MÜHENDİSLİK ve DOĞA BİLİMLERİ FAKÜLTESİ

Uluslararası Mühendislik Araştırma ve Geliştirme Dergisi
International Journal of Engineering Research and Development

UMAGD, (2024) 16(1), 221-226.

10.29137/umagd.1379171

Cilt/Volume:16 Sayı/Issue:01 Ocak/January 2024

Araştırma Makalesi / Research Article

Classification of Starch Adulteration in Milk Using Spectroscopic Data and Machine Learning

Spektroskopik Veri ve Makine Öğrenimi Kullanarak Sütteki Nişasta Sahteciliğinin Sınıflandırılması

Yeliz Durgun[*1]

Department of Biomedical Equipment Technology Program. Turhal Vocational School, 60300, Tokat, TURKEY

***Başvuru/Received**: 20/10/2023* ***Kabul/Accepted**: 23/12/2023* ***Çevrimiçi Basım/Published Online:** 31/01/2024*

***Son Versiyon/Final Version**: 31/01/2024*

Öz

In this comprehensive research, an in-depth evaluation of several machine learning algorithms, including Multilayer Perceptron, IBk, KStar, M5Rules, and RandomForest, is conducted to ascertain their effectiveness in detecting adulteration in milk products using spectroscopic data. The algorithms were rigorously deployed and assessed through a series of controlled experiments involving both raw and adulterated milk samples. Notably, IBk and KStar algorithms emerged with a perfect accuracy rate of 100% in identifying adulteration, highlighting their superior capability in this domain. Additionally, the Decision Table algorithm also performed exceptionally well, achieving a remarkable correlation coefficient of 0.9871. These promising results emphasize the undeniable potential of machine learning algorithms as reliable and precise tools for detecting adulteration in milk. Such technological interventions play a critical role in elevating the safety and quality standards of milk and milk-based products in the market. Moreover, the deployment of these advanced machine-learning techniques provides an invaluable layer of consumer protection, plays a significant role in combating widespread fraudulent practices in the milk industry, and ensures compliance with stringent food safety standards. These methodologies could be indispensable for both industry players and regulatory bodies, significantly contributing to the safeguarding of public health.

Key Words

"Spectroscopy, Classification Models, Milk Adulteration, Food Safety"

Abstract

Bu kapsamlı araştırmada, Çok Katmanlı Algılayıcı, IBk, KStar, M5Rules ve RandomForest gibi çeşitli makine öğrenimi algoritmaları, spektroskopik veri kullanarak süt ürünlerindeki sahteciliği tespit etme etkinlikleri açısından derinlemesine değerlendirilmiştir. Algoritmalar, ham ve sahte süt örnekleri üzerinde yapılan kontrollü deneyler aracılığıyla titizlikle uygulanmış ve değerlendirilmiştir. Özellikle, IBk ve KStar algoritmaları, sahteciliği tespit etmede % 100'lik mükemmel bir doğruluk oranı ile öne çıkmıştır, bu da bu alandaki üstün yeteneklerini vurgulamaktadır. Ek olarak, Karar Tablosu algoritması da son derece iyi bir performans göstererek, 0.9871'lik dikkate değer bir korelasyon katsayısı elde etmiştir. Bu umut verici sonuçlar, makine öğrenimi algoritmalarının sütteki sahteciliği tespit etme konusunda güvenilir ve kesin araçlar olarak inkar edilemez potansiyellerini vurgulamaktadır. Bu tür teknolojik müdahaleler, piyasada bulunan süt ve süt ürünlerinin güvenlik ve kalite standartlarını yükseltmede kritik bir rol oynamaktadır. Ayrıca, bu gelişmiş makine öğrenimi tekniklerinin uygulanması, tüketicilere değerli bir koruma katmanı sağlamakta, süt endüstrisinde yaygın sahtecilik uygulamalarıyla mücadelede önemli bir rol oynamakta ve katı gıda güvenliği standartlarına uyumu sağlamaktadır. Bu yöntemler, endüstri oyuncuları ve düzenleyici kurumlar için vazgeçilmez olabilir, kamusal sağlığın korunmasına önemli ölçüde katkı sağlayabilir.

