



Research article

Rapid detection and quantification of melamine, urea, sucrose, water, and milk powder adulteration in pasteurized milk using Fourier transform infrared (FTIR) spectroscopy coupled with modern statistical machine learning algorithms

Chu Chu^{a,1}, Haitong Wang^{a,1}, Xuelu Luo^{a,1}, Yikai Fan^a, Liangkang Nan^a, Chao Du^a, Dengying Gao^a, Peipei Wen^a, Dongwei Wang^a, Zhuo Yang^a, Guochang Yang^a, Li Liu^a, Yongqing Li^a, Bo Hu^b, Abula Zunongjiang^b, Shujun Zhang^{a,*}

^a Key Lab of Agricultural Animal Genetics, Breeding and Reproduction of Ministry of Education, Huazhong Agricultural University, Wuhan, 430070, China

^b Quality Standards Institute of Animal Husbandry, Xinjiang Academy of Animal Science, Urumqi, Xinjiang, 830012, China

ARTICLE INFO

Keywords:

Pasteurized milk

Fourier transform infrared (FTIR) spectroscopy

Modern statistical machine learning algorithms

Adulteration

ABSTRACT

There is an evident requirement for a rapid, efficient, and simple method to screen the authenticity of milk products in the market. Fourier transform infrared (FTIR) spectroscopy stands out as a promising solution. This work employed FTIR spectroscopy and modern statistical machine learning algorithms for the identification and quantification of pasteurized milk adulteration. Comparative results demonstrate modern statistical machine learning algorithms will improve the ability of FTIR spectroscopy to predict milk adulteration compared to partial least square (PLS). To discern the types of substances utilized in milk adulteration, a top-performing multi-classification model was established using multi-layer perceptron (MLP) algorithm, delivering an impressive prediction accuracy of 97.4 %. For quantification purposes, bayesian regularized neural networks (BRNN) provided the best results for the determination of both melamine, urea and milk powder adulteration, while extreme gradient boosting (XGB) and projection pursuit regression (PPR) gave better results in predicting sucrose and water adulteration levels, respectively. The regression models provided suitable predictive accuracy with the ratio of performance to deviation (RPD) values higher than 3. The proposed methodology proved to be a cost-effective and fast tool for screening the authenticity of pasteurized milk in the market.

1. Introduction

Pasteurization is a milk preservation method primarily reliant on heat treatment to eliminate or deactivate potentially harmful microorganisms [1]. The process yields pasteurized milk, a product renowned for its rich nutritional content and its delightful, fresh,

* Corresponding author.

E-mail address: sjxiaozhang@mail.hzau.edu.cn (S. Zhang).

¹ These authors contributed equally: Chu Chu, Haitong Wang, Xuelu Luo.

<https://doi.org/10.1016/j.heliyon.2024.e32720>

Received 17 April 2024; Received in revised form 7 June 2024; Accepted 7 June 2024

Available online 8 June 2024

2405-8440/© 2024 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC license (<http://creativecommons.org/licenses/by-nc/4.0/>).

and unadulterated flavor. It stands as the foremost choice for liquid milk consumption in many developed nations boasting thriving dairy industries. However, it is also a target for illicit tampering by unscrupulous milk vendors and traders.

Water is a frequently encountered adulterant in milk, serving to increase its volume but at the cost of diluting its nutrient content. Consequently, unscrupulous traders often resort to the addition of substances like melamine, urea, sucrose, and milk powder to mask the resultant loss of nutrients, density, and sweetness [2]. Melamine is a nitrogen-rich toxic compound frequently employed to artificially boost the apparent protein content of milk [3]. Excessive melamine consumption can result in severe health consequences, including renal failure and even fatality [4]. The maximum residue limits (MRL) for melamine residues in milk fluctuate, ranging from 0.05 mg/kg to 2.5 mg/kg, contingent upon local regulations. For instance, the European Union (EU), the European Food Safety Authority (EFSA), and Canada have set the MRL at 2.5 mg/kg [5–7], while the World Health Organization (WHO) recommends a more stringent MRL of 1 mg/kg [8]. Urea is another adulterant utilized to inflate the apparent protein content, consistency, and shelf life of milk [9,10]. However, excessive urea in milk places a considerable strain on the kidneys [11], particularly when urea adulteration reaches 500 mg/L [12]. The inclusion of sucrose is often intended to enhance the density and sweetness of milk. The reconstitution of milk refers to the practice of incorporating skim or whole milk powder into milk, either partially or as a complete replacement. This deceptive tactic helps to gain economic benefits due to the extended sell-by date of powdered milk, and its cost-effective storage and transportation compared to milk. While sucrose and milk powder adulteration in milk may not present direct threats to human life, they can severely impact the economy and undermine consumer confidence in the dairy industry.

Various analytical methods have been developed to uncover milk adulteration, encompassing electrophoresis technology [13], chromatography technology [14], and mass spectrometry. While these methods often show lower detection limits and high accuracy, they suffer from being time-intensive, intricate, and costly in terms of equipment and testing expenses. Fourier transform infrared (FTIR) spectroscopy emerges as a promising remedy to address the deficiencies of conventional detection techniques. Its advantages, such as minimal preprocessing requirements, cost-effectiveness, rapidity, non-destructiveness, environmentally friendliness, and real-time capability, make it a standout candidate [15]. Numerous studies have been conducted to identify the presence of water, milk powder, starch, formaldehyde, sucrose, melamine, urea, and sodium bicarbonate in milk through the analysis of FTIR spectroscopy [2, 16,17]. The subjects of these investigations primarily encompass raw milk, milk powder, ultra-high-temperature milk (UHT), and mixed liquid milk.

The infrared spectroscopy has wide bands and severe overlap between bands [18]. Consequently, the application of machine learning technology becomes imperative for effective data mining. While machine learning and multivariate analysis have found extensive utility in detecting milk adulteration, the predominant modeling techniques are linear methods including linear discriminant analysis (LDA), multiple linear regression (MLR), principal component regression (PCR), and partial least squares (PLS), making them less suitable when a nonlinear relationship exists between the predicted traits and spectral data. Research has underscored that in cases of low adulteration level, a nonlinear relationship exists between melamine and urea content and the spectral data [6,19,20]. Random forests, gradient boosting machines, bayesian regularized artificial neural networks, and multi-layer perceptrons are representative of nonlinear modeling algorithms, and are likewise important components of modern statistical machine learning algorithms, which can unveil intricate relationships between predictor variables and predictive features, potentially enhancing the accuracy of trait predictions. Notably, Balabin et al. [19] first ventured into applying five algorithms, which included artificial neural networks, for FTIR-based detection of melamine adulteration in dairy products. However, the study primarily focused on a quantitative model. Hence, this study aims to address three primary objectives: (1) Investigate the feasibility of utilizing FTIR spectroscopy for both quantitative and qualitative detection of pasteurized milk adulteration with melamine, urea, sucrose, water, and milk powder by employing an array of modern statistical machine learning techniques. (2) Evaluate parameters across various models to demonstrate the superior performance of models developed using modern statistical machine learning algorithms in comparison to PLS. (3) Compare the performance of models established using different spectral preprocessing techniques and modeling algorithms on the validation set to identify the best model. These models should enable fast, sensitive, reliable, and robust detection of adulterated pasteurized milk, identify the adulteration substances, and accurately predict their concentration.

2. Materials and methods

2.1. Preparation of the adulterated milk samples

Two brands of pure pasteurized milks (PPM) were bought from supermarkets in Hubei, China, which are officially authorized, including Mengniu pure milk (Mengniu Dairy Co., Ltd., Inner Mongolia) and Guangming pure milk (Bright Dairy & Food Co., Ltd., Shanghai). The origin of all samples was guaranteed by the suppliers. A total of seventy PPM were purchased and confirmed not to be adulterated. Immediately after purchase, the fresh milk samples were conveyed to the laboratory at Huazhong Agricultural University and stored at a temperature of 4 °C.

Adulteration mixtures were concocted by intentionally introducing varying level of melamine, urea, sucrose, distilled water, and milk powder into the PPM samples. Each mixture was spiked with one of five adulterants. Nescafe milk powder was purchased from a local supermarket and has been confirmed to be melamine-free. In pursuit of practical applicability and considering the real situations of the dairy industry, the mixing concentration of adulterants was determined based on the reality. Police investigations in Brazil revealed that adulterated liquid milk often contained 10–15 % (v/v) added water [21]. Consequently, the concentration of water in this study ranged from 5 to 50 % (v/v). In the case of urea, it was estimated that over 120 mg/L needed to be added to fresh milk to raise the nitrogen content by 1 %, which is economically feasible [21]. Therefore, the concentration of urea in this study spanned from 100 to 1000 mg/L (0.1–1 g/L), encompassing the critical threshold of 500 mg/L, known to be capable of inducing cancer and kidney

failure. Considering the pronounced health risks associated with melamine and the fact that EU and EFSA have imposed a limit of less than 2.5 mg/L^{5,7} and WHO [8] has set a stringent limit of less than 1 mg/kg. Accordingly, the melamine adulteration levels in this study were established within the range of 1–500 mg/L. Table 1 presents the distribution range and sample quantities for the concentration of adulterants in milk.

