sustainability

MDPI

Article

Utilization of FTIR and machine learning for evaluating gluten-free bread contaminated with wheat flour

Adedeji, A. Akinbode[1,*], Okeke, Abuchi[1] and Rady, Ahmed[2]

[1] Department of Biosystems and Agricultural Engineering, University of Kentucky, Lexington, KY, 40506, USA
[2] Food, Water, Waste Research Group, Faculty of Engineering, University of Nottingham, Nottingham, United Kingdom
* Correspondence: Author: akinbode.adedeji@uky.edu; 859-218-4355

Abstract: In this study, Fourier Transform Infrared (FTIR) spectroscopy coupled with machine learning (ML) approaches were applied to detect and quantify wheat flour (WF) contamination in gluten-free cornbread. Samples of corn flour (CF) were contaminated with WF in the range of 0 – 10 % with a 0.5 % increment. The flour samples were baked into bread using basic bread formulation and ground into fine particle size for homogeneity, and FTIR spectra of the ground samples were obtained and standardized before modeling. For constructing the classification model, majority voting-based ensemble learning (stack of k-nearest neighbor [KNN], random forest, and support vector classifier) was implemented to detect and quantify WF in the cornbread samples. KNN regressor was determined to the best predictive model to quantify wheat contaminant based on the majority-vote ensemble. The optimal classification model for the test set showed an F1 score, true positive rate (TPR), and false negative rate (FNR) of 1.0, 1.0, and 0.0, respectively. For the quantification models, the coefficient of determination and root mean square error for the prediction set (R^2_P and RMSEP) were 0.99 and 0.34, respectively. These results show the feasibility of utilizing FTIR along with supervised learning algorithms for rapid offline evaluation of wheat flour contamination in gluten-free products.

Keywords: Celiac Disease; Cornbread; Ensemble learning; Gluten; Machine learning; Wheat flour; FTIR

Citation: Adedeji; A. Akinbode1,*; Okeke; Abuchi1; Rady; Ahmed2 Utilization of FTIR and machine learning for evaluating gluten-free bread contaminated with wheat flour. *Sustainability* **2023**, *15*, x. https://doi.org/10.3390/xxxxx

Academic Editor(s):

Received: date
Accepted: date
Published: date

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2023 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Sustainable food production has become a global need for the current and future generations to lessen the need for pollution-generating, traditional food production. Food allergy is reported to affect 3 – 5 % of adults and 8 % of children worldwide [1]. The medical and caregiving cost of food allergies in 2013 in the US was estimated between $19 and $25 billion [2]. Gluten-free products have a vast number of customers and some of which cannot tolerate the presence of gluten in foods. Gluten proteins including gliadin and glutenin in wheat and whole wheat are the major allergens in wheat. Celiac disease, gluten intolerance or sensitivity are examples of the immunological or physiological consequences of consuming such allergens [3]. In 2012, it was estimated that 0.3-0.9% of the US population suffers from the celiac disease [4]. For people with gluten-related disorders, a strict gluten-free diet is essential to properly manage their reactions. The United States Food and Drug Administration (FDA) identified gluten-free diet as any food containing less than or equal to 20 ppm of gluten [5]. Cross-contamination in the food industry is not an unusual matter. The severity of such incidents can cause a further collapse of food businesses through product recalls and even legal disputes. Cross-contamination do occur in gluten-free production lines, which leads to the gluten ratio being higher than 20 ppm

in the final products. One of the common incidents of cross-contamination is caused by inadequacy of cleaning of processing lines that are used to manufacture gluten-containing (i.e., wheat, barley, rye, and their crossbred varieties), and gluten-free products [6, 7]. Cross-contamination in food factories can be identified as the accidental mixing of one or more ingredients of a food product [7]. Baked products are generally examples where cross-contamination develops through undeliberate mixing of raw materials with gluten-contained flours such as wheat, barley, or rye; or through the inadequate cleaning of processing lines that handle both gluten-contained and gluten-free products. In bakeries where bread is made, different gluten and gluten-free flours may be processed in the same factory; where the same equipment (e.g., milling machine, kneading rolls, fermenting drum, etc.) or kitchen utensils are used for both. The likelihood of cross-contamination exists if proper cleaning is not followed, or measures are taken to ascertain the gluten-free nature of the ingredients or processing lines. Food losses and/or food waste originated from cross-contaminated food represent a major problem for the food businesses as well as the environment and healthcare considering the need to hospitalize patients suffering from food allergens, foodborne, and other diseases resulting from food products that are adulterated or contaminated. Therefore, detection of cross-contamination in food processing lines is a key factor to ensure that the product labeled as gluten-free is suitable for consumption by consumers having gluten-related disorders.

Enzyme-linked immunosorbent assay (ELISA) is the most common analytical technique for identifying gluten in contaminated food products [8]. However, such a method is cumbersome and requires highly skilled personnel to conduct. The global market size value of gluten-free bakeries was estimated to be USD 1.81 billion, and total revenue is expected to reach USD 4.15 billion by 2030. The US share of such a market in 2022 was USD 449.0 million [9]. Thus, considering the relatively large production as well as the demand for gluten-free baked products, ensuring safety through fast and accurate inspection is crucial in all processing steps, which achieve the targets of quality control.