Anahtar Kelimeler

"Spektroskopi , Sınıflandırma Modelleri, Süt Sahteciliği, Gıda Güvenliği"

***Sorumlu Yazar:** yeliz.durgun@gop.edu.tr*

1. Giriş

Food fraud represents a global economic and food safety issue for consumers, industry, and governments (Spink et al., 2016; Spink et al., 2019).This situation arises when food products, their components, or packaging are deliberately added, altered, or tampered with for the purpose of economic gain (Spink and Moyer, 2011a ; Manning and Soon,2016) . Food fraud is perpetrated to increase the apparent value of the product or reduce production costs and can lead to acute health consequences, including allergies, illnesses, or even deaths resulting from mislabeling (Pointing et al., 2020 ; Sammut et al., 2021 ; Visciano and Schirone, 2021; Spink and Moyer, 2011b). The prevalence and lack of control over adulteration in the milk and dairy products sector lead to the introduction of low-quality products into the market (Thangaraju, Modupalli, and Natarajan, 2021 ; Nacul and Revoredo-Giha, 2022).Quick detection of adulteration in these products is of critical importance. Typical adulterants found in milk can be categorized into four main groups: protein-based, nitrogen-based, carbohydrate-based adulterants, and chemical preservatives (Dugyala, Pradhan, and Basavaraj, 2023a). Carbohydrate-based adulterants can increase the total solid content of milk; substances like starch not only increase the apparent protein, fat, and lactose content of the milk but also mask the reduction in specific gravity.

Starch masks the reduction in specific gravity caused by the dilution of water, and is therefore used as a secondary adulterant to water (Singh and Gandhi, 2015; Dugyala, Pradhan, and Basavaraj, 2023b ; Nascimento et al., 2017).This characteristic makes starch adulteration more lucrative compared to other carbohydrate-based milk adulterants, as it can conceal the effects of dilution and partial removal of native macromolecules (e.g., fat) in milk. While the motivation behind starch adulteration is economic gain, its consequences have negative effects on consumer health (Banti, 2020a ; Banti, 2020b).Excessive starch in milk can cause diarrhea due to undigested residues in the intestines. In diabetic individuals, the accumulation of starch in the body can lead to serious health issues and even death. Therefore, the adulteration and accurate detection of starch in milk are crucial for global health and nutrition (Bojarczuk et al., 2022 ; Weinstein et al., 1961 ; Reddy, Venkatesh, and Reddy, 2017).

In various studies conducted on starch adulteration in milk, different methods have been used for its detection. The Iodine test, one of the most commonly used methods, works by observing the formation of a starch-iodine complex in milk containing starch.

This reaction turns the milk into a blue-black color, indicating the presence of starch (Chauhan et al., 2019). However, besides the Iodine test, there are also potentiometric and amperometric detection as well as iodine titration (Banks, Greenwood, and Muir, 1971) and near-infrared spectroscopy (Borin et al., 2006) such alternative methods have also been used for starch detection.These methods have some shortcomings, such as long sample preparation times, complex instrument requirements, the use of harmful chemicals, and challenges in data interpretation (Jha et al., 2016). In light of these shortcomings, it is of great importance to develop simple, cost-effective, and quantitative methods for the effective detection of starch in milk. In this context, new approaches have been developed for the detection of adulterants in milk through deposit modeling on solid surfaces. In studies examining sessile drop evaporation, a liquid drop is deposited on a solid surface, forming a convex dome (Sadek et al., 2015), it has been used as a model system for detecting soluble and insoluble adulterants in milk (Harindran, Hashmi, and Madhurima, 2022 ; Kumar and Dash, 2021). These studies demonstrate the diversity of methodologies used for starch detection in milk and the ongoing dynamism in research in this field.