2.2. Spectral measurements

All samples were vortexed for 5 s to mix the adulterant well with the milk and ensure homogeneity prior to spectroscopic determination. Spectral data for PPM and adulterated pasteurized milk samples were collected by the MilkoScan FT+ (Foss, Hillerød, Denmark) instrument. The spectral data for each sample contained information on 1060 wave points, representing the infrared light absorption of the milk sample in the infrared region of 5, 008–925 cm⁻¹.

2.3. Statistical analysis

All statistical analyses were performed using R software version 4.2.2. The model developing process has five main steps: spectral preprocessing, selection of modeling bands, division of the data set, model building using algorithms, and model performance evaluation.

2.3.1. Data pre-processing

The spectra were preprocessed before developing the prediction model. Four of the most commonly used spectral preprocessing algorithms were used in this study, i.e., first-order derivative (1D), second-order derivative (2D), standard normal variable (SNV) and Savitzky-Golay convolutional smoothing (SG), and the R package that performs these preprocessing algorithms is `prospectr`.

The regions from 2, 968 to 5, 008 cm⁻¹ and 1, 692 to 1, 604 cm⁻¹ were considered noisy and removed from the dataset, and the region from 1, 773 to 2, 802 cm⁻¹ did not contain valuable information and was similarly removed [22]. Finally, the remaining 244 wavepoints were used for subsequent analyses (2, 968 to 2, 802 cm⁻¹, 1, 773 to 1, 692 cm⁻¹, and 925 to 1, 604 cm⁻¹) [23]. The pre-processed spectral data were scaled before modeling.

2.3.2. Calibration models

Firstly, The dataset was randomly split into two subsets: a calibration set (comprising 80 % of the total samples) and a validation set (comprising 20 % of the total samples). The calibration set is used to develop the prediction model and the validation set is used to evaluate the performance of models. It was ensured that at least one negative sample was included in the validation set. The dividing of dataset was conducted using the `createDataPartition()` function from the R package `caret`.

Secondly, prediction models were developed using the preprocessed FTIR spectroscopy (244 wavepoints) as explanatory variables and the type of adulterant present in the PPM or adulterant level in PPM samples as predictive traits. Two types of models were involved in this study: (1) multi-classification model: to determine if a sample was adulterated or not and to identify which adulterant was in the sample. (2) regression model: To quantitatively predict adulterant level in milk samples. 8 classification machine learning algorithms were used, including partial least squares discriminant analysis (PLSDA), support vector machine (`svmLinear` and `svmRadial`), gradient boosting machine (GBM), Random Forest (RF), extreme gradient boosting (XGB), and multi-layer perceptron (MLP), and decision tree (DT). Thirteen regression machine learning algorithms, including partial least squares regression (PLSR), SVM (`svmLinear` and `svmRadial`), spike and slab regression (SSR), projection pursuit regression (PPR), bayesian regularized neural networks (BRNN), RF, GBM, XGB, ridge regression (RR), least absolute shrinkage and selection operator (LASSO), elastic net regression (EN), and DT. Among them, `svmRadial`, GBM, RF, XGB, MLP, PPR, BRNN and DT belong to nonlinear modelling algorithms and PLS, SSR, `svmLinear`, RR, LASSO and EN belong to linear modelling algorithms. All the machine learning algorithms used in this study use R package `caret`.

The maximum number of PLS latent variables was set to 20. The number of the hidden layer for the BRNN was varied from 1 to 4. The number of the `mtry` for the RF was 3, 10, 20, 50, 100, 300, 700, 1000, and 2000. The computation of SVM was based on support

Table 1

Distribution of the levels of adulterants in milk samples and the number of samples analysed according to the adulterant added.

Group ^a	N1	N2	The range distribution of the concentration of the adulterants				
			Melamine	Urea	Sucrose	Water	Milk powder
Control	–	70	–	–	–	–	–
A	5	350	1–500 mg/L	–	–	–	–
B	5	350	–	0.1–1 g/L	–	–	–
C	5	350	–	–	0.2–5 % (w/v)	–	–
D	5	350	–	–	–	5–50 % (v/v)	–
E	4	280	–	–	–	–	5–50 % (w/v)

N1=Number of adulterant concentration levels; N2=Number of samples. ^a Control represents pure milk; A represents melamine-adulterated milk samples; B represents urea-adulterated milk samples; C represents sucrose-adulterated milk samples; D represents water-adulterated milk samples; E represents milk powder-adulterated milk samples.

vector machine with kernel or radial basis function kernel and was implemented using the method = “svmLinear” or “svmRadial” as arguments in the train function of the CARET package. For “svmLinear”, the tested C values were 0.01, 0.05, 0.1, 0.25, 0.5, 0.75, 1, 1.25, 1.5, 1.75, 2, and 5. For the “svmRadial” kernel, the tested C values were 0.01, 0.05, 0.1, 0.25, 0.5, 0.75, 1, 1.25, 1.5, 1.75, 2, and 5; the sigma values were 0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 0.1, 0.25, 0.5, 0.75, 0.90. The parameters of the other algorithms are randomized using the “tuneLength = 10” function in Caret.

2.3.3. Model performance evaluation and optimal model selection

For problems involving multi-class classification, model performance evaluation relies on model accuracy.

For quantitative models, the coefficient of determination of calibration (R_c^2), the coefficient of determination of validation (R_v^2), the root mean square error of calibration (RMSEC), the root mean square error value of validation (RMSEV) and the ratio of performance to deviation (RPD) were used as model performance assessment metrics. Tamaki et al. [24] summarised the various performance parameters of the quantitative regression model: an R^2 of 0.66–0.80, 0.81–0.90 and > 0.90 indicate fair, good and excellent model predictions respectively. Williams [25] used RPD to grade the performance of the model: very poor prediction (RPD < 2.3), poor

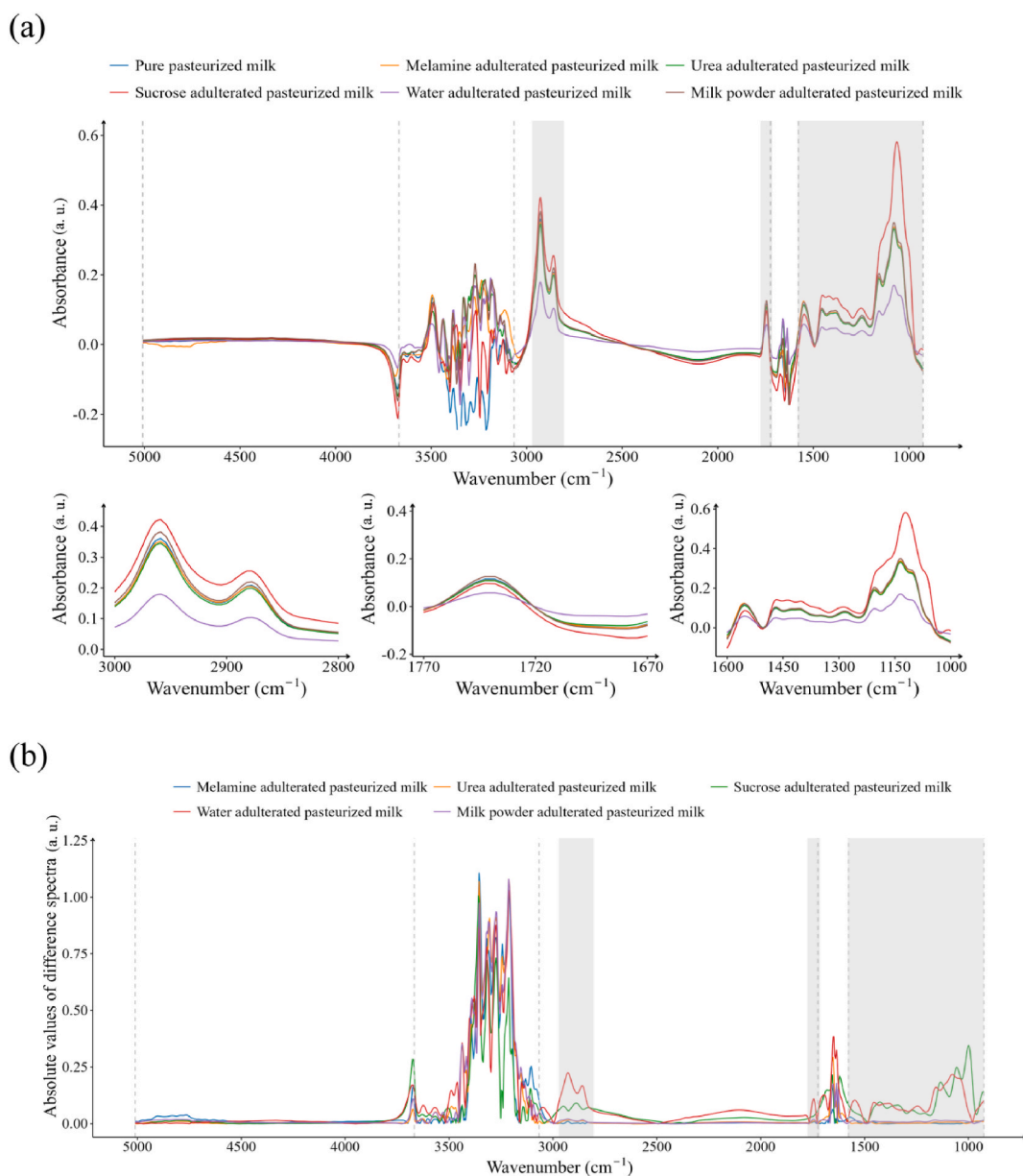


Fig. 1. (a) Average spectra of all pure and adulterated (highest adulteration level) milk samples. (b) Absolute values of the difference between the spectra of pure milk and the adulterated milk samples at the highest level. Spectral region used in the models is marked by the dashed squares.

prediction ($2.4 < RPD < 3.0$), fair prediction ($3.1 < RPD < 4.9$), good prediction ($5.0 < RPD < 6.4$), very good prediction ($6.5 < RPD < 8.0$) and excellent prediction ($RPD > 8.0$). Therefore, a robust and accurate quantitative prediction model should have an RPD value greater than 8.0, an R^2 value close to 1 and the smallest possible prediction error value.