Nondestructive methods based on optical sensors have been successfully applied for the quality evaluation of perishable and processed food products [10]. Several studies researched the applications of optical sensing for determining the quality characteristics and/or gluten content of grains' flour and bread. Sørensen [11] investigated the evaluation of nutrients in a wide range of commercially-produced bread varieties using near-infrared spectroscopy (NIRS) which yielded results with a coefficient of determination (R^2) values ranging from 0.76 to 0.99 for the different food constituents analyzed. Identification of low gluten concentrations (< 4.5%, w/w) in gluten-free flour and the batter was conducted via NIRS (1100-2500 nm) and modified partial least squares (MPLS) and the validation models yielded validation models with R^2 values of up to 0.967 for the flour and 82.5% for the batter [12]. Ahmad, Nache [13] studied the application of fluorescence spectroscopy with an excitation parameter ($\lambda_{Excitation}$) of 270-550 nm, and the emission parameter ($\lambda_{Emission}$) of 310-590 nm to detect low gluten levels (< 5%, w/w) in commercial gluten-free flour. The partial least squares regression (PLSR) prediction models had R^2 values of 0.90 and root mean square error of prediction (RMSEP) of as low as 0.46%. Quantifying the total protein and wet gluten in wheat flour was investigated by utilizing NIRS (800-3030 nm) and synergy interval support vector regression (siSVR) and siPLSR along with different spectral preprocessing methods [14]. The deduced prediction models had as high R^2 as 0.906 for total protein and 0.850 for gluten, and as low RMSEP as 0.925 for total protein and 1.024 for gluten.

Among optical noninvasive sensors, Fourier-transform infrared (FTIR) spectroscopy provides rapid, accurate, and comprehensive identification and a presenting fingerprint of the chemical composition of the food material with minimal sample preparation [15]. FTIR surpasses other optical systems by replacing the prism with an interferometer, which helps resolve overlapping infrared bandwidth for complex samples [15, 16]. Fourier-trans-

form near-infrared (FT-NIR) (835-2502 nm) resulted in comparable results to NIRS instrumentation (450-2498 nm) for reliable prediction of quality attributes of wheat grain (protein, moisture, and hardness index) and flour (protein, ash, and amylose) with R^2 ranging from 0.80 to 0.99 for NIRS and 0.83 to for FT-NIR [17]. Raman spectroscopy (300-3700 cm^{-1}), diffuse reflectance-FT-NIR (3700-9000 cm^{-1}), and attenuated total reflectance (ATR) FTIR (400-4000 cm^{-1}) coupled with PLSR were compared to develop prediction models for gluten in wheat flour [18]. The models developed by cross-validation showed higher correlation coefficient (r) values obtained from FTIR (0.962) compared with diffuse reflectance-FR-NIR (0.937), and Raman (0.9058) sensors. However, the relative standard error of prediction (RSEP) for the independent test set was higher for the FTIR (5.69); whereas the RSEP values were 3.54 for the diffuse reflectance-FT-NIR, and 3.24 for the Raman. Applications of FTIR in the food domain also included assessing cholesterol [19] calcium [20], and lactose [21] in powdered milk; authentication of tea varieties [22]; assessing the quality of green tea [23], black tea [24]; talcum [25], and polyphenols [26] in tea powers; and determining antioxidant capacity [27], adulteration [28]; evaluating the adulteration [29], and defects [30] in ground and/or roasted coffee. To our knowledge, no previous study searched the utilization of FTIR for determining gluten in gluten-free bread. Therefore, the objective of this study was to develop an FTIR spectroscopy methodology that utilizes supervised machine learning algorithms to detect and quantify wheat flour contamination in gluten-free cornbread. Advanced machine learning algorithms that include ensemble learning were explored in this study.

2. Materials and Methods

2.1. Ingredients for breadmaking

Corn flour (CF) and wheat flour (WF) used in this study were purchased from Bob's Red Mill Natural Foods (Milwaukie, OR, US). The formulation used with the basic bread ingredient was adopted from [31]. The cornbread formulation includes corn flour (100 %) and the following ingredients based on the corn flour weight: water (70 %), dried yeast (2 %), salt (2 %), sugar (2 %), vegetable fat (3 %); and 0 – 10 % wheat flour was used for contamination at 0.5 % increment.

2.2. Cornbread baking

The CF was mixed with the aforementioned formulated ingredients at the WF contamination levels. A kitchen mixer (KitchenAid, Model KV25G0X, Benton Harbor, MI, USA) was used for bread dough mixing for two speed levels. The first level had 60 rpm speed and the mixing was for 1 min. The second level had a speed of 95 rpm and a mixing time of 6 min. The dough was scrapped every 2 min. The dough was poured into aluminum baking pans and proofed for 35 min at 40°C, then baked for 1 h at 190.6 °C in a convection oven (Hobart, HR202, OH, U.S.A). This was conducted as a result of the batter-like system of the dough due to its consistency. The baked bread loaves were kept for 1 h at room temperature (24 °C) to cool, then ground to fine particle size (< 150 μm) using a commercial laboratory blender (Waring Commercial 7010BU Lab Blender) for 40 s to obtain a homogenous mixture. In the end, 21 different bread samples (20 g each) were obtained based on the WF addition levels.

2.3. FTIR spectra data pre-processing

FTIR spectra measurements were carried out on the ground cornbread samples using Attenuated Total Reflectance Fourier Transform Infrared (ATR-FTIR) spectrometer (Nicolet iS50, ThermoFisher Scientific, Massachusetts, USA). The spectra were acquired in the range of 4000 – 450 cm^{-1} with a spectral resolution of 4 cm^{-1} comprising 887 wavenumbers. Each sample was scanned 32 times and the average spectrum was recorded. The data obtained were then pre-processed using the standard scaling (SC) method or standardization [32]. The SC is obtained by subtracting the mean of the feature vectors (μ) from every

data point (X) and then dividing each column by the corresponding element in the vector's standard deviation (σ). Generally, standard scaling makes the data increasingly interpretable, because the normal estimation of Y when x (the mean or centered X) is zero represents the expected value of Y when X is at its mean with a standard deviation of 1. This transforms the data to have a resulting distribution of a mean of 0 and a standard deviation of 1 [33].

$$x_{ij} = \frac{(X_{ij} - \mu_{ij})}{\sigma_{ij}} \quad (4.1)$$

For the classification and prediction approaches, the data was split into a training set (70%) and a test set (30%).

