.

4.2.2. FTIR spectra of pure edible oils

The FTIR spectra of the five types of pure edible oil (camellia oil 1, camellia oil 2, corn oil, soybean oil, and peanut oil) in the mid-infrared region of 4000 cm^{-1} to 650 cm^{-1} are shown in Fig. 9. Almost all edible oils consist of triacylglycerol (92%), low concentrations of di- and mono-acylglycerols (5%), and low levels of other components (Jamwal et al., 2020). Thus, the spectra of these oil samples exhibit numerous similarities among absorbance bands in this study. The spectra for all oil samples exhibit distinct peaks at wavenumbers 3006, 2921, 2852, 1743, 1655, 1463, 1376, 1236, 1159, 1118, 1097, 1031, 966, 914, 869, and 723 cm^{-1}, indicating the presence of various specific functional groups. The results align with the findings of Han et al. who used FTIR spectroscopy to authenticate and quantify camellia oil adulteration, further validating our results and demonstrating the robustness of FTIR spectroscopy in identifying specific functional groups across different types of edible oils (Han et al., 2020). However, the subtle variations in peak intensities at other wavenumbers suggest the distinct compositional

Table 7
Functional groups in the edible oils and effect on the FTIR spectrum: wavenumber, functional group, and mode of vibration.

Wavenumbers (cm^{-1})	Functional Group	Mode of Vibration
3006	= C–H(cis)	Stretching in Unsaturated Fatty Acids
2921	−C–H(CH_2)	Asymmetric Stretching in Saturated Fatty Acids
2852	−C–H(CH_2)	Symmetric Stretching in Saturated Fatty Acids
1743	−C=O	Stretching in Esters
1655	−C=C−	Stretching in Unsaturated Fatty Acids
1463	−C–H(CH_2)	Bending in Saturated Fatty Acids
1417	O–H	Asymmetric Stretching in Carboxyl Groups (possible from free fatty acids or moisture)
1376	−C–H(CH_3)	Bending in Saturated Fatty Acids
1236	C–O–C	Stretching in Esters (e.g., Triglycerides)
1159	C–O–C	Stretching in Esters (e.g., Triglycerides) or Phospholipids
1118	−C–O	Stretching in Alcohols or Esters
1097	−C–O	Stretching in Alcohols or Esters
1031	C–O–C	Stretching in Esters
966	= C–H	Stretching in Unsaturated Fatty Acids
914	= C–H	Stretching in Unsaturated Fatty Acids
869	= C–H	Bending in Unsaturated Fatty Acids
723	−HC=CH−	Wagging or Twisting in Saturated Fatty Acids

Table 8
Major fatty acid composition of pure edible oil samples analyzed using gas chromatography.

Fatty Acids	Common Names	Camellia oil 1	Camellia oil 2	Corn oil	Soybean oil	Peanut oil
C14:0	Myristic Acid	ND	ND	0.0800	ND	ND
C16:0	Palmitic Acid	7.6200	7.5500	12.7000	11.1300	11.0600
C16:1	Palmitoleic Acid	0.1000	0.0900	0.1000	0.1100	ND
C18:0	Stearic Acid	3.8900	4.2200	3.4700	5.8100	4.2700
C18:1	Oleic Acid	75.0800	74.2500	24.6900	19.6800	53.1800
C18:2	Linoleic Acid	11.1700	10.9700	56.2900	54.7600	22.1800
C18:3	Alpha-Linolenic Acid	0.8500	0.8900	0.5500	5.4800	0.2500
C20:0	Arachidic Acid	0.2100	0.3000	0.9400	0.5800	1.2700
C20:1	Gadoleic Acid	ND	ND	0.3200	0.4200	0.9600

ND represents not detected.

intricacies of each oil type. For example, the peaks related to ester groups (1743 cm^{-1}) and aliphatic hydrocarbons (966 cm^{-1}, 914 cm^{-1}) seem relatively more pronounced in corn oil and soybean oil, suggesting the higher concentrations of these components. The details of functional groups responsible for FTIR absorption bands are listed in Table 7.