RMSE was calculated according to Eq. (1), where n is the number of samples in the dataset, y_i is the true value, and \hat{y}_i is the predicted value. RPD was calculated by Eq. (2), where SD is the standard deviation of the true value. In addition, to assess the prediction error of the model for each adulteration concentration, the percentage absolute error (PAE) was used and was calculated as Eq. (3) [20].

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (1)$$

$$RPD = \frac{SD}{RMSEV} \quad (2)$$

$$PAE(\text{in}\%) = \frac{|y_i - \hat{y}_i|}{y_i} \times 100 \quad (3)$$

3. Results and discussion

3.1. FTIR spectroscopy of pure and adulterated milk

The FTIR spectroscopy of milk consists of 1060 individual spectral points in the range of 925 to 5008 cm^{-1} , which are divided into short-wavelength infrared region (SWIR), mid-wavelength infrared region (MWIR), and long-wavelength infrared region (LWIR). In more detailed terms, it is divided into five regions: SWIR (5010 to 3673 cm^{-1}), SWIR-MWIR (3669 to 3052 cm^{-1}), MWIR-1 (3048 to 1701 cm^{-1}), MWIR-2 (1698 to 1585 cm^{-1}), and MWIR-LWIR (1582 – 925 cm^{-1}) [26].

Fig. 1 displays the average spectra and difference spectra of PPM and adulterated milk. In comparison to PPM, milk samples adulterated with sucrose exhibit an increase in peak amplitude, while milk samples adulterated with water display a decrease. The peak amplitudes of milk samples adulterated with melamine, urea, or milk powder remain relatively unchanged. In this study, it was found that the spectral differences between PPM and adulterated milk mainly existed in the MWIR-1, MWIR-2, SWIR-MWIR and MWIR-LWIR regions. The MWIR-2 and SWIR-MWIR regions are associated with water absorption, and these spectral features increase the variability of absorbance between milk samples. These two regions are often excluded when predicting milk compositions and cow's physiological status [27]. Upon eliminating the water absorption region, the most significant differences become apparent within MWIR-LWIR region, corresponding to sugar absorption (1141 – 972 cm^{-1}) and milk protein absorption (1640 – 1500 cm^{-1}). The main absorption peaks in the MWIR-1 region are the C–H, C=O, C–N and N–H bonds [28], all of which are related to the milk fat content. In this region, a number of peaks with large differences in absorbance were detected. The first significant spectral region was located at 2993 to 2802 cm^{-1} , which is related to the C–H bond vibration of Fat-B [29], and the second significant spectral region was located at 1778 to 1692 cm^{-1} , which is related to the carbonyl vibration of Fat-A [29].

3.2. Multi-classification model used for distinguishing adulterant types

The focus on identifying the presence of adulterants takes precedence over quantification [3]. To address this, a multi-classification model, using FTIR spectroscopy from 1401 milk samples, was established to determine if a sample was adulterated or not and to detect adulterant types in pasteurized milk. The detailed discrimination results for the samples in the validation set are presented in Table 2.

Table 2

Accuracies for all classes and each individual class (pure, melamine, urea, sucrose, water, and milk powder) in validation set for multi-classification models.

Index	Number	PLS	LSVM	RSVM	GBM	RF	XGB	MLP	DT
Pretr. ^a	-	SG	SNV	None	1D	1D	1D	1D	None
Total accuracy ^b	349	0.937	0.788	0.725	0.968	0.957	0.963	0.974	0.906
Pure	16	0.875	0.688	0.813	0.750	0.813	0.813	0.938	0.545
Melamine	69	1.000	1.000	0.826	1.000	1.000	0.986	1.000	1.000
Urea	73	0.973	0.507	0.219	0.986	1.000	0.986	1.000	0.965
Sucrose	70	0.986	0.971	1.000	1.000	1.000	1.000	1.000	0.971
Water	70	0.786	0.957	1.000	0.814	0.886	0.914	0.943	0.886
Milk powder	51	0.961	0.451	0.529	0.980	0.922	0.961	0.922	0.802

PLS = partial least squares; LSVM = svmLinear (support vector machine with kernel); RSVM = svmRadial (support vector machine with radial basis function kernel); RF = random forest, and GBM = gradient boosting machine; XGB = eXtreme gradient boosting; MLP = multi-layer perceptron; DT = Decision Tree; 1D = first-order derivative; 2D = second-order derivative; SNV = standard normal variate; SG = Savitzky–Goltsy convolution smoothing.

^a Pretr. means pretreatment method of FTIR spectroscopy. Only the best spectral pre-processing results are shown.

^b Accuracy refers to the proportion of category A that is correctly predicted to be category A, i.e. Sensitivity and Specificity.

The results revealed that for multi-classification tasks, the GBM, RF, XGB, and MLP algorithms outperformed the PLS algorithm, improving the classification accuracy in validation set by 3.1 %, 2.0 %, 2.6 %, and 3.7 %, respectively. SVM algorithm exhibited the weakest performance. SVM was originally designed as a binary classification tool and is therefore better suited for binary classification problems [19]. In addition, we find that ensemble technique such as XGB and RF outperform DT algorithms. The optimal model achieved a prediction accuracy of 97.4 % through the MLP modeling algorithm combined with the 1D spectral preprocessing technique. Notably, this model misclassified only nine milk samples in the validation set (Table 3). Interestingly, the misclassification pattern occurred between PPM, milk powder-adulterated milk, and water-adulterated milk (Table 3), indicating that the spectral features of these 3 sample types are difficult to extract, leading to potential confusion in predictions.

In the interest of safeguarding consumer health, stringent control measures are placed on false negative results (samples incorrectly predicted as unadulterated) to prevent adulterated milk from infiltrating the market. Conversely, false positive results, where the model erroneously identifies an unadulterated sample as adulterated, are less concerning, as all positively identified samples can be further verified through reference methods [2]. From Table 2, it can be seen that the MLP algorithm is the best predictor for pure milk, with 93.8 % of the pure milk samples correctly classified (i.e., model specificity is 93.8 %). In addition, this model also predicts adulterated milk very well with 99.7 % of the adulterated milk samples correctly classified (i.e., model sensitivity is 99.7 %). From Table 3, it can be seen that there is one false negative and one false positive in the prediction results of the validation set, with false negative rates (false negative samples/actually positive samples) and false positive rates (false positive samples/actually negative samples) of 0.3 % and 6.2 %, respectively. The performance of the model application needs to be verified by more data.

In a previous study by Santos et al. [30], FTIR spectroscopy and SIMCA were used to detect six adulterants in milk, achieving multi-classification accuracy of 90 % for pure milk and 98 % for urea [30]. In another study by Gondim et al. [31], similar technology was employed to detect more than ten adulterants in milk, but the multi-classification model only achieved classification accuracies of 66 % for pure milk and 38 % for sucrose.

3.3. The quantitative regression model of predicting adulteration level

Regression models were established using thirteen machine learning algorithms to predict the levels of adulterants in pasteurized milk. With the exception of the regression model used to predict the adulteration levels of milk powder, which was trained on 280 samples and validated on 70 samples, the modeling and validation sets for the other adulterant levels prediction models consisted of 336 and 84 samples, respectively. Table 4 provides an overview of the prediction performance for each machine learning algorithm under the optimal spectral preprocessing method. The spectral data underwent various preprocessing techniques, with many models achieving their best performance using SG or SNV.

For models predicting the level of adulterated melamine, urea, sucrose, or milk powder in milk, the performance of linear modeling algorithms was consistently similar to the PLS algorithm, while nonlinear modeling algorithms consistently outperformed them, especially BRNN, RF, GBM, and XGB (Table 4). However, when predicting the level of water added to pasteurized milk, a different pattern emerged: RF, GBM, DT, and XGB, which had consistently performed well, now performed worse than the six linear modeling algorithms, and RSVM also underperformed LSVM (Table 4). This suggests a potential linear relationship between FTIR spectroscopy and the level of water added, but a nonlinear relationship between FTIR spectroscopy and the level of adulterated melamine, urea, sucrose, and milk powder in milk. Similar patterns have been observed in research using Raman spectroscopy and near-infrared spectroscopy to predict milk adulteration containing melamine and urea [6,19,20]. In most cases, the DT algorithm is always worse than the three ensemble techniques (RF, GBM and XGB). Out of the thirteen modeling algorithms, BRNN consistently demonstrated superior prediction performance, even when the trait being predicted was the level of water added, which may exhibit a linear relationship with FTIR spectroscopy. Research has indicated that compared to PLS, neural networks (NN) can offer improved predictions [32].