2.4. Development of machine learning models

Classification and regression models were developed using scikit learn 0.22.2 (machine learning module in python) which is an open-source robust library that provides a range of python-based supervised and unsupervised learning algorithms with the capability to deploy machine learning models from prototypes to production systems.

2.4.1. Feature engineering

The spectra data obtained from the ATR-FTIR is considered a high-dimensional dataset where the number of features is relatively larger than the number of samples. Such data also contains redundant features, which increases the propensity of overfitting the model. Therefore, feature reduction was implemented using principal component analysis (PCA). PCA is based on mapping the original domain of the features into a new set of uncorrelated features called principal components (PCs). The number of PCs is then selected based on the desired maximum amount of variance explained which was chosen to be 100 % of the total variance [33, 34].

2.4.2. Classification analysis

To detect whether a bread sample is contaminated with wheat flour during the baking process, a classification model was developed by training different individual classifiers and applying an ensemble learning technique. The ensemble learning method incorporates the combination of different learning algorithms, usually weak, to obtain a higher-performance meta-model. Experimental evidence indicates that ensemble learning often yields more accurate classification than a single learning algorithm [35]. In this study, majority voting-based ensemble learning was utilized. Different supervised learning algorithms were stacked including random forest (RF), support vector machine (SVM), and k-nearest neighbor (KNN). Each individual classifier model was trained using 70 % of the dataset and then made a classification (vote) on the test (30 % of the dataset) instances. The sample was assigned to the class that received more than half of the votes. This was similar to the method used by Bouziane, Messabih [36] to predict protein secondary structure and was found to outperform individual classifiers. The model was evaluated based on the confusion matrix parameters obtained with emphasis on the false negative rate (FNR), true positive rate (TPR), and the F1 score [33]. These parameters are defined as:

$$\mathrm{TPR} = \frac{\mathrm{TP}}{(\mathrm{TP}+\mathrm{FN})} \quad (4.2)$$

$$\mathrm{FNR} = \frac{\mathrm{FN}}{(\mathrm{FN}+\mathrm{TP})} \quad (4.3)$$

$$\mathrm{F1\ score} = 2\frac{P*R}{P+R} \quad (4.4)$$

where TP = true positive, FN = false positive, P = precision, and R = recall.

2.4.3. Prediction of gluten level in contaminated cornbread

Prediction models were developed using the ensemble learning concept mentioned in section 2.4.2 but for the regression analysis. Several individual regression models were merged and the best-performing model is selected as in section 2.4.2 using regression metrics stated later in this section. Supervised learning models for regression used in this project included k-nearest neighbors (KNN) regressor, random forest (RF) regressor, decision tree (dct) regressor, SVM regressor, and partial least square regressor (PLSR) [33]. The efficiency of prediction models was evaluated by assessing the coefficient of determination and root mean square error for the training set (R^2_T, RMSET), and the test or prediction set (R^2_P, RMSEP). Thus, the best training model was characterized by higher R^2_T and the lower root means square error RMSET. Cross-validation was performed to tune and determine the optimal hyper-parameter values of the training models, and their learning curves were obtained [33]. This was to ensure that the models were not under or overfitting the data. The optimized training model was then applied to the test or prediction set and the values of R^2_P, and RMSEP were reported.

3. Results and Discussion

3.1. FTIR spectra characteristics of the ground cornbread samples

Figures 1a and 1b display the FTIR spectra of the samples contaminated with 0.5 % wheat flour when it is in raw form (flour), 1a, and after being baked (bread), 1b. It is noted that there were evident differences in the absorbance values when comparing the figures at the region of 1860 cm^{-1} – 1480 cm^{-1}, which includes most of the amide I (1690 cm^{-1} – 1600 cm^{-1}), and amide II (1580 cm^{-1} – 1480 cm^{-1}) characteristic bands that are sensitive to the secondary structure content of proteins. These regions maintain a smooth formation with unique peaks due to CO- and NH+ bonds or other potential (CC and CN) stretching vibrations in Figure [37]. However, in Figure 1b the spectra have noisy deformation within the aforementioned regions and this may be due to protein denaturation due to heat application during the baking process and other conversion processes such as mixing with other ingredients (e.g. salt) [38]. Other differences could be seen in the intensities of the peak and trough. Figure 1c shows the spectra of baked cornbread contaminated with different levels of wheat flour (0.5 - 10%) which is similar to the baked cornbread spectra in Figure 1b.