In this comprehensive study, the efficacy of various machine learning algorithms in detecting adulteration in dairy products is meticulously evaluated. A series of sophisticated algorithms, including Multilayer Perceptron, IBk, KStar, M5Rules, and RandomForest, have been carefully deployed and examined through a series of experiments conducted on raw and adulterated milk samples.Our hypothesis is that these machine learning algorithms can achieve high accuracy rates in detecting adulteration in milk. This hypothesis has been logically developed based on data and discussions in previous literature, and this study empirically evaluates the effectiveness of the mentioned algorithms in adulteration detection.

Specifically, it has been determined that the IBk and KStar algorithms stand out for precisely detecting adulteration with a 100% accuracy rate, excelling in this area.Additionally, the Decision Table algorithm also deserves special mention; this algorithm possesses a unique ability to precisely and accurately classify samples, with a noteworthy correlation coefficient of 0.9871.

2.Materials and Methods

2.1 Experimental design

In our study, spectroscopic data derived from raw milk and milk samples adulterated with starch at various ratios were meticulously analyzed. For each milk sample, hundreds of spectral data points were gathered, spanning wavelengths from 410nm to 940nm.

Measurements were taken from different points within the liquid under mild vibrations to ensure homogeneous sampling from each specimen.

Our specialized mobile spectral setup, depicted in Figure 1, is specifically designed for the rapid analysis of milk samples and incorporates the AS7265x family of sensors. This sensor suite comprises three chips, with each chip hosting six channels, collectively delivering 18 VIS and NIR channels that cover wavelengths from 410nm to 940nm with a Full Width at Half Maximum (FWHM) of 20nm.