Gas chromatography (GC) is a routine method used for fatty acid profiling. GC analysis results revealed that the fatty acid profiles of these oils fell within the range defined by Chinese national standards. As shown in Table 8, camellia oil 1 and camellia oil 2 exhibited pronounced concentrations of oleic acid (C18:1), rendering them distinct from corn, soybean, and peanut oils. This finding was corroborated by a markedly high relative density, identifying oleic acid as the predominant fatty acid in camellia oils. The similarity in fatty acid profiles, particularly the high concentration of oleic acid, not only indicates the quality of camellia oils but also echoes findings from studies on other high-quality oils, such as olive oil. Similar to findings from this study, research by Nuon et el. has also revealed that the fatty acid composition of olive oils emphasizes a universal characteristic among high-quality edible oils——that is, the presence of distinct yet beneficial, fatty acid profiles (Rodrigues et al., 2021). By contrast, corn and soybean oils exhibited more variations in their fatty acid profiles, except for oleic acid, which showed low relative densities. The high relative density of oleic acid in camellia oils may imply a health-promoting profile, congruent with the recognized cardiovascular benefits of unsaturated fatty acids. Both samples of camellia oil samples exhibited a high degree of consistency in their fatty acid profiles, strengthening the reliability of the source. Moreover, variations in fatty acid profiles led to a discrepancy in the FTIR spectra, which could contribute to the classification of edible oil types and the detection of adulteration levels.

In summary, the FTIR spectra offer valuable insights into the compositional and structural nuances among different oil samples, serving as a foundational step for further targeted analyses and applications in food science and technology.

5. Conclusions

This study conducted a comprehensive investigation into the identification of different oils at varying adulteration ratios in camellia oil, leveraging the capabilities of E-nose technology and FTIR spectroscopy. PCA failed to efficiently analyze the E-nose data for distinguishing different classes within a dataset, emphasizing the significance of employing machine learning algorithms for accurate adulteration detection. Various machine learning models were used to evaluate their efficacy in detecting adulteration ratios.

The proposed TES model consistently outperformed other models across multiple datasets. The efficacy of the TES model in discerning adulteration ratios across varied datasets is attributed to its composite structure, which adeptly combining the sequential data processing strength of the Transformer Encoder with the sophisticated regression functionality of the SVM. The Transformer Encoder excels in parsing the intricate, time-sensitive patterns inherent in E-nose data, reflecting the presence of adulterants. Its self-attention mechanism analytically accentuates salient features, providing a detailed VOC profile. Simultaneously, the SVM, with its nonlinear characteristic, capably navigates the high-dimensional feature space, contributing to the robust performance of the model in predicting adulteration ratios. The combined effect of these two components within the TES model leads to high R^2 and low RMSE and MAE. The TES model is not only precise in detecting adulteration ratios but is also versatile and reliable across various adulteration scenarios, endorsing its application for quality control in the camellia oil industry.

The application of FTIR spectroscopy in this study emphasizes its precision in identifying adulteration in camellia oils, exhibiting the sensitivity to discern subtle differences in spectral patterns attributable to variations in adulterants. The ability of the FTIR technique to discern these differences across a range of adulteration ratios suggests its suitability for quality control and assurance in the camellia oil industry. Moreover, the distinctive spectral lines provide a basis not only for detecting the presence of adulterants but also for potentially identifying the types of oils used in adulteration. Nonetheless, the application of FTIR spectroscopy is limited by practical concerns in real-world scenarios. The requirement for detailed sample preparation, the need for trained personnel, and the high costs associated with FTIR equipment significantly impede its adoption for large-scale, rapid screening processes. In comparison, the E-nose exhibits less stringent operational demands and lower cost, emerging as a more suitable option for routine, high-throughput screening. The E-nose is particularly advantageous in settings requiring rapid and efficient quality control measures.

Characterized by operational simplicity and cost-effectiveness, the E-nose provides an effective tool for mass screening in real-world scenarios. FTIR spectroscopy excels in detailed analysis; meanwhile, the E-nose exhibits superior adaptability to large-scale preliminary screening, demonstrating the agility required for dynamic industry demands. By demonstrating the efficacy of E-nose technology coupled with advanced machine learning models, this research contributes to the broader goal of ensuring food safety and authenticity, thereby protecting consumer interests and supporting the continued development of the camellia oil industry.

Declaration of competing interest

We declare that we have no financial and personal relationships with other people or organizations that can inappropriately influence our work, there is no professional or other personal interest of any nature or kind in any product, service and/or company that could be construed as influencing the position presented in, or the review of, the manuscript entitled, “Rapid quantitative authentication and analysis of camellia oil adulterated with edible oils by electronic nose and FTIR spectroscopy”.

Data availability

The authors do not have permission to share data.