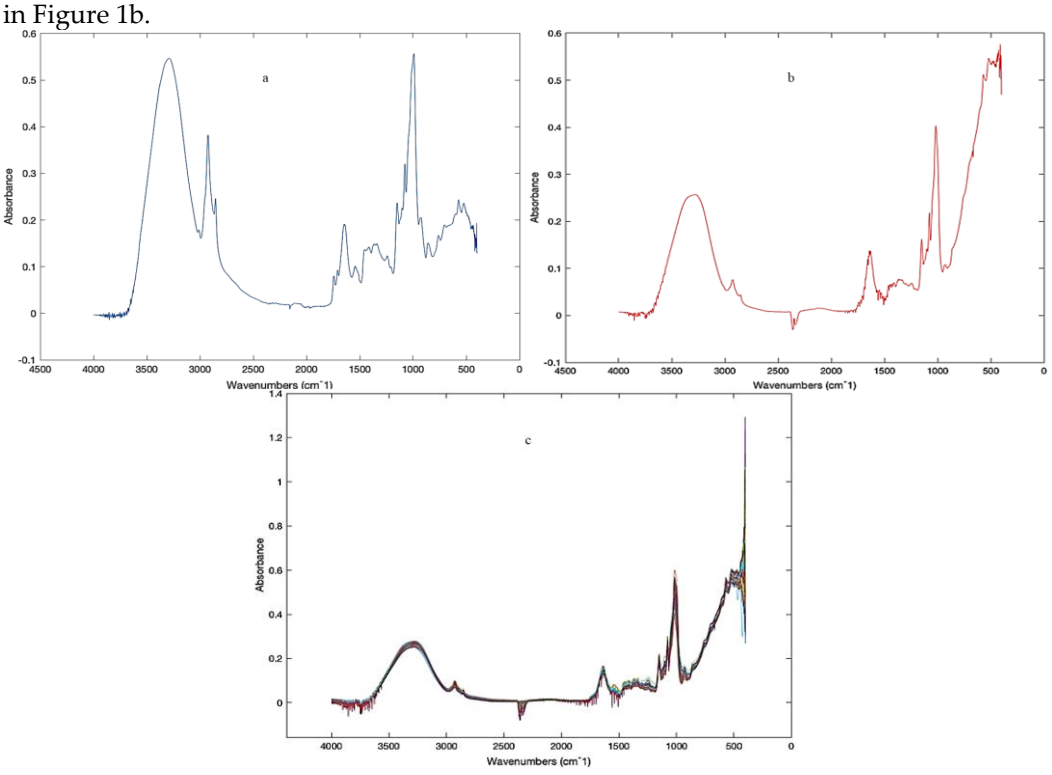


Figure 1. (a) FTIR-spectra of the corn flour contaminated with 0.5% wheat flour (raw sample); (b) FTIR-spectra of baked cornbread contaminated with 0.5 % wheat flour; (c) FTIR-spectra of baked cornbread mixed with different levels of wheat flour (0.5 – 10 % at 0.5 % increment).

3.2. Classification analyses results

The application of PCA reduced the number of features from 887 to 20 principal components (PCs), which explained all (100 %) of the variance in the data (Figure 2). The reduced features (i.,e. PCs) were utilized to develop a binary classification models that included two classes: class 1 (pure cornbread) and class 2 (contaminated cornbread with wheat). Among all the classifier methods used including RF classifier, SVM classifier, KNN classifier, and majority voting-based ensemble learning (that

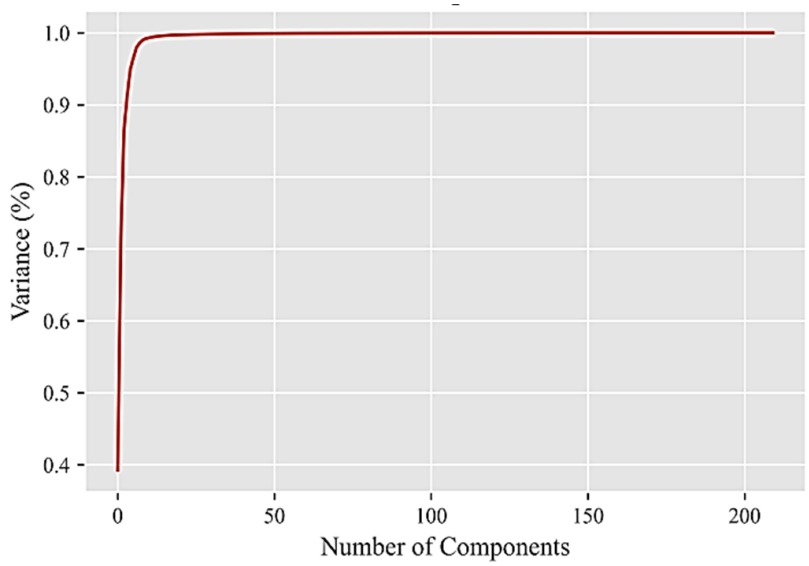


Figure 2. A plot of the number of principal components (PCs) as a function of the variance explained in the samples.

involves stacking the individual learning algorithms), the ensemble method outperformed individual classifiers, which is clearly illustrated by the confusion matrix which was applied to assess the predictive capabilities of classification models [33, 39, 40]. Tables 1 and 2 present the confusion matrix and its parameters obtained for the training dataset (70 % of the dataset). The false negative rate (0), true-positive rate (1), and F1-score (1) values obtained indicate a perfect performance of the model to identify the contaminated samples. Table 3 and Table 4 present the confusion matrix parameters obtained after the model was applied to the separate test data that contributed to 30% of the all samples. The model was able to accurately classify samples belonging to each of the classes with a TPR, FNR, and F1-score of 1, 0, and 1, respectively [33]. One possible reason for the perfect classification for both the training and test set data is the fewer classes (two) in this study. However, the result shows the ability of the ensemble classifiers to learn every feature in the binary classes of the samples used, also when subjected to the test set data the performance was reliable and consistent. There is empirical evidence in the literature that ensemble learning methods adjust for the deficiencies of individual classifiers delivering improved and more accurate results in most cases it is applied [41]. Kim et al.[42], reported that the ensemble learning improved the performance of machine learning in predicting the sugar content in various citrus species. Also, Oztuk et al. [43], reported that the application of the voting-based ensemble learning method improved the detection of 19 classes of grain flours studied under rotational motion.

Table 1. Confusion matrix parameters for the majority voting-based learning classification training model.

Class|Metric	TPR	FPR	TNR	FNR	Err	P	F1_score
Class 1	1.0	0.0	1.0	0.0	0.0	1.0	1.0
Class 2	1.0	0.0	1.0	0.0	0.0	1.0	1.0

TPR: True positive rate, FPR: False positive rate, TNR: True negative rate, FNR: False negative rate, Err: Error, P: Precision, F1: scores for a measure of accuracy. Class 1: No Contamination, Class 2: Contaminated with Wheat.