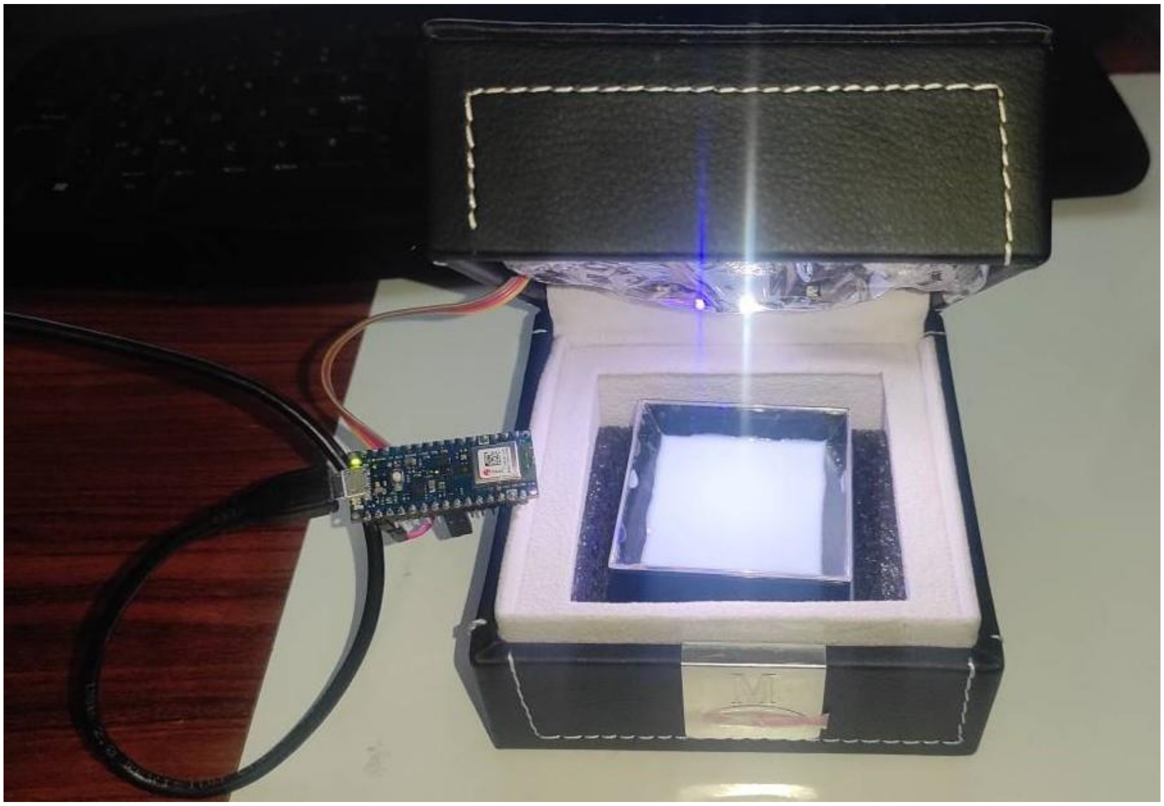

Figure 1. The Mobile Spectral Setup

The AS7265x family features interference filters directly integrated onto standard CMOS silicon and each sensor in the set is equipped with an integrated microcontroller, boasting low power consumption. Communication with these sensors can be established via UART or I2C interfaces, and each sensor comes with two integrated LED drivers.

The measurements were conducted inside a light-proof box measuring 12cm x 12cm x 10cm. The AS7265x sensor suite was positioned approximately 1 cm above the milk samples. Data acquisition was facilitated using an Arduino Nano 33 BLE. The acquired spectroscopic data underwent statistical analyses, encompassing discriminant analysis to differentiate between various milk samples.

2.2 Statistical analysis

In the present study, we utilized both raw milk and milk samples adulterated with varying ratios of starch for analysis. Discrimination of the samples was accomplished using the WEKA machine learning application (Machine Learning Group, University of Waikato). Differences in the spectroscopic data between raw and adulterated milk samples were meticulously analyzed, with the procedural flowchart presented in Figure 2.