The selection of the best model was based on the criteria of having the lowest RMSEV, the highest R^2_V and RPD. The optimal model for predicting sucrose level in pasteurized milk was established using the XGB modeling algorithm and the 1D spectral preprocessing method. This model exhibited remarkable predictive accuracy, with a near-zero RMSE, slopes and R^2_V approaching 1, and an extraordinarily high RPD value. For predicting the level of adulterated water in pasteurized milk, the model established using the PPR algorithm and SNV spectral preprocessing method also demonstrated excellent predictive capabilities. It achieved an RMSEV of 1.248 % (v/v), an R^2_V of 0.995, and an RPD of 14.087. In the case of predicting the level of adulterated melamine or urea in pasteurized milk, the most effective models were established using the BRNN modeling algorithm and the SG spectral preprocessing method. The

Table 3

Classification matrix of validation set for a MLP (Multi-Layer Perceptron) multiclass classification.

Real	Predicted					
	Pure	Melamine	Urea	Sucrose	Water	Milk powder
Pure	15	0	0	0	0	1
Melamine	0	69	0	0	0	0
Urea	0	0	73	0	0	0
Sucrose	0	0	0	70	0	0
Water	1	0	0	0	66	3
Milk powder	0	0	0	0	4	47

Table 4

Statistical parameters of regression models from FTIR spectroscopy data for the quantitative prediction of adulterated melamine, urea, sucrose, water and milk powder content in pasteurized milk.

Adulterant	Modeling	Pretr. ^a	Calibration set		Validation set			
			RMSEC	R _C ²	RMSEV	R _V ²	RPD	
Melamine	PLSR	SNV	21.478	0.985	22.079	0.985	8.121	
	LSVM	SG	21.188	0.996	21.000	0.986	8.538	
	RSVM	SG	16.043	0.996	11.506	0.997	15.584	
	SSR	SNV	23.481	0.982	21.649	0.985	8.282	
	PPR	SNV	1.254	1.000	6.033	0.999	29.719	
	BRNN	SG	1.792	1.000	4.437	0.999	40.414	
	RF	SNV	7.742	0.998	5.094	0.999	35.195	
	GBM	SNV	11.279	0.996	7.544	0.998	23.768	
	XGB	1D	0.001	1.000	6.158	0.999	29.114	
	DT	None	12.548	0.995	14.869	0.993	12.059	
	RR	SNV	24.715	0.981	23.175	0.983	7.737	
	LASSO	SNV	25.476	0.979	23.343	0.983	7.681	
	EN	SNV	21.338	0.986	21.022	0.986	8.529	
	Urea	PLSR	None	0.107	0.910	0.111	0.902	3.201
		LSVM	SNV	0.108	0.911	0.118	0.889	3.010
RSVM		SG	0.047	0.984	0.054	0.978	6.644	
SSR		SNV	0.118	0.891	0.112	0.905	3.169	
PPR		None	0.044	0.985	0.096	0.929	3.707	
BRNN		SG	0.023	0.996	0.049	0.981	7.330	
RF		None	0.036	0.991	0.053	0.979	6.689	
GBM		None	0.006	1.000	0.061	0.971	5.862	
XGB		None	0.001	1.000	0.051	0.980	6.983	
DT		SG	0.071	0.960	0.074	0.958	4.794	
RR		1D	0.121	0.886	0.114	0.898	3.132	
LASSO		SNV	0.096	0.927	0.109	0.905	3.257	
EN		SG	0.107	0.909	0.109	0.906	3.257	
Sucrose		PLSR	2D	0.071	0.998	0.054	0.999	32.548
		LSVM	None	0.085	0.998	0.081	0.998	21.831
	RSVM	SNV	0.106	0.997	0.117	0.996	15.078	
	SSR	2D	0.069	0.998	0.048	0.999	36.421	
	PPR	1D	0.017	1.000	0.043	0.999	41.021	
	BRNN	SNV	0.013	1.000	0.023	1.000	76.526	
	RF	1D	0.008	1.000	0.018	1.000	98.449	
	GBM	2D	0.000	1.000	0.017	1.000	105.281	
	XGB	1D	0.000	1.000	0.000	1.000	29487.494	
	DT	SG	0.017	1.000	0.022	1.000	79.888	
	RR	2D	0.072	0.998	0.052	0.999	33.819	
	LASSO	2D	0.074	0.998	0.052	0.999	33.982	
	EN	1D	0.073	0.998	0.054	0.999	32.680	
	Water	PLSR	1D	1.693	0.991	1.407	0.994	12.500
		LSVM	1D	1.853	0.989	1.474	0.993	11.929
RSVM		1D	1.544	0.993	1.896	0.988	9.276	
SSR		2D	2.040	0.987	1.616	0.992	10.881	
PPR		SNV	0.274	1.000	1.248	0.995	14.087	
BRNN		None	1.379	0.994	1.373	0.994	12.813	
RF		1D	0.866	0.998	1.607	0.992	10.943	
GBM		1D	0.263	1.000	1.769	0.990	9.939	
XGB		2D	0.011	1.000	1.865	0.989	9.428	
DT		2D	1.128	0.996	1.738	0.990	10.778	
RR		2D	2.157	0.985	1.760	0.990	9.991	
LASSO		1D	1.938	0.988	1.528	0.992	11.510	
EN		1D	1.760	0.990	1.378	0.994	12.763	
Milk powder		PLSR	SG	5.279	0.912	6.069	0.889	2.950
		LSVM	SNV	5.293	0.912	6.714	0.862	2.667
	RSVM	SNV	1.662	0.993	4.731	0.936	3.784	
	SSR	SNV	6.036	0.888	6.941	0.848	2.579	
	PPR	SNV	0.872	0.998	5.598	0.904	3.199	
	BRNN	SG	2.566	0.979	4.295	0.944	4.168	
	RF	SNV	3.040	0.984	5.833	0.917	3.070	
	GBM	SNV	0.153	1.000	5.630	0.905	3.180	
	XGB	SNV	0.051	1.000	6.665	0.870	2.686	
	DT	None	8.336	0.780	8.501	0.786	2.106	
	RR	SNV	5.966	0.890	7.042	0.844	2.543	
	LASSO	None	5.121	0.917	6.183	0.887	2.896	
	EN	None	5.337	0.910	6.247	0.884	2.866	

PLSR=Partial least squares regression; LSVM = svmLinear (support vector machine with kernel); RSVM = svmRadial (support vector machine with radial basis function kernel); SSR= Spike and Slab Regression, PPR= Projection Pursuit Regression; BRNN=Bayesian Regularized Neural Networks; RF= Random Forest; GBM = Gradient Boosting Machine; XGB = eXtreme Gradient Boosting; RR= Ridge Regression; LASSO = Least Absolute Shrinkage and Selection Operator; EN = Elastic net Regression; DT = Decision Tree; 1D = first-order derivative; 2D = second-order derivative; SNV = standard normal variate; SG = Savitzky–Goltsy convolution smoothing. R_c^2 = coefficient of determination for calibration; R_v^2 = coefficient of determination for external validation; RMSEC = root mean square error of calibration; RMSEV = root mean square error of external validation; RPD = the ratio of performance to deviation; RPD = Residual predictive deviation.

^a Pretr. means pretreatment method of FTIR spectroscopy. Only the best spectral pre-processing results are shown.

melamine content prediction model exhibited RMSEV, R_v^2 , and RPD values of 4.437 mg/L, 0.999, and 40.414, respectively. The model for predicting the adulterated urea in milk yielded RMSEV, R_v^2 , and RPD values of 0.049 g/L, 0.981, and 7.330, respectively. Moreover, the above combined methods (BRNN modeling algorithm and the SG spectral preprocessing method) also delivered the best results for predicting the level of adulterated milk powder in pasteurized milk. While the prediction performance was good, with RMSEV, R_v^2 , and RPD values of 4.295 % (w/v), 0.944, and 4.168, respectively, the accuracy was slightly lower compared to the models used for predicting the spiked contents of melamine, urea, sucrose, and water. In a word, the order of FTIR spectroscopy prediction performance for the five adulterants from high to low was sucrose, melamine, water, urea, and milk powder (Table 4). The R_v^2 values for the optimal prediction models of melamine, sucrose, and water adulteration levels all exceeded 0.99, with RPD values surpassing 8. Manley [33] suggested that prediction models with RPD values greater than 8 are valuable for various analytical applications. The RPD of the best prediction model for adulterated urea level, while slightly lower at 7.33, still indicates its usability for process control [34]. This slight variation can be attributed to the complex composition and properties of milk, as milk naturally contains urea, with concentrations ranging from 180 to 400 mg/L¹¹. Consequently, adulterated urea predictions are inherently more challenging using FTIR spectroscopy. Detecting milk powder in fresh milk is also a formidable analytical task due to the similar chemical composition of milk powder and milk. Nevertheless, the optimal model established in this study for predicting the level of milk powder added achieved an R_v^2 higher than 0.90, an RPD exceeding 3, and an RMSE smaller than the added ratio. These statistical parameters indicate that the model can provide reasonably accurate predictions of the level of milk powder added. However, this model is only suitable for screening purposes [34].