Table 2. Confusion matrix for the majority voting-based ensemble learning. classification training model

Actual Class	Class 1	Class 2
Classified as Class 1	138	0
Classified as Class 2	0	142
Classified as Unassigned	0	0

Class 1: No Contamination, Class 2: Contaminated with Wheat.

Table 3. Confusion matrix parameters for the classification test model.

Class|Metric	TPR	FPR	TNR	FNR	Err	P	F1_score
Class 1	1.0	0.0	1.0	0.0	0.0	1.0	1.0
Class 2	1.0	0.0	1.0	0.0	0.0	1.0	1.0

TPR: True positive rate, FPR: False positive rate, TNR: True negative rate, FNR: False negative rate, Err: Error, P: Precision, F1: scores for a measure of accuracy. Class 1: No Contamination, Class 2: Contaminated with Wheat.

Table 4. Confusion matrix for the majority voting-based ensemble learning classification test model.

Actual Class	Class 1	Class 2
Classified as Class 1	62	0
Classified as Class 2	0	58
Classified as Unassigned	0	0

Class 1: No Contamination, Class 2: Contaminated with wheat

Table.5. Prediction analysis on a different learning algorithm.

Learning Algorithm	**Hyper-parameters**	**Training**		**Prediction**	
		R^2_T	**RMSET**	**R^2_P**	**RMSEP**
Random Forest (rf)	n_estimators = 991	1.00	0.00	0.516	2.064
K-Nearest Neighbors (KNN)	neighbors = 4, metric = 'manhattan'	0.98	0.41	0.99	0.34

Decision Tree (dct)	max_depth = 6	0.99	0.29	0.57	1.93
Support Vector Machine (svr)	gamma = 0.03	0.90	0.98	0.75	1.47
Partial Least Square Regression (PLSR)	n_components = 30	0.99	0.08	0.97	0.52
Ensemble Method (voting)	(rf, KNN, dct, svr, wt = none)	0.99	1.29	0.81	1.29

Wt: weight, n_estimators: no. of estimators, n_components: no. of components, R^2_T: R^2 of training, R^2_P: R^2 of prediction, RMSET: root mean square error of training, RMSEP: root mean square error of prediction

3.3. Prediction of wheat level in cornbread

Table 5 presents the evaluation parameters for the predictive learning algorithm used including RF regressor, KNN regressor, Decision trees, SVM regressor, PLSR, and ensemble learning. The results for KNN and PLSR are very close in performance with vales of R^2_T = 0.98 (KNN), 0.99 (PLSR), R^2_P = 0.99 (KNN), 0.97 (PLSR), and RMSET = 0.41 (KNN), 0.08 (PLSR), and RMSEP = 0.[illegible] (KNN), 0.52 (PLSR), respectively [33]. The results indicate that both learning algorithms have the potential to quantify the level of the wheat flour contaminant in the cornbread samples within the percentage levels used. It is worth stating that KNN regressor resulted in lower error value for the prediction set compared with the PLSR model. Thus, based on both the values of R^2_P (0.99) and RMSEP (0.[illegible] KNN was selected as the best-performing learning algorithm for predicting wheat flour contamination levels in baked cornbread. Figures 3A to 3E show the learning curves obtained as a function of the number of the hyper-parameters tuned for the individual algorithms [33]. The aim from such a function was to choose the optimal hyper-parameters that yield a balance between the bias and variance in other to prevent overfitting [33]. As seen in Figure 3 for the best learning algorithm (KNN), the learning gap between the scores measured on the training and cross-validation set is minimized and insignificant. It was observed that as the number of neighbors in the KNN algorithm increases, the changes remain constant until approaching the value of 6 where the scores started dropping leading to a lower prediction accuracy of the model [33]. Therefore, we can conclude from this that the number of neighbors ranging from 1 to 5 is more effective for the KNN model to predict or quantify the percentage contamination of the wheat flour in the cornbread [33].