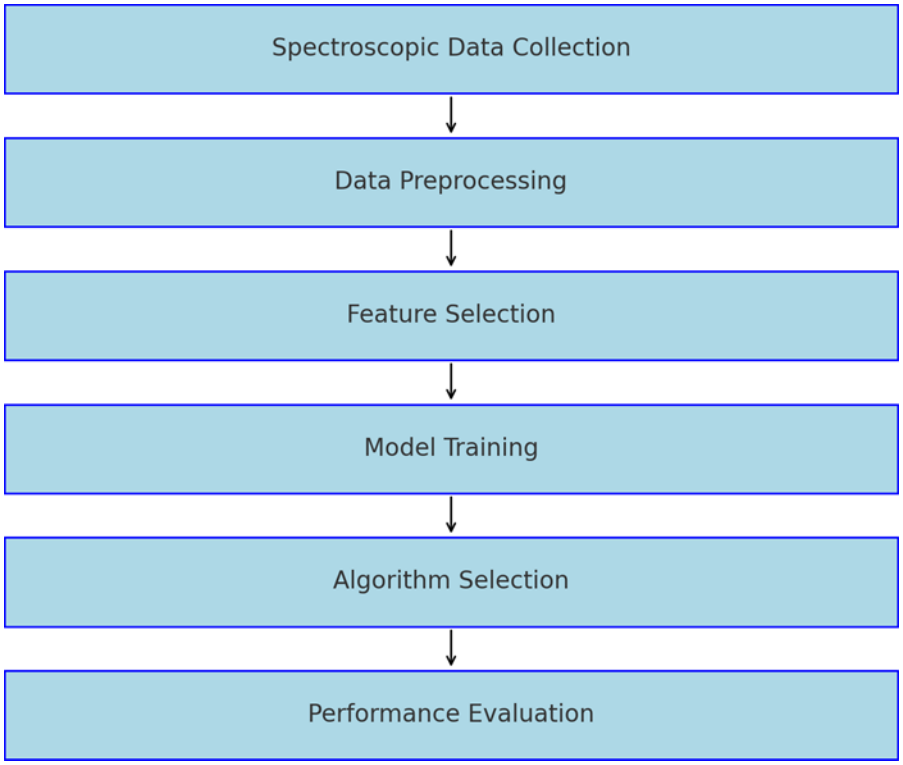

Figure 2. Flowchart Outlining The Process for Discriminating Between Raw and Starch- Adulterated Milk Samples Using Spectroscopic Data and Machine Learning Algorithms.

Upon acquisition of the spectroscopic data, the initial step of the analysis involved attribute selection, undertaken using the Ranker search method coupled with the OneR Attribute Evaluator. This process facilitated the identification and selection of spectroscopic data with maximal discriminative power between the sample types. Discriminative models were subsequently developed based on the selected features, employing a tenfold cross-validation mode. Machine learning algorithms from the Functions, Trees, Rules, and Lazy groups were utilized for this purpose. Within each group, algorithms that yielded the most satisfactory discrimination performance metrics were chosen for the study.

The outcomes of the analyses were documented as confusion matrices, encompassing accuracy for each sample, average accuracy, time elapsed for model construction, Kappa statistic, mean absolute error, root mean squared error, and relative absolute error. These performance metrics were computed utilizing the WEKA application.

3. Results and Discussion

The raw milk and adulterated milk samples were unequivocally distinguished through the developed models: Multi-Class Classifier from the MultilayerPerceptron group, IBk and KStar from the Lazy group, M5Rules from the Rules group, and RandomForest from the Trees group (refer to Table 1). All models achieved high accuracy levels in detecting adulteration, with IBk and KStar models reaching 100% accuracy. The IBk and KStar algorithms obtained a perfect Kappa statistic of 1.0, demonstrating completely accurate classifications. These algorithms also recorded zero mean absolute error, root mean squared error, and relative absolute error. The model generated using the IBk algorithm boasted the shortest training time at virtually zero seconds, whereas the model constructed with the KStar algorithm took the longest time, clocking in at 1.43 seconds, which is still remarkably swift.

Table 1. Performance Metrics of Various Classification Algorithms on Milk Adulteration Detection (Approximately 100% Accuracy)

Algorithm	**Time to Build Model(s)**	**Classification Accuracy**	**Mean Absolute Error(%)**	**Root Mean Squared Error**	**Relative Absolute Error (%)**	**Kappa Statistic**
Multilayer Perceptron	0.87	100	0.0691	0.0858	5.54	0.999
IBk	0.00	100	0.0000	0.0000	0.00	1.000
KStar	1.43	100	0.0000	0.0000	0.00	1.000
M5Rules	0.06	99.91	0.0333	0.0629	2.67	0.9991
RandomForest	0.07	99.96	0.0092	0.0385	0.73	0.9996

The Multilayer Perceptron,M5Rules, and RandomForest algorithms also provided high classification accuracies above 99.9%, with Kappa statistics nearly reaching perfection.

The raw milk and adulterated milk samples were unequivocally distinguished using the developed models: Gaussian Processes and Linear Regression from the Functions group, Locally Weighted Learning (LWL) from the Lazy group, and Decision Table from the Rules group (refer to Table 2). Each model exhibited different levels of accuracy in classifying the samples.

The Decision Table algorithm achieved an extraordinary correlation coefficient of 0.9871, underscoring its capability to classify samples with precision. It also recorded the lowest mean absolute error, root mean squared error, and relative absolute error among the four models. Both Gaussian Processes and Linear Regression models provided strong classification results with correlation coefficients above 0.95, whereas the LWL model is noteworthy for its rapid model-building time. The specific requirements and constraints of the task should guide the selection of the appropriate model for classifying milk samples.