Scatter plots of observed vs. predicted values obtained from the best model for detection of adulterated melamine, urea, sucrose, water, and milk powder in milk were plotted (Fig. 2). These plots display an evident relation between the predicted and observed adulterant levels. The scatter plots of all five best models have a regression line that basically coincides with $y = x$, with slope close to 1 (the intercept and deviation are negligible), further demonstrating the superior performance of the regression model.

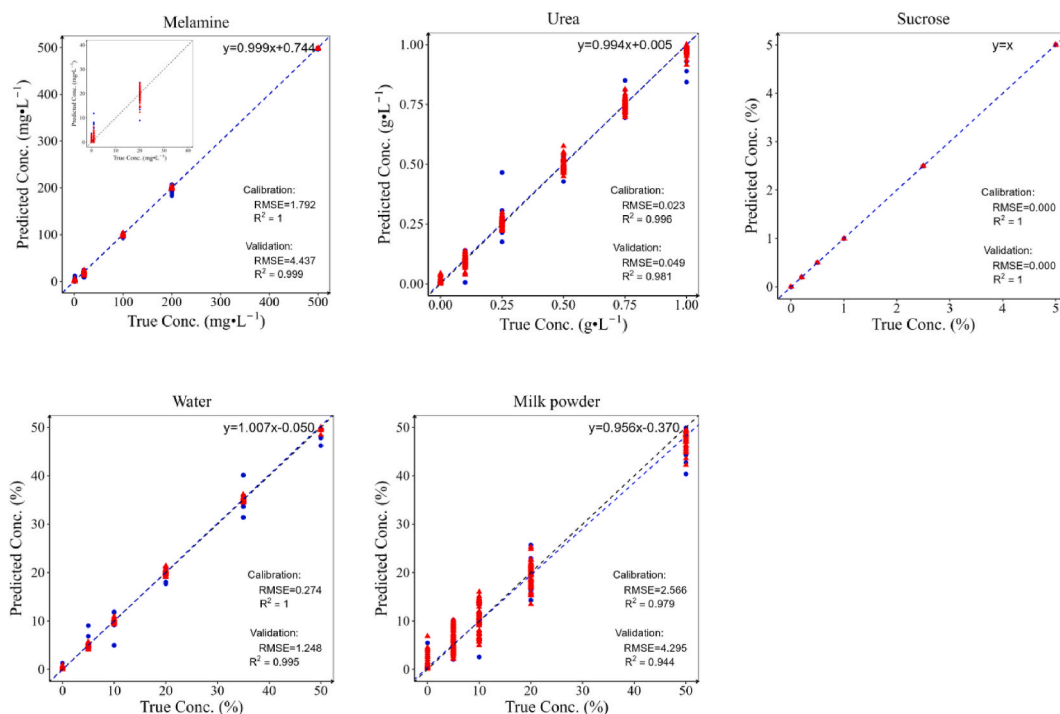


Fig. 2. The optimal regression model curves for training (circles) and validation (triangle) data milk sets spiked with melamine, urea, sucrose, water and (e) milk powder. The inset shows a zoomed view of the plots in the lower concentration ranges (0–20 mg/L). Blue circle: validation set; red triangle: calibration set; blue dashed line: trend line in validation set; black dashed line: $y = x$. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

To date, there is a noticeable absence of literature employing FTIR spectroscopy for the exclusive investigation of pasteurized milk adulteration, and this work marks an initial step in leveraging FTIR spectroscopy technology for such purposes. Moreover, to our knowledge, this article represents the first endeavor to combine FTIR spectroscopy with a diverse array of modern statistical machine learning algorithms, facilitating the simultaneous quantitative and qualitative detection of adulterants like melamine, urea, sucrose, water, and milk powder in pasteurized milk. Past studies have predominantly utilized Raman spectroscopy and Near infrared (NIR) spectroscopy for predicting milk adulterants [20,34–39]. The FTIR model developed in this study shows superior predictive capabilities when compared to Raman spectroscopy and NIR spectroscopy. The mid-infrared (MIR) spectroscopy is the most important region in FTIR spectroscopy [40]. Crucial distinctions between MIR and NIR spectroscopy influence their adeptness at detecting specific adulterants. Generally, NIR detection is more challenging due to the markedly lower absorbance intensity in NIR spectra in comparison to MIR. Moreover, NIR exhibits less sensitivity since its absorption bands are wider and tend to overlap [41]. Santos et al. [42] similarly ascertained that the MIR system outperforms the NIR system in discerning milk adulteration. Currently, several studies aim to employ FTIR spectroscopy for assessing milk adulteration levels. The prediction error for adulterated sucrose content was 0.202 % (w/v) [43]. The prediction error for adulterated urea content predictions spanned from 0.245 g/L to 3.03 g/L [10,44,45]. Predictions for milk powder content exhibited an error rate of 3 % (w/v) [46]. Except for the milk powder content prediction model, the models introduced in this study surpass earlier literature and provide markedly improved predictive performance. The strength of this study lies in its relatively large data size and the endeavor to explore an array of modeling algorithms and spectral preprocessing techniques, thereby uncovering valuable information within FTIR spectroscopy and constructing a robust prediction model.

Table 5

Comparison of the Limit of detection (LOD) of the developed technique with previous ones reported in literature.

Reference	Analytical approach	Data analysis technique	LOD ^a
Melamine			
This work	FTIR spectroscopy	BRNN	13.311 mg/L
Hilding-Ohlsson et al. [49]	cyclic voltammetry	K nearest neighbors	85.0 mg/L*
Hu et al. [6]	Raman spectroscopy	Linear regression model	0.265 mg/L
Wu et al. [36]	NIR spectroscopy	PLS	100 mg/L*
Yang et al. [50]	2D IR/NIR correlation spectra	multi-way PLS	10 mg/L*
Nieuwoudt et al. [34]	Raman spectroscopy	PLS	100–260 mg/L
Nieuwoudt et al. [35]	Raman spectroscopy	PLS	154–522 mg/L
Chen et al. [4]	NIR spectroscopy	one-class PLS	10 mg/L*
Barreto et al. [51]	fluorescence spectroscopy	Parallel factor analysis	120.6 mg/L
Jin et al. [52]	2T2D auto-correlation spectra	least squares support vector machine	10 mg/L*
Urea			
This work	FTIR spectroscopy	BRNN	0.147 g/L
Hilding-Ohlsson et al. [49]	cyclic voltammetry	K nearest neighbors	0.12 g/L*
Santos et al. [30]	MIR microspectroscopy	SIMCA	0.78 g/L*
Jha et al. [45]	ATR-FTIR spectroscopy	SIMCA	0.1 g/L*
Khan et al. [20]	Raman spectroscopy	PLS	0.5 g/L*
Nieuwoudt et al. [34]	Raman spectroscopy	PLS	0.12–0.49 g/L
Nieuwoudt et al. [35]	portable mini-Raman	PLS	0.136–0.498 g/L
De Toledo et al. [53]	diffuse reflectance spectroscopy	unknow	0.0066 g/L
Mabood et al. [12]	NIR spectroscopy	PLS	1 g/L*
Huang et al. [54]	Temperature-perturbed 2D correlation spectra	multi-way PLS	0.0002 g/L*
Sharifi et al. [55]	Vis-SWNIR	artificial neural networks	2 g/L
Tan et al. [10]	ATR-FTIR spectroscopy	Discriminant Analysis	5 g/L*
Sucrose			
This work	FTIR spectroscopy	XGB	Close to 0 % w/v
Nieuwoudt et al. [34]	Raman spectroscopy	PLS	0.72–2.5 % w/v
Nieuwoudt et al. [35]	Raman spectroscopy	PLS	0.70–3.6 % w/v
Gondim et al. [31]	MIR spectroscopy	SIMCA	0.54 %*w/v
Balan et al. [43]	FTIR spectroscopy	SIMCA	0.480 % w/v
Water			
This work	FTIR spectroscopy	PPR	3.744 % v/v
Kasemsumran et al. [56]	NIR spectroscopy	PLS	1 % v/v*
Gondim et al. [31]	MIR spectroscopy	SIMCA	15 % w/v*
Kamboj et al. [38]	NIR spectroscopy	PLS	5 % v/w*
Milk powder			
This work	FTIR spectroscopy	BRNN	12.885 % w/v
Guan et al. [57]	Tryptofan fluorescence values	unknow	10 % w/v*
Du et al. [46]	FTIR spectroscopy	PLS	0.5 % w/v*
Nikolaou et al. [9]	electro-analytical cyclic voltammetry test	PLS	5.8 % v/v

FTIR spectroscopy = Fourier-transform infrared spectroscopy; 2D IR/NIR correlation spectra = Two-dimensional hetero-spectral mid-infrared and near-infrared correlation spectroscopy; NIR spectroscopy = near-infrared spectroscopy; MIR spectroscopy = mid-infrared spectroscopy; 2T2D auto-correlation spectra = two-trace two-dimensional auto-correlation spectra; ATR-FTIR = Attenuated total reflection-Fourier transform infrared spectroscopy; Vis-SWNIR = visible to short-wave near infrared; PLS = partial least squares; SIMCA = Soft Independent Modeling of Class Analogy; BRNN = bayesian regularized neural network; XGB = extreme gradient boosting. ^a * indicates that LOD is calculated as the lowest value detectable by the classification model.