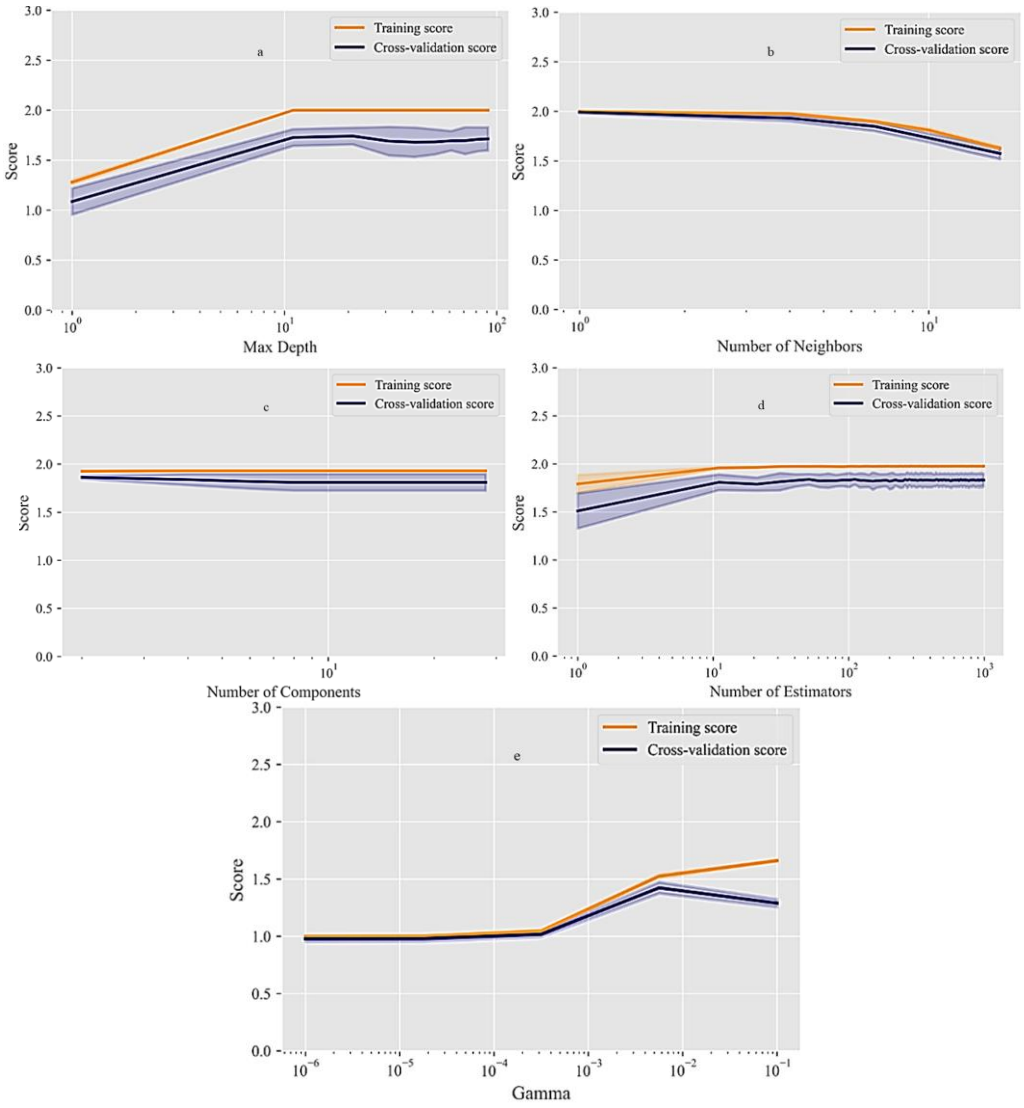


Figure 3. Validation curve for decision tree regressor (A), KNN (B), PLSR (C), random forest regressor (D), and support vector machine (E).

The learning gap between the scores measured on the training and cross-validation set reached the minimum or optimum level. It was observed that as we increase the number of neighbors for the algorithm the changes remain constant until approaching the value of 6 where the scores started dropping leading to a lower accuracy of the model. Therefore, we can conclude that the number of neighbors ranging from 1 to 5 is more effective for the KNN model to predict or quantify the percentage contamination of the wheat flour in the cornbread [33]. Cui et al. [44] followed a similar approach for optimizing the hyperparameters (i.e. gagging fraction and frequency) for the gradient boosting decision tree ensemble learning to enhance regression models based on FTIR data to predict the decontamination rate of cold plasma-treated Chitosan-DNA films.

4. Conclusion

FTIR spectroscopy has always played an important role in the food industry with regard to food safety inspection and quality assessment. In this study, we used FTIR spectroscopy coupled with supervised machine learning approaches to detect and quantify wheat flour (gluten) contamination in the range of 0.5 - 10% in gluten-free cornbread that falls within the threshold of 20 ppm of containment that elicits an allergenic reaction in wheat/gluten sensitive individual. This study shows that the ensemble learning method

performed best than the rest of the individual supervised ML algorithms in detecting wheat/gluten contaminants in gluten-free cornbread. The KNN regressor emerged as the most promising technique in quantifying the level of the WF contamination with a prediction coefficient of determination and root mean squared error of 0.99 and 0.34, respectively. Therefore, the results obtained from this study indicate the potential and the effectiveness of FTIR spectroscopy with supervised learning techniques for rapid authenticating of cornbread contaminated with what flour. More generally, the study can indeed help develop a protocol for quality assurance in bakery products factories. This help reduces cross-contamination in the baking industry and consequently provides less food waste for the currently critical food supply chain that dramatically suffers from high prices of energy, lack of water resources, domestic disputes, and migration. A future research consideration is the development of a handheld system and/or a mobile phone application that would allow stakeholders and consumers to scan their foods to run a quick test for wheat/gluten allergen contamination.

Author Contributions: Conceptualization, A.A.A.; methodology, A.A.A., A.O and A.R; software, A.O and A.R; validation, A.A.A.; formal analysis, A.O.; investigation, A.O. and A.A.A.; resources, A.A.A.; data curation, A.O.; and University of Kentucky; writing—original draft preparation, A.O. and A.A.A.; writing—review and editing, A.A.A. and A.R.; visualization, A.O.; supervision, A.A.A. and A.R.; project administration, A.A.A.; funding acquisition, A.A.A.,. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Institute of Food and Agriculture (NIFA) U.S.D.A. Multistate Project #: 1024529

Data Availability Statement: All data used in this project belong to the United State government and the administering institution, University of Kentucky (UK), and can be requested through the UK Library.

Acknowledgments: The authors would like to acknowledge the Kentucky Agricultural Experiment Station and for supporting and sponsoring this work.