Acknowledgements

This work was supported by the Natural Science Foundation of Guangdong Province (Grant No.2021A1515011970) and the National Natural Science Foundation of China (Grant No.62276174).

References

Abid, A., Zhang, M.J., Bagaria, V.K., Zou, J., 2018. Exploring patterns enriched in a dataset with contrastive principal component analysis. Nat. Commun. 9, 2134.

Bunaciu, A.A., Vu, D.H., Aboul-Enein, H.Y., 2023. Edible oil discrimination by fourier transform infrared (ftir) spectroscopy and chemometrics. Anal. Lett. 1–11.

Cao, J., Jiang, X., Chen, Q., Zhang, H., Sun, H., Zhang, W.-M., Li, C., 2020. Oxidative stabilities of olive and camellia oils: possible mechanism of aldehydes formation in oleic acid triglyceride at high temperature. Lebensm. Wiss. Technol. 118, 108858.

Cervantes, J., Garcia-Lamont, F., Rodríguez-Mazahua, L., Lopez, A., 2020. A comprehensive survey on support vector machine classification: applications, challenges and trends. Neurocomputing 408, 189–215.

Chao, Y.-S., Wu, H.-C., Wu, C.-J., Chen, W.-C., 2018. Principal component approximation and interpretation in health survey and biobank data. Front. Digital Humanities 5, 11.

Deris, A.M., Zain, A.M., Sallehuddin, R., 2011. Overview of support vector machine in modeling machining performances. Procedia Eng. 24, 308–312.

Guo, L., Guo, Y., Wu, P., Lu, F., Zhu, J., Ma, H., Chen, Y., Zhang, T., 2020. Camellia oil lowering blood pressure in spontaneously hypertension rats. J. Funct.Foods 70, 103915.

Han, J., Sun, R., Zeng, X., Zhang, J., Xing, R., Sun, C., Chen, Y., 2020. Rapid classification and quantification of camellia (camellia oleifera abel.) oil blended with rapeseed oil using ftir-atr spectroscopy. Molecules 25, 2036.

He, J., Wu, X., Yu, Z., 2021. Microwave pretreatment of camellia (camellia oleifera abel.) seeds: effect on oil flavor. Food Chem. 364, 130388.

He, W., Lei, T., 2020. Identification of camellia oil using ft-ir spectroscopy and chemometrics based on both isolated unsaponifiables and vegetable oils. Spectrochim. Acta Mol. Biomol. Spectrosc. 228, 117839.

Huang, T., Jiang, J., Cao, Y., Huang, J., Zhang, F., Cui, G., 2023. Camellia oil (camellia oleifera abel.) treatment improves high-fat diet-induced atherosclerosis in apolipoprotein e (apoe)-/- mice. Biosci. Microbiota, Food Health 42, 56–64.

Jamwal, R., Kumari, S., Balan, B., Dhaulaniya, A.S., Kelly, S., Cannavan, A., Singh, D.K., et al., 2020. Attenuated total reflectance–fourier transform infrared (atr–ftir) spectroscopy coupled with chemometrics for rapid detection of argemone oil adulteration in mustard oil. Lebensm. Wiss. Technol. 120, 108945.

Jamwal, R., Kumari, S., Sharma, S., Kelly, S., Cannavan, A., Singh, D.K., et al., 2021. Recent trends in the use of ftir spectroscopy integrated with chemometrics for the detection of edible oil adulteration. Vib. Spectrosc. 113, 103222.

Jiménez-Carvelo, A.M., Osorio, M.T., Koidis, A., González-Casado, A., Cuadros-Rodríguez, L., 2017. Chemometric classification and quantification of olive oil in blends with any edible vegetable oils using ftir-atr and Raman spectroscopy. Lebensm. Wiss. Technol. 86, 174–184.

Jolliffe, I.T., Cadima, J., 2016. Principal component analysis: a review and recent developments. Phil. Trans. Math. Phys. Eng. Sci. 374, 20150202.

Karami, H., Rasekh, M., Mirzaee-Ghaleh, E., 2020. Comparison of chemometrics and aocs official methods for predicting the shelf life of edible oil. Chemometr. Intell. Lab. Syst. 206, 104165.

Lever, J., Krzywinski, M., Altman, N., 2017. Points of significance: principal component analysis. Nat. Methods 14, 641–643.

Li, Z., Liu, A., Du, Q., Zhu, W., Liu, H., Naeem, A., Guan, Y., Chen, L., Ming, L., 2022. Bioactive substances and therapeutic potential of camellia oil: an overview. Food Biosci. 49, 101855.