Limits of detection (LOD) corresponds to the lowest amount of analyte in the sample, which can be detected but not necessarily quantitated under stated experimental conditions [47]. Certain studies have suggested that, for multivariable regression models like PLSR, the LOD can be roughly estimated as $3 \times \text{RMSEP}$ [34,48]. In light of these considerations, the LODs for the best prediction models developed in this article were determined to be 13.311 mg/L for melamine, 0.147 g/L for urea, 0 % (w/v) for sucrose, 3.744 % (v/v) for water, and 12.885 % (w/v) for milk powder. Table 5 provides a summary of LOD comparisons between the model established in this article and the detection limits reported in prior literature. In most cases, the detection limits in this study are notably lower than those of other spectroscopic analysis methods, such as Raman spectroscopy and NIR spectroscopy.

Furthermore, it is worthwhile to investigate the variations in predictive model performance across samples with differing adulteration concentrations. The PAE, denoting the prediction error between the predicted and actual values, serves as a valuable metric in this context. The adulterant content significantly influences the regression model's prediction accuracy. Notably, the PAE diminishes as the adulterant concentration increases, observed in both the calibration and validation sets (Table 6). With the exception of sucrose, the remaining four adulterants exhibit higher PAEs at very low concentrations (level 1), indicating that the relative uncertainty of predictions is higher. This increase in PAE may be attributed to a relatively weaker signal-to-noise ratio at lower concentrations. This pattern aligns with finding from Khan et al. [20], who revealed that the performance of Raman spectroscopy in identifying adulterated milk correlates positively with the adulterated solution's concentration. In sum, the regression models presented in this article can successfully predict adulteration level with an accuracy exceeding 90 % (a PAE of less than 10 %) for melamine >100 mg/L, urea >0.5 g/L, sucrose >0.2 %, water >5 %, and milk powder >50 %.

The models developed in this study allow for large-scale screening of pasteurized milk sold in the market for adulteration. Compared to previous studies, this study not only has an advantage in terms of data volume, but also generates a more accurate predictive model by comparing different combinations of spectral preprocessing algorithms and machine learning algorithms. However, the models developed in this study were based on only two brands of (more popular) pasteurized milk, which means it is likely not applicable to other brands of pasteurized milk. Therefore, a more comprehensive dataset to sample variations in type (broader calibration data) is also needed to build a more general model to predict adulteration of pasteurized milk. In addition, the predictive performance of the model is unknown when multiple adulterants are added to the milk, which is a shortcoming of this study. In most cases, there may be more than one adulterant in milk.

4. Conclusion

In this work, we demonstrate the feasibility of FTIR spectroscopy in conjunction with modern statistical machine learning algorithms in the identification and quantification of adulteration of pasteurized milk by the addition of melamine, urea, sucrose, water, and milk powder. In contrast to traditional linear methods like PLS, modern non-linear statistical machine learning algorithms exhibit high predictive performance, with BRNN standing out. The established models in this work are not limited to laboratory FTIR spectrometers but can also be integrated with portable spectrometers or sensors for on-farm self-inspections and regulatory authorities' random inspections, thus bolstering milk safety testing efforts on a larger scale. Furthermore, the development of software, applications, and online service platforms holds the potential to enhance transparency and openness in adulteration detection, fostering consumer trust, bolstering confidence in the dairy industry, and serving as a deterrent to adulteration by dairy farms and producers. Future studies should expand scopes by encompassing a broader range of adulteration concentrations, encompassing various pasteurized milk brands and production batches, and examining scenarios involving the simultaneous adulteration of milk with multiple adulterants. This approach will enhance the understanding of FTIR spectroscopy patterns of milk, and also align more closely with real-world production conditions.

CRedit authorship contribution statement

Chu Chu: Writing – review & editing, Writing – original draft, Visualization, Software, Data curation, Conceptualization. **Haitong Wang:** Writing – review & editing, Writing – original draft, Validation, Investigation, Data curation. **Xuelu Luo:** Writing – review & editing, Writing – original draft, Validation, Investigation, Data curation. **Yikai Fan:** Writing – review & editing, Visualization.

Table 6

Percentage absolute error (PAE, %) at each Melamine, urea, sucrose, water, and milk powder concentration.

Adulteration Level ^a	Calibration set					Validation set				
	Melamine	Urea	Sucrose	Water	Milk powder	Melamine	Urea	Sucrose	Water	Milk powder
level1	163.600	16.095	0.018	3.681	42.231	386.934	22.011	0.018	10.450	53.188
level2	9.573	6.181	0.004	1.430	23.326	16.660	18.149	0.004	7.407	35.386
level3	1.455	4.212	0.001	0.919	9.645	2.886	4.861	0.001	1.976	19.636
level4	0.547	2.592	0.001	0.419	4.749	2.783	4.462	0.001	2.559	10.013
level5	0.210	2.076	0.002	0.217		0.259	5.123	0.002	1.277	

^a The level of adulteration gradually increases from level 1 to level 5. The concentration of adulterated melamine is 1 mg/L, 20 mg/L, 100 mg/L, 200 mg/L, and 500 mg/L. The concentration of adulterated urea is 0.1 g/L, 0.25 g/L, 0.5 g/L, 0.75 g/L, and 1 g/L. The concentration of adulterated sucrose is 0.2 %, 0.5 %, 1 %, 2.5 %, and 5 % (w/v). The concentration of adulterated water is 5 %, 10 %, 20 %, 35 %, and 50 % (v/v). The concentration of adulterated milk powder is 5 %, 10 %, 20 %, and 50 % (w/v).

Liangkang Nan: Supervision, Methodology, Data curation. **Chao Du:** Writing – review & editing, Methodology. **Dengying Gao:** Writing – review & editing. **Peipei Wen:** Writing – review & editing, Validation. **Dongwei Wang:** Writing – review & editing. **Zhuo Yang:** Writing – review & editing. **Guochang Yang:** Writing – review & editing. **Li Liu:** Writing – review & editing. **Yongqing Li:** Writing – review & editing. **Bo Hu:** Writing – review & editing. **Abula Zunongjiang:** Writing – review & editing. **Shujun Zhang:** Writing – review & editing, Supervision, Project administration, Funding acquisition, Conceptualization.

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.

Acknowledgements

Financial assistance from the Fundamental Research Funds for the Inter-Governmental International Science and Technology Cooperation Project of the State Key Research and Development Program (2021YFE0115500), National Center of Technology Innovation for Dairy (2022-KYGG-3), the Central Universities (2662023DKPY001), and International Science and Technology Cooperation Project of Hubei Province (2022EHB043) are highly appreciated.