Conflicts of Interest: The authors declare no conflict of interest.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

References

1. Gendel, S.M., *Comparison of international food allergen labeling regulations.* Regulatory Toxicology and Pharmacology, 2012. **63**(2): p. 279-285.DOI:Â https://doi.org/10.1016/j.yrtph.2012.04.007.
2. Croote, D. and S.R. Quake, *Food allergen detection by mass spectrometry: the role of systems biology.* NPJ systems biology and applications, 2016. **2**(1): p. 1-10.DOI:Â doi:10.1038/npjsba.2016.22;.
3. Murray, J.A., *Gluten - sensitive enteropathy*, in *Food Allergy: Adverse Reactions to Foods and Food Additives*, D.D. Metcalfe, H.A. Sampson, and R.A. Simon, Editors. 2008. p. 211-222.
4. Leonard, M.M., et al., *Celiac disease and nonceliac gluten sensitivity: a review.* Jama, 2017. **318**(7): p. 647-656.DOI:Â doi: 10.1001/jama.2017.97.
5. Allred, L.K. and B.W. Ritter, *Recognition of gliadin and glutenin fractions in four commercial gluten assays.* Journal of AOAC International, 2010. **93**(1): p. 190-196.DOI:Â DOI:10.1093/JAOAC/93.1.190.

6. Wieser, H., et al., *Food safety and cross-contamination of gluten-free products: A narrative review.* Nutrients, 2021. **13**(7): p. 2244.DOI:Â doi: 10.3390/nu13072244.
7. Taylor, S.L. and J.L. Baumert, *Cross-contamination of foods and implications for food allergic patients.* Current allergy and asthma reports, 2010. **10**: p. 265-270.DOI:Â doi: 10.1007/s11882-010-0112-4.
8. Osorio, C.E., J.H. Mejías, and S. Rustgi, *Gluten detection methods and their critical role in assuring safe diets for celiac patients.* Nutrients, 2019. **11**(12): p. 2920.DOI:Â Gluten detection methods and their critical role in assuring safe diets for celiac patients.
9. Research, G.V. *Gluten-free Bakery Market Size, Share & Trends Analysis Report By Product (Biscuits & Cookies, Bread, Cakes), By Distribution Channel (Online, Supermarkets & Hypermarkets), By Region, And Segment Forecasts, 2022 - 2030*. 2023 [cited 2023 03/02/2023]; Available from: https://www.grandviewresearch.com/industry-analysis/gluten-free-bakery-market-report#:~:text=North%20America%20dominated%20the%20gluten,due%20to%20rising%20health%20concerns.
10. José Blasco, E.M.G. and a.C.Z. Da-Wen Sun, *Vision Systems*, in *Optical monitoring of fresh and processed agricultural crops*, M. Zude, Editor. 2008, CRC press: Boca Raton, FL, USA.
11. Sørensen, L., *Application of reflectance near infrared spectroscopy for bread analyses.* Food Chemistry, 2009. **113**(4): p. 1318-1322.DOI:Â https://doi.org/10.1016/j.foodchem.2008.08.065.
12. Albanell, E., B. Miñarro, and N. Carrasco, *Detection of low-level gluten content in flour and batter by near infrared reflectance spectroscopy (NIRS).* Journal of Cereal Science, 2012. **56**(2): p. 490-495.DOI:Â https://doi.org/10.1016/j.jcs.2012.06.011.
13. Ahmad, M.H., M. Nache, and B. Hitzmann, *Potential of fluorescence spectroscopy in detection of low-levels of gluten in flour: A preliminary study.* Food Control, 2017. **73**: p. 401-405.DOI:Â https://doi.org/10.1016/j.foodcont.2016.08.028.
14. Chen, J., S. Zhu, and G. Zhao, *Rapid determination of total protein and wet gluten in commercial wheat flour using siSVR-NIR.* Food Chemistry, 2017. **221**: p. 1939-1946.DOI:Â https://doi.org/10.1016/j.foodchem.2016.11.155.
15. Wenning, M. and S. Scherer, *Identification of microorganisms by FTIR spectroscopy: perspectives and limitations of the method.* Applied microbiology and biotechnology, 2013. **97**: p. 7111-7120.DOI:Â https://doi.org/10.1007/s00253-013-5087-3.
16. Rodriguez-Saona, L.E. and M.E. Allendorf, *Use of FTIR for rapid authentication and detection of adulteration of food.* Annual review of food science and technology, 2011. **2**: p. 467-483.DOI:Â https://doi.org/10.1146/annurev-food-022510-133750.
17. Armstrong, P., et al., *Comparison of dispersive and Fourier-transform NIR instruments for measuring grain and flour attributes.* Applied Engineering in Agriculture, 2006. **22**(3): p. 453-457.DOI:Â doi: 10.13031/2013.20448.
18. Czaja, T., S. Mazurek, and R. Szostak, *Quantification of gluten in wheat flour by FT-Raman spectroscopy.* Food Chemistry, 2016. **211**: p. 560-563.DOI:Â https://doi.org/10.1016/j.foodchem.2016.05.108.
19. Chitra, J., M. Ghosh, and H. Mishra, *Rapid quantification of cholesterol in dairy powders using Fourier transform near infrared spectroscopy and chemometrics.* Food Control, 2017. **78**: p. 342-349.DOI:Â https://doi.org/10.1016/j.foodcont.2016.10.008.
20. Wu, D., et al., *Determination of calcium content in powdered milk using near and mid-infrared spectroscopy with variable selection and chemometrics.* Food and Bioprocess Technology, 2012. **5**: p. 1402-1410.DOI:Â https://doi.org/10.1007/s11947-010-0492-4.
21. Lei, Y., et al., *Analysis of crystallized lactose in milk powder by Fourier-transform infrared spectroscopy combined with two-dimensional correlation infrared spectroscopy.* Journal of Molecular Structure, 2010. **974**(1-3): p. 88-93.DOI:Â https://doi.org/10.1016/j.molstruc.2009.12.030.
22. Wu, X., et al., *Discrimination of tea varieties using FTIR spectroscopy and allied Gustafson-Kessel clustering.* Computers and Electronics in Agriculture, 2018. **147**: p. 64-69.DOI:Â https://doi.org/10.1016/j.compag.2018.02.014.
23. Ikeda, T., et al., *Prediction of Japanese green tea ranking by Fourier transform near-infrared reflectance spectroscopy.* Journal of agricultural and food chemistry, 2007. **55**(24): p. 9908-9912.DOI:Â https://doi.org/10.1021/jf0717642.
24. Ren, G., et al., *Quantitative analysis and geographical traceability of black tea using Fourier transform near-infrared spectroscopy (FT-NIRS).* Food Research International, 2013. **53**(2): p. 822-826.DOI:Â https://doi.org/10.1016/j.foodres.2012.10.032.