Table 2. Performance Metrics of Various Classification Algorithms on Milk Adulteration Detection (Approximately 95% Accuracy)

Algorithm	Time to Build Model(s)	Classification Accuracy	Mean Absolute Error(%)	Root Mean Squared Error	Relative Absolute Error (%)	Kappa Statistic
Multilayer Perceptron	0.87	100	0.0691	0.0858	5.54	0.999
IBk	0.00	100	0.0000	0.0000	0.00	1.000
KStar	1.43	100	0.0000	0.0000	0.00	1.000
M5Rules	0.06	99.91	0.0333	0.0629	2.67	0.9991
RandomForest	0.07	99.96	0.0092	0.0385	0.73	0.9996

4. Conclusions

Certain and unequivocal distinctions were made between raw and adulterated milk samples through the utilization of developed models. Initially, the Multi-Class Classifier from the Multilayer Perceptron group, IBk and KStar from the Lazy group, M5Rules from the Rules group, and RandomForest from the Trees group were employed (refer to Table 1). All these models achieved high accuracy levels in detecting adulteration in milk, with IBk and KStar models reaching a 100% accuracy rate.

The IBk and KStar algorithms obtained a perfect Kappa statistic of 1.0, demonstrating completely accurate classifications. These algorithms also recorded zero mean absolute error, root mean squared error, and relative absolute error. The model generated using the IBk algorithm boasted the shortest training time, taking virtually zero seconds, whereas the model constructed with the KStar algorithm documented the longest training time, clocking in at 1.43 seconds. The Multilayer Perceptron, M5Rules, and RandomForest algorithms also provided high classification accuracies above 99.9%, with Kappa statistics nearly reaching perfection.

Secondly, the Gaussian Processes and Linear Regression models from the Functions group, Locally Weighted Learning (LWL) from the Lazy group, and Decision Table from the Rules group were examined (refer to Table 2). Each of these models exhibited different levels of accuracy in classifying the samples.

The Decision Table algorithm achieved an extraordinary correlation coefficient of 0.9871, underscoring its capability to classify samples with precision. It also recorded the lowest mean absolute error, root mean squared error and relative absolute error among the four models. Both Gaussian Processes and Linear Regression models provided strong classification results with correlation coefficients above 0.95, whereas the LWL model is noteworthy for its rapid model-building time.

These findings demonstrate the effectiveness and accuracy of various machine learning algorithms used in detecting adulteration in milk. The results indicate that these algorithms can be utilized to ensure milk safety and quality, offer safe products to consumers, and prevent fraud in the milk industry. Future studies plan on testing more algorithms and evaluating their capacity to detect adulterations in different milk products.

References

Banks, W, C T Greenwood, & D D Muir. (1971) “The Characterization of Starch and Its Components. Part 3. The Technique of Semi-Micro, Differential, Potentiometric, Iodine Titration, and the Factors Affecting It.” Starch-St{ä}rke 23 (4): 118–124.

Banti, Misgana. (2020a). “Food adulteration and some methods of detection, review.”International Journal of Nutrition and Food Sciences 9 (3): 86–94.

Bojarczuk, Adrianna, Sylwia Sk, Amin Mousavi Khaneghah, & Krystian Marsza. (2022). “Health benefits of resistant starch: A review of the literature.” Journal of functional foods 93:105094.

Borin, Alessandra, Marco Flores Ferrao, Cesar Mello, Danilo Althmann Maretto, & Ronei Jesus Poppi. (2006). "Least-squares support vector machines and near infrared spectroscopy for quantification of common adulterants in powdered milk." Analytica chimica acta 579 (1): 25–32.

Chauhan, Sneh Lata, Kruti Debnath Mandal Priyanka, Babul Rudra Paul, & Chinmoy Maji. (2019). "Adulteration of milk: A Review." International Journal of Chemical Studies 7 (1): 2055–2057.