References

- [1] A. Karmaker, P.C. Das, A. Iqbal, Quality assessment of different commercial and local milk available in the local markets of selected area of Bangladesh, *J Adv Vet Anim Res* 7 (1) (2019 Nov 20) 26–33, <https://doi.org/10.5455/javar.2020.g389>. PMID: 32219106; PMCID: PMC7096109.
- [2] B.G. Botelho, N. Reis, L.S. Oliveira, M.M. Sena, Development and analytical validation of a screening method for simultaneous detection of five adulterants in raw milk using mid-infrared spectroscopy and PLS-DA, *Food Chem.* 181 (2015 Aug 15) 31–37, <https://doi.org/10.1016/j.foodchem.2015.02.077>. Epub 2015 Feb 20. PMID: 25794717.
- [3] Ghaffar Al-Lafi Abdul, AL-Naser Isam, Application of 2D-COS-FTIR spectroscopic analysis to milk powder adulteration: detection of melamine, *J. Food Compos. Anal.* (2022) 104720, <https://doi.org/10.1016/j.jfca.2022.104720>.
- [4] H. Chen, C. Tan, Z. Lin, T. Wu, Detection of melamine adulteration in milk by near-infrared spectroscopy and one-class partial least squares, *Spectrochim. Acta Mol. Biomol. Spectrosc.* 173 (2017 Feb 15) 832–836, <https://doi.org/10.1016/j.saa.2016.10.051>. Epub 2016 Oct 28. PMID: 27816741.
- [5] S. Jawaid, F.N. Talpur, S.T. Sherazi, S.M. Nizamani, A.A. Khaskheli, Rapid detection of melamine adulteration in dairy milk by SB-ATR-Fourier transform infrared spectroscopy, *Food Chem.* 141 (3) (2013 Dec 1) 3066–3071, <https://doi.org/10.1016/j.foodchem.2013.05.106>. Epub 2013 Jun 2. PMID: 23871060.
- [6] Y. Hu, S. Feng, F. Gao, E.C. Li-Chan, E. Grant, X. Lu, Detection of melamine in milk using molecularly imprinted polymers-surface enhanced Raman spectroscopy, *Food Chem.* 176 (2015 Jun 1) 123–129, <https://doi.org/10.1016/j.foodchem.2014.12.051>. Epub 2014 Dec 20. PMID: 25624214.
- [7] C.F. Nascimento, P.M. Santos, E.R. Pereira-Filho, F.R.P. Rocha, Recent advances on determination of milk adulterants, *Food Chem.* 221 (2017 Apr 15) 1232–1244, <https://doi.org/10.1016/j.foodchem.2016.11.034>. Epub 2016 Nov 9. PMID: 27979084.
- [8] F. World Health Organization, A. O. of the U. N., Proposed draft maximum levels for melamine in food and feed (N13-2009), in: *Uniform Procedure for the Elaboration of Codex Standards and Related Texts*, March, vol. 26, 2010.
- [9] P. Nikolaou, E. Deskoulidis, E. Topoglidis, A.T. Kakoulidou, F. Tsopelas, Application of chemometrics for detection and modeling of adulteration of fresh cow milk with reconstituted skim milk powder using voltammetric fingerprinting on a graphite/SiO₂ hybrid electrode, *Talanta* 206 (2020 Jan 1) 120223, <https://doi.org/10.1016/j.talanta.2019.120223>. Epub 2019 Aug 5. PMID: 31514874.
- [10] E. Tan, N. Binti Julmohammad, W.Y. Koh, M.S. Abdullah Sanii, B. Rasti, Application of ATR-FTIR incorporated with multivariate data analysis for discrimination and quantification of urea as an adulterant in UHT milk, *Foods* 12 (15) (2023 Jul 27) 2855, <https://doi.org/10.3390/foods12152855>. PMID: 37569123; PMCID: PMC10417858.
- [11] H. Toledo-Alvarado, A.I. Vazquez, G. de Los Campos, R.J. Tempelman, G. Bittante, A. Cecchinato, Diagnosing pregnancy status using infrared spectra and milk composition in dairy cows, *J. Dairy Sci.* 101 (3) (2018 Mar) 2496–2505, <https://doi.org/10.3168/jds.2017-13647>. Epub 2017 Dec 28. PMID: 29290427.
- [12] F. Mabood, L. Ali, R. Boque, G. Abbas, F. Jabeeen, Q.M.I. Haq, J. Hussain, A.M. Hamaed, Z. Naureen, M. Al-Nabhani, M.Z. Khan, A. Khan, A. Al-Harrasi, Robust Fourier transformed infrared spectroscopy coupled with multivariate methods for detection and quantification of urea adulteration in fresh milk samples, *Food Sci. Nutr.* 8 (10) (2019 Sep 30) 5249–5258, <https://doi.org/10.1002/fsn3.987>. PMID: 33133527; PMCID: PMC7590340.
- [13] F. Trimboli, N. Costanzo, V. Lopreato, C. Ceni, V.M. Morittu, A. Spina, D. Britti, Detection of buffalo milk adulteration with cow milk by capillary electrophoresis analysis, *J. Dairy Sci.* 102 (7) (2019 Jul) 5962–5970, <https://doi.org/10.3168/jds.2018-16194>. Epub 2019 May 2. PMID: 31056332.
- [14] S. Wang, H. Chen, B. Sun, Recent progress in food flavor analysis using gas chromatography-ion mobility spectrometry (GC-IMS), *Food Chem.* 315 (2020 Jun 15) 126158, <https://doi.org/10.1016/j.foodchem.2019.126158>. Epub 2020 Jan 7. PMID: 32014672.
- [15] Hernández B. Contla, N. Lopez-Villalobos, M. Vignes, Identifying health status in grazing dairy cows from milk mid-infrared spectroscopy by using machine learning methods, *Animals (Basel)*. 11 (8) (2021 Jul 21) 2154, <https://doi.org/10.3390/ani11082154>. PMID: 34438612; PMCID: PMC8388516.
- [16] P. Jaiswal, S.N. Jha, A. Borah, A. Gautam, M.K. Grewal, G. Jindal, Detection and quantification of soymilk in cow-buffalo milk using attenuated total reflectance fourier transform infrared spectroscopy (ATR-FTIR), *Food Chem.* 168 (2015 Feb 1) 41–47, <https://doi.org/10.1016/j.foodchem.2014.07.010>. Epub 2014 Jul 11. PMID: 25172681.
- [17] P. Jaiswal, S.N. Jha, J. Kaur, A. Borah, Detection and quantification of anionic detergent (lissapol) in milk using attenuated total reflectance-Fourier Transform Infrared spectroscopy, *Food Chem.* 221 (2017 Apr 15) 815–821, <https://doi.org/10.1016/j.foodchem.2016.11.095>. Epub 2016 Nov 22. PMID: 27979278.
- [18] G. Huang, L.M. Yuan, W. Shi, X. Chen, X. Chen, Using one-class autoencoder for adulteration detection of milk powder by infrared spectrum, *Food Chem.* 372 (2022 Mar 15) 131219, <https://doi.org/10.1016/j.foodchem.2021.131219>. Epub 2021 Sep 24. PMID: 34601417.
- [19] R.M. Balabin, S.V. Smirnov, Melamine detection by mid- and near-infrared (MIR/NIR) spectroscopy: a quick and sensitive method for dairy products analysis including liquid milk, infant formula, and milk powder, *Talanta* 85 (1) (2011 Jul 15) 562–568, <https://doi.org/10.1016/j.talanta.2011.04.026>. Epub 2011 Apr 19. PMID: 21645742.
- [20] K.M. Khan, H. Krishna, S.K. Majumder, et al., Detection of urea adulteration in milk using near-infrared Raman spectroscopy, *Food Anal. Methods* 8 (2015) 93–102, <https://doi.org/10.1007/s12161-014-9873-z>.
- [21] G. Abernethy, K. Higgs, Rapid detection of economic adulterants in fresh milk by liquid chromatography-tandem mass spectrometry, *J. Chromatogr. A* 1288 (2013 May 3) 10–20, <https://doi.org/10.1016/j.chroma.2013.02.022>. Epub 2013 Feb 16. PMID: 23540766.
- [22] V. Bonfatti, F. Tiezzi, F. Miglior, P. Carnier, Comparison of Bayesian regression models and partial least squares regression for the development of infrared prediction equations, *J. Dairy Sci.* 100 (2017) 7306–7319, <https://doi.org/10.3168/jds.2016-12203>.
- [23] C. Chu, H. Wang, X. Luo, P. Wen, L. Nan, C. Du, Y. Fan, D. Wang, Z. Yang, G. Yang, L. Liu, Y. Li, B. Hu, Z. Abula, S. Zhang, Possible alternatives: identifying and quantifying adulteration in Buffalo, goat, and camel milk using mid-infrared spectroscopy combined with modern statistical machine learning methods, *Foods* 12 (20) (2023 Oct 21) 3856, <https://doi.org/10.3390/foods12203856>. PMID: 37893749; PMCID: PMC10606090.