25. Li, X., Y. Zhang, and Y. He, *Rapid detection of talcum powder in tea using FT-IR spectroscopy coupled with chemometrics.* Scientific Reports, 2016. **6**(1): p. 30313.DOI:Â https://doi.org/10.1038/srep30313.
26. Li, X., et al., *Determination of tea polyphenols content by infrared spectroscopy coupled with iPLS and random frog techniques.* Computers and Electronics in Agriculture, 2015. **112**: p. 28-35.DOI:Â https://doi.org/10.1016/j.compag.2015.01.005.
27. Batista, N.N., et al., *Antioxidant capacity of cocoa beans and chocolate assessed by FTIR.* Food Research International, 2016. **90**: p. 313-319.DOI:Â https://doi.org/10.1016/j.foodres.2016.10.028.
28. Teye, E., et al., *Feasibility study on the use of Fourier transform near-infrared spectroscopy together with chemometrics to discriminate and quantify adulteration in cocoa beans.* Food research international, 2014. **55**: p. 288-293.
29. Reis, N., A.S. Franca, and L.S. Oliveira, *Performance of diffuse reflectance infrared Fourier transform spectroscopy and chemometrics for detection of multiple adulterants in roasted and ground coffee.* LWT-Food Science and Technology, 2013. **53**(2): p. 395-401.DOI:Â https://doi.org/10.1016/j.lwt.2013.04.008.
30. Craig, A.P., et al., *Fourier transform infrared spectroscopy and near infrared spectroscopy for the quantification of defects in roasted coffees.* Talanta, 2015. **134**: p. 379-386.DOI:Â https://doi.org/10.1016/j.talanta.2014.11.038.
31. Mondal, A. and A. Datta, *Bread baking–A review.* Journal of food engineering, 2008. **86**(4): p. 465-474.DOI:Â https://doi.org/10.1016/j.jfoodeng.2007.11.014.
32. Varmuza, K. and P. Filzmoser, *Introduction to multivariate statistical analysis in chemometrics*. 2009: CRC press.DOI:Â https://doi.org/10.1201/9781420059496.
33. Okeke, A., *Fourier Transform Infrared Spectroscopy (As A Rapid Method) Coupled With Machine Learning Approaches For Detection And Quantification Of Gluten Contaminations In Grain-Based Foods*, in *Biosystems and Agricultural Engineering* 2020, University of Kentucky: Lexington.DOI: https://doi.org/10.13023/etd.2020.353.
34. Howley, T., et al. *The effect of principal component analysis on machine learning accuracy with high dimensional spectral data.* in *International Conference on Innovative Techniques and Applications of Artificial Intelligence*. 2005. Springer.
35. Dietterich, T.G., *Ensemble learning.* The handbook of brain theory and neural networks, 2002. **2**: p. 110-125.
36. Bouziane, H., B. Messabih, and A. Chouarfia, *Profiles and majority voting-based ensemble method for protein secondary structure prediction.* Evolutionary Bioinformatics, 2011. **7**: p. EBO. S7931.DOI:Â https://doi.org/10.4137/EBO.S7931.
37. Jabs, A., *Determination of secondary structure in proteins by fourier transform infrared spectroscopy (FTIR).* Jena Library of Biologica Macromolecules, 2005.
38. Neill, G., H. Ala'a, and T. Magee, *Optimisation of time/temperature treatment, for heat treated soft wheat flour.* Journal of Food Engineering, 2012. **113**(3): p. 422-426.
39. Jeong, S., et al., *Unravelling the physicochemical features of US wheat flours over the past two decades by machine learning analysis.* LWT, 2022. **169**: p. 114036.DOI:Â https://doi.org/10.1016/j.lwt.2022.114036.
40. Marom, N.D., L. Rokach, and A. Shmilovici. *Using the confusion matrix for improving ensemble classifiers*. in *2010 IEEE 26-th Convention of Electrical and Electronics Engineers in Israel*. 2010. IEEE.DOI:Â 10.1109/EEEI.2010.5662159.
41. Olenskyj, A.G., et al., *End-to-end prediction of uniaxial compression profiles of apples during in vitro digestion using time-series micro-computed tomography and deep learning.* Journal of Food Engineering, 2022. **325**: p. 111014.DOI:Â https://doi.org/10.1016/j.jfoodeng.2022.111014.
42. Kim, S.-Y., et al., *Application of ensemble neural-network method to integrated sugar content prediction model for citrus fruit using Vis/NIR spectroscopy.* Journal of Food Engineering, 2023. **338**: p. 111254.DOI:Â https://doi.org/10.1016/j.jfoodeng.2022.111254.
43. Ozturk, S., et al., *Near-infrared spectroscopy and machine learning for classification of food powders during a continuous process.* Journal of Food Engineering, 2023. **341**: p. 111339.DOI:Â https://doi.org/10.1016/j.jfoodeng.2022.111339.

44. Cui, H., et al., *DNA-based surrogates for the validation of microbial inactivation using cold atmospheric pressure plasma and plasma-activated water processing*. Journal of Food Engineering, 2023. **339**: p. 111267.DOI:Â https://doi.org/10.1016/j.jfoodeng.2022.111267.