Lin, T., Wang, Y., Liu, X., Qiu, X., 2022. A Survey of Transformers. AI Open.

Majchrzak, T., Wojnowski, W., Dymerski, T., Gębicki, J., Namieśnik, J., 2018. Electronic noses in classification and quality control of edible oils: a review. Food Chem. 246, 192–201.

Meng, X., Yin, C., Yuan, L., Zhang, Y., Ju, Y., Xin, K., Chen, W., Lv, K., Hu, L., 2023. Rapid detection of adulteration of olive oil with soybean oil combined with chemometrics by fourier transform infrared, visible-near-infrared and excitation-emission matrix fluorescence spectroscopy: a comparative study. Food Chem. 405, 134828.

Njoman, M.F., Nugroho, G., Chandra, S.D.P., Permana, Y., Suhadi, S., Mujiono, M., Hermawan, A.D., Sugiono, S., 2017. The vulnerability of human sensory evaluation and the promising senses instrumentation. Br. Food J. 119, 2145–2160.

Orken, M., Dina, O., Keylan, A., Tolganay, T., Mohamed, O., 2022. A study of transformer-based end-to-end speech recognition system for Kazakh language. Sci. Rep. 12, 8337.

Parsons, K.J., Cooper, W.J., Albertson, R.C., 2009. Limits of principal components analysis for producing a common trait space: implications for inferring selection, contingency, and chance in evolution. PLoS One 4, e7957.

Rahardja, S., Wang, M., Nguyen, B.P., Fränti, P., Rahardja, S., 2022. A lightweight classification of adaptor proteins using transformer networks. BMC Bioinf. 23, 1–14.

Rodrigues, N., Casal, S., Pinho, T., Cruz, R., Peres, A.M., Baptista, P., Pereira, J.A., 2021. Fatty acid composition from olive oils of Portuguese centenarian trees is highly dependent on olive cultivar and crop year. Foods 10, 496.

Saha, D., Manickavasagan, A., 2021. Machine learning techniques for analysis of hyperspectral images to determine quality of food products: a review. Curr. Res. Food Sci. 4, 28–44.

Shehab, M., Abualigah, L., Shambour, Q., Abu-Hashem, M.A., Shambour, M.K.Y., Alsalibi, A.I., Gandomi, A.H., 2022. Machine learning in medical applications: a review of state-of-the-art methods. Comput. Biol. Med. 145, 105458.

Shi, T., Wu, G., Jin, Q., Wang, X., 2020. Camellia oil authentication: a comparative analysis and recent analytical techniques developed for its assessment. a review. Trends Food Sci. Technol. 97, 88–99.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I., 2017. Attention is all you need. Adv. Neural Inf. Process. Syst. 30.

Wei, X., Shao, X., Wei, Y., Cheong, L., Pan, L., Tu, K., 2018. Rapid detection of adulterated peony seed oil by electronic nose. J. Food Sci. Technol. 55, 2152–2159.

Windarsih, A., Indrianingsih, A.W., Apriyana, W., Rohman, A., 2023. Rapid detection of pork oil adulteration in snakehead fish oil using ftir-atr spectroscopy and chemometrics for halal authentication. Chem. Pap. 1–12.

Xu, L., Yu, X., Liu, L., Zhang, R., 2016. A novel method for qualitative analysis of edible oil oxidation using an electronic nose. Food Chem. 202, 229–235.

Yao, Y., Cui, H., Liu, Y., Li, L., Zhang, L., Chen, X., et al., 2015. Pmsvm: an optimized support vector machine classification algorithm based on pca and multilevel grid search methods. Math. Probl Eng. 2015.

Ye, Q., Meng, X., 2022. Highly efficient authentication of edible oils by ftir spectroscopy coupled with chemometrics. Food Chem. 385, 132661.

Zhang, F., Zhu, F., Chen, B., Su, E., Chen, Y., Cao, F., 2022a. Composition, bioactive substances, extraction technologies and the influences on characteristics of camellia oleifera oil: a review. Food Res. Int. 156, 111159.

Zhang, L., Hong, X., Arandjelović, O., Zhao, G., 2022b. Short and long range relation based spatio-temporal transformer for micro-expression recognition. IEEE Transact. Affect. Comput. 13, 1973–1985.

Zou, H., Hastie, T., Tibshirani, R., 2006. Sparse principal component analysis. J. Comput. Graph Stat. 15, 265–286.