- [24] Y. Tamaki, G. Mazza, Rapid determination of lignin content of straw using fourier transform mid-infrared spectroscopy, *J. Agric. Food Chem.* 59 (2) (2011 Jan 26) 504–512, <https://doi.org/10.1021/jf1036678>. Epub 2010 Dec 22. PMID: 21175187.
- [25] P. Williams, P. Implementation of Near-Infrared Technology. *Near-Infrared Technology in the Agricultural and Food Industries, and ed., 2001.*
- [26] H. Toledo-Alvarado, A.I. Vazquez, G. de Los Campos, R.J. Tempelman, G. Bittante, A. Cecchinato, Diagnosing pregnancy status using infrared spectra and milk composition in dairy cows, *J. Dairy Sci.* 101 (3) (2018 Mar) 2496–2505, <https://doi.org/10.3168/jds.2017-13647>. Epub 2017 Dec 28. PMID: 29290427.
- [27] H. Soyeurt, F. Dehareng, N. Gengler, S. McParland, E. Wall, D.P. Berry, M. Coffey, P. Dardenne, Mid-infrared prediction of bovine milk fatty acids across multiple breeds, production systems, and countries, *J. Dairy Sci.* 94 (2011) 1657–1667.
- [28] G. Bittante, A. Cecchinato, Genetic analysis of the Fourier-transform infrared spectra of bovine milk with emphasis on individual wavelengths related to specific chemical bonds, *J. Dairy Sci.* 96 (9) (2013) 5991–6006.
- [29] K.E. Kaylegian, J.M. Lynch, J.R. Fleming, et al., Influence of fatty acid chain length and unsaturation on mid-infrared milk analysis1, *J. Dairy Sci.* 92 (6) (2009) 2485–2501.
- [30] P.M. Santos, E.R. Pereira-Filho, L.E. Rodriguez-Saona, Rapid detection and quantification of milk adulteration using infrared microspectroscopy and chemometrics analysis, *Food Chem.* 138 (1) (2013 May 1) 19–24, <https://doi.org/10.1016/j.foodchem.2012.10.024>. Epub 2012 Oct 30. PMID: 23265450.
- [31] C.S. Gondim, R.G. Junqueira, S.V.C. Souza, I. Ruisánchez, M.P. Callao, Detection of several common adulterants in raw milk by MID-infrared spectroscopy and one-class and multi-class multivariate strategies, *Food Chem.* 230 (2017 Sep 1) 68–75, <https://doi.org/10.1016/j.foodchem.2017.03.022>. Epub 2017 Mar 6. PMID: 28407966.
- [32] S. Shadpour, T.C.S. Chud, D. Hailemariam, G. Plastow, H.R. Oliveira, P. Stothard, J. Lassen, F. Miglior, C.F. Baes, D. Tulpan, F.S. Schenkel, Predicting methane emission in Canadian Holstein dairy cattle using milk mid-infrared reflectance spectroscopy and other commonly available predictors via artificial neural networks, *J. Dairy Sci.* 105 (10) (2022) 8272–8285, <https://doi.org/10.3168/jds.2021-21176>.
- [33] M. Manley, Near-infrared spectroscopy and hyperspectral imaging: non-destructive analysis of biological materials, *Chem. Soc. Rev.* 43 (24) (2014) 8200–8214, <https://doi.org/10.1039/c4cs00062e>.
- [34] M.K. Nieuwoudt, S.E. Holroyd, C.M. McGoverin, M.C. Simpson, D.E. Williams, Raman spectroscopy as an effective screening method for detecting adulteration of milk with small nitrogen-rich molecules and sucrose, = *Apr, J. Dairy Sci.* 99 (4) (2016) 2520–2536, <https://doi.org/10.3168/jds.2015-10342>. Epub 2016 Feb 10. PMID: 26874427.
- [35] M.K. Nieuwoudt, S.E. Holroyd, C.M. McGoverin, M.C. Simpson, D.E. Williams, Rapid, sensitive, and reproducible screening of liquid milk for adulterants using a portable Raman spectrometer and a simple, optimized sample well, = *Oct, J. Dairy Sci.* 99 (10) (2016) 7821–7831, <https://doi.org/10.3168/jds.2016-11100>. Epub 2016 Jul 27. PMID: 27474982.
- [36] T. Wu, H. Chen, Z. Lin, C. Tan, Identification and quantitation of melamine in milk by near-infrared spectroscopy and chemometrics, *Journal of Spectroscopy* 2016 (2016) 1–8, <https://doi.org/10.1155/2016/6184987>.
- [37] F. Mabood, L. Ali, R. Boque, G. Abbas, F. Jabeen, Q.M.I. Haq, J. Hussain, A.M. Hamaed, Z. Naureen, M. Al-Nabhani, M.Z. Khan, A. Khan, A. Al-Harrasi, Robust Fourier transformed infrared spectroscopy coupled with multivariate methods for detection and quantification of urea adulteration in fresh milk samples, *Food Sci. Nutr.* 8 (10) (2019 Sep 30) 5249–5258, <https://doi.org/10.1002/fsn3.987>. PMID: 33133527; PMCID: PMC7590340.
- [38] U. Kamboj, N. Kaushal, S. Mishra, N. Munjal, Application of selective near infrared spectroscopy for qualitative and quantitative prediction of water adulteration in milk, *Mater. Today: Proc.* 24 (2020) 2449–2456, <https://doi.org/10.1016/j.matpr.2020.03.775>.
- [39] M.A. Musa, S. Yang, Detection and quantification of cow milk adulteration using portable near-infrared spectroscopy combined with chemometrics African, *J. Agric. Res.* 17 (2) (2021) 198–207, <https://doi.org/10.5897/AJAR2020.15321>.
- [40] JC Jr Machado, M.A. Faria, I.M.P.L.V.O. Ferreira, R.N.M.J. Páscoa, J.A. Lopes, Varietal discrimination of hop pellets by near and mid infrared spectroscopy, *Talanta* 180 (2018 Apr 1) 69–75, <https://doi.org/10.1016/j.talanta.2017.12.030>. Epub 2017 Dec 14. PMID: 29332835.
- [41] R. Valand, S. Tanna, G. Lawson, L. Bengtström, A review of Fourier Transform Infrared (FTIR) spectroscopy used in food adulteration and authenticity investigations, *Food Addit. Contam. Part A Chem Anal Control Expo Risk Assess* 37 (1) (2020 Jan) 19–38, <https://doi.org/10.1080/19440049.2019.1675909>. Epub 2019 Oct 15. PMID: 31613710.
- [42] P.M. Santos, E.R. Pereira-Filho, L.E. Rodriguez-Saona, Application of hand-held and portable infrared spectrometers in bovine milk analysis, *J. Agric. Food Chem.* 61 (6) (2013 Feb 13) 1205–1211, <https://doi.org/10.1021/jf303814g>. Epub 2013 Jan 31. PMID: 23339381.
- [43] B. Balan, A.S. Dhulaniya, R. Jamwal, A. Yadav, S. Kelly, A. Cannavan, D.K. Singh, Rapid detection and quantification of sucrose adulteration in cow milk using Attenuated total reflectance-Fourier transform infrared spectroscopy coupled with multivariate analysis, *Spectrochim. Acta Mol. Biomol. Spectrosc.* 240 (2020 Oct 15) 118628, <https://doi.org/10.1016/j.saa.2020.118628>. Epub 2020 Jun 19. PMID: 32599485.
- [44] P.M. Santos, E.R. Pereira-Filho, L.E. Rodriguez-Saona, Rapid detection and quantification of milk adulteration using infrared microspectroscopy and chemometrics analysis, *May 1, Food Chem.* 138 (1) (2013) 19–24, <https://doi.org/10.1016/j.foodchem.2012.10.024>. Epub 2012 Oct 30. PMID: 23265450.
- [45] S.N. Jha, P. Jaiswal, A. Borah, A.K. Gautam, N. Srivastava, Detection and quantification of urea in milk using attenuated total reflectance-fourier transform infrared spectroscopy, *Food Bioprocess Technol.* 8 (4) (2014) 926–933, <https://doi.org/10.1007/s11947-014-1455-y>.
- [46] L. Du, W. Lu, B. Gao, J. Wang, L. Lucy Yu, Authenticating raw from reconstituted milk using fourier transform infrared spectroscopy and chemometrics, *J. Food Qual.* 2019 (2019) 1–6, <https://doi.org/10.1155/2019/5487890>.
- [47] V.B. Gupta Shrivastava, Methods for the determination of limit of detection and limit of quantitation of the analytical methods, *Chronicles Young Sci.* 2 (2011) 21.
- [48] T.M.P. Cattaneo, S.E. Holroyd, The use of near infrared spectroscopy for determination of adulteration and contamination in milk and milk powder: updating knowledge, *J. Near Infrared Spectrosc.* 21 (5) (2013) 341–349, <https://doi.org/10.1255/jnirs.1077>.
- [49] A. Hilding-Ohlsson, J.A. Fauerbach, N.J. Sacco, M.C. Bonetto, E. Cortón, Voltamperometric discrimination of urea and melamine adulterated skimmed milk powder, *Sensors* 12 (9) (2012) 12220–12234, <https://doi.org/10.3390/s120912220>. Epub 2012 Sep 5. PMID: 23112709; PMCID: PMC3478836.
- [50] R. Yang, R. Liu, G. Dong, K. Xu, Y. Yang, W. Zhang, Two-dimensional hetero-spectral mid-infrared and near-infrared correlation spectroscopy for discrimination adulterated milk, *Spectrochim. Acta Mol. Biomol. Spectrosc.* 157 (2016 Mar 15) 50–54, <https://doi.org/10.1016/j.saa.2015.12.017>. Epub 2015 Dec 15. PMID: 26714285.
- [51] M.C. Barreto, R.G. Braga, S.G. Lemos, W.D. Fragoso, Determination of melamine in milk by fluorescence spectroscopy and second-order calibration, *Food Chem.* 364 (2021 Dec 1) 130407, <https://doi.org/10.1016/j.foodchem.2021.130407>. Epub 2021 Jun 18. PMID: 34182362.
- [52] H. Jin, G.M. Dong, H.Y. Wu, Y.R. Yang, M.Y. Huang, M.Y. Wang, R.J. Yang, Identification of adulterated milk based on auto-correlation spectra, *Spectrochim. Acta Mol. Biomol. Spectrosc.* 286 (2023 Feb 5) 121987, <https://doi.org/10.1016/j.saa.2022.121987>. Epub 2022 Oct 17. PMID: 36265304.
- [53] P.R.A.B. De Toledo, A.T. Toci, H.R. Pezza, L. Pezza, Fast and simple method for identification of adulteration of cow's milk with urea using diffuse reflectance spectroscopy, *Anal. Methods* 9 (45) (2017) 6357–6364, <https://doi.org/10.1039/c7ay02354e>.
- [54] M.Y. Huang, R.J. Yang, Z.Y. Zheng, H.Y. Wu, Y.R. Yang, Discrimination of adulterated milk using temperature-perturbed two-dimensional infrared correlation spectroscopy and multivariate analysis, *Spectrochim. Acta Mol. Biomol. Spectrosc.* 278 (2022 Oct 5) 121342, <https://doi.org/10.1016/j.saa.2022.121342>. Epub 2022 May 4. PMID: 35550994.
- [55] F. Sharifi, M. Naderi-Boldaji, M. Ghasemi-Varnamkhasi, K. Kheiralipour, M. Ghasemi, A. Maleki, Feasibility study of detecting some milk adulterations using a LED-based Vis-SWIR photocooustic spectroscopy system, *Food Chem.* 424 (2023 Oct 30) 136411, <https://doi.org/10.1016/j.foodchem.2023.136411>. Epub 2023 May 19. PMID: 37229900.
- [56] S. Kasemsunran, W. Thanapase, A. Kiatsoonthon, Feasibility of near-infrared spectroscopy to detect and to quantify adulterants in cow milk, *Anal. Sci.* 23 (2007) 907–910, <https://doi.org/10.2116/analsci.23.907>.
- [57] R.F. Guan, D.H. Liu, X.Q. Ye, K. Yang, Use of fluorometry for determination of skim milk powder adulteration in fresh milk, *J. Zhejiang Univ. - Sci. B* 6 (11) (2005 Nov) 1101–1106, <https://doi.org/10.1631/jzus.2005.B1101>. PMID: 16252345; PMCID: PMC1390658.