Dugyala, Venkateshwar Rao, Shantanu Pradhan, & Madivala G Basavaraj. (2023a). "Sessile drop evaporation approach to detect starch adulteration in milk." Food Control 143:109272.

Harindran, Aswini, Sabin Hashmi, & V Madhurima. (2022). "Pattern formation of dried droplets of milk during different processes and classifying them using artificial neural networks." Journal of Dispersion Science and Technology 43 (12): 1838–1847.

Jha, Shyam Narayan, Pranita Jaiswal, Manpreet Kaur Grewal, Mansha Gupta, & Rishi Bhardwaj. (2016). "Detection of adulterants and contaminants in liquid foods—a review." Critical reviews in food science and nutrition 56 (10): 1662–1684.

Kumar, Virkeshwar, & Susmita Dash. (2021). "Evaporation-based low-cost method for the detection of adulterant in milk." ACS omega 6 (41): 27200–27207.

Manning, Louise & Jan Mei Soon. (2016). "Food safety, food fraud, and food defense: a fast evolving literature." Journal of food science 81 (4): 823–834.

Nacul, Hasibi Zavala, & Cesar Revoredo-Giha. (2022). "Food safety and the informal milk supply chain in Kenya." Agriculture & Food Security 11 (1): 8.

Nascimento, Carina F, Poliana M Santos, Edenir Rodrigues Pereira-Filho, & Fábio RP Rocha. (2017). "Recent advances on determination of milk adulterants." Food chemistry 221:1232–1244.

Pointing, John, Yunes Ramadan Al-Teinaz, John Lever, Mary Critchley, & Stuart Spear.(2020). "Food fraud." The Halal Food Handbook, pp. 321–329.

Reddy, D Maheswara, K Venkatesh, & C Venkata Sesha Reddy. (2017). "Adulteration of milk and its detection: a review." International Journal of Chemical Studies 5 (4): 613–617.

Sadek, Celine, Pierre Schuck, Yannick Fallourd, Nicolas Pradeau, Cecile Le Floch-Fouere, & Romain Jeantet. (2015). "Drying of a single droplet to investigate process-structure-function relationships: a review." Dairy Science and Samp; Technology 95 (6): 771–794.

Sammut, Jesmond, Karthik Gopi, Neil Saintilan, & Debashish Mazumder. (2021).Facing the challenges of food fraud in the global food system.

Singh, Parminder, & Neeraj Gandhi. (2015). "Milk preservatives and adulterants: processing, regulatory and safety issues." Food Reviews International 31 (3): 236–261.

Spink, John, Brian Bedard, John Keogh, Douglas C Moyer, Joe Scimeca, & Akhila Vasan. (2019). "International survey of food fraud and related terminology: Preliminary results and discussion." Journal of food science 84 (10): 2705–2718.

Spink, John, Neal D Fortin, Douglas C Moyer, Hong Miao, & Yongning Wu. (2016). "Food fraud prevention: policy, strategy, and decision-making-implementation steps for a Government Agency or industry." Chimia 70 (5): 320–320.

Spink, John, &Douglas C Moyer. (2011a). "Defining the public health threat of food fraud." Journal of food science 76 (9): 157–163.

Spink, John, & Douglas C Moyer. (2011b). "Defining the public health threat of food fraud." Journal of food science 76 (9): 157–163.

Thangaraju, Suka, Nikitha Modupalli, & Venkatachalapathy Natarajan. (2021). "Food adulteration and its impacts on our health/balanced nutrition." Food Chemistry: The Role of Additives, Preservatives and Adulteration, pp. 189–216.

Visciano, Pierina, & Maria Schirone. (2021). "Food frauds: Global incidents and misleading situations." Trends in Food Science & Technology 114:424–442.

Weinstein, Louis, Robert E Olson, Theodore B Van Itallie, Elizabeth Caso, Doris Johnson, & Franz J Ingelfinger. (1961). "Diet as related to gastrointestinal function." JAMA 176 (11): 935–941.