JFBE 6(2): 56-62, 2023

Journal of Food and Bioprocess Engineering

Journal homepage: https://jfabe.ut.ac.ir

Original research

Lime juice adulteration detection by spectroscopy and machine learning

Zahra Alaei Roozbahani [a, c], Mohsen Labbafi [a,*] Ali Aghakhani1[a], Saeed Izadi [b]

[a] *1Department of Food Science and Technology, College of Agriculture and Natural Resources, University of Tehran, Karaj, Iran*
[b] *Simon Fraser University, Canada*
[c] *Department of Food Science and Technology Standard Research Institute (SRI), Karaj, Iran*

ABSTRACT

Fruit juices, and especially lime juice, belong to the most targeted food commodities for fraud. Therefore, reliable and cost-effective analytical methodologies need to be developed to guarantee lime juice authenticity and quality. The manifestation of machine learning techniques (MLT) has paved the way for fast and reliable processing and analysis of food and juice data for more effective use of inexpensive, readily available, and easy-to-use equipment such as UV/Vis spectrometers for quality control. The study aimed to investigate UV/Vis spectrometry and MLT to detect at least 10% of water, acid, and sugar added to lime juice. For this purpose, 26 lime samples, including Mexican and Persian lime, were collected from the orchards of four main lime-cultivated areas in Iran to prepare pure lime juice samples (as authentic samples). To investigate adulterated lime juice, four types of treatment were defined by adding acid, sugar, a mix of acid and sugar solution, and water at different volume proportions (10, 20, 30, 40, and 50 % v/v) to pure lime juice samples. Each treatment was repeated eight times. The absorption rate of different adulterated and pure lime juice samples was measured at different wavelengths in the 210–550 nm range. The evaluation results of different MLTs showed that the accuracy of separating samples using absorption data by decision tree (DT), k-nearest neighbor (k-NN), random forest (RF), multilayer perceptron (MLP), and support vector machine (SVM) were 75%, 79%, 80%, 87%, and 92%, respectively. SVM had the highest level of accuracy in separating adulterated lime juice samples. Also, this model's performance criteria (sensitivity and F-score) were higher than other models for identifying adulterated samples using absorption data. This is the first time that the common adulterations in lime juice were identified by rapid and accessible screening methods using UV/Vis spectroscopy and MLT with high accuracy, precision, and sensitivity.

Keywords: lime juice; UV/Vis spectroscopy; machine learning; SVM; MLP; k-NN

Received 26 Aug 2023; Revised 13 Sep 2023; Accepted 16 Nov 2023

1. Introduction

Lime juice is a rich source of nutrients such as flavonoids, vitamin C, minerals, and organic acids, which significantly affect human health (González-Molina et al., 2010; Jandrić & Cannavan, 2017; Sanches et al., 2022). For the industrial production of lime juice, 'Mexican' lime (*Citrus aurantifolia*) and 'Persian' lime (*Citrus latifolia*) is used, which are grown in very few areas due to their high sensitivity to low temperatures. Iran is ranked the 8th producer among the top 10 producers in the world, with a production of 3,450,000 tons of citrus fruits per year (FAO, 2020).

Juices are on the list of 10 food products at risk of adulteration (Dasenaki & Thomaidis, 2019). Food adulteration negatively impacts consumers' and producers' nutrition, health, and economy (Chang et al., 2016; Jandrić & Cannavan, 2017). Therefore, the introduction and guarantee of the quality and health of juices are essential. Although, authentication and identification of juice adulteration are complicated due to the varieties, cultivation regions, production methods, and adulteration techniques (Dasenaki & Thomaidis, 2019).

According to studies conducted in the identification of adulteration and authentication of lime juice, we can refer to a study published, in 2018, the quality of extracts of two varieties of 'key' and 'kaffir' lime was investigated, and their aromatic compounds were compared by device two-dimensional gas chromatography coupled with time-of-flight mass spectrometry (GC × GC-TOF-MS)

*Corresponding author.
E-mail address: mlabbafi@ut.ac.ir (M. Labbafi).
https://doi.org/10.22059/JFABE.2023.364398.1151

(Lubinska-Szczygieł et al., 2018). In another study, researchers used high-performance liquid chromatography coupled with isotope ratio mass spectrometry (HPLC-CO-IRMS) of carbon 13 to carbon 12, to investigate the type of acid, sugar, and geographical origin of lime and lemon juices (Guyon et al., 2014). In 2022, high-performance liquid chromatography coupled with mass spectrometry (HPLC-MS) has been proposed to authenticate the main acids of lime juice (AliAbadi et al., 2022). Shafiee and Minaei presented a method to distinguish samples of synthetic lime juice from natural lime juice using Near-Infrared Spectroscopy (NIR) and machine learning science. According to this study, the support vector machine (SVM) classified synthetic and natural lime juice using NIR spectra (Shafiee & Minaei, 2018).

Identification of adulteration and authentication is one of the most important aspects of food quality control. Most proposed methods require expensive equipment that is not easily available in most regions (Fidelis et al., 2017). Due to the demanding food market, using a simple, fast, accessible, and inexpensive method in quality control laboratories is much needed. Hence, spectroscopic techniques, such as UV coupled with multivariate analysis methods such as principal component analysis (PCA), principal component regression (PCR), and partial least squares regression (PLSR), have been focused on investigating various food products and juices in recent years (Dasenaki & Thomaidis, 2019; Ropodi et al., 2016). This approach has been considered to authenticate Tequila (Pérez-Caballero et al., 2017), vinegar (Ríos-Reina, Azcarate, Camiña, Callejón, & Amigo, 2019), pomegranate juice (Boggia et al., 2013), and apple juice (Chang et al., 2016).

Conventional statistical analysis methods for data are based on a data model that is limited by the assumption of a specific model. However, machine learning can extract information directly from the data itself and provide a more accurate representation of the natural data mechanism. Despite the high performance of machine learning algorithms, only a few studies have used machine learning algorithms to authenticate fruit juices. (Dasenaki & Thomaidis, 2019). So far, spectroscopy and machine learning tools have been coupled in only a few studies to determine water, sugar, and acids added to lime juice. The objective of this research is to identify adulteration in the lime juice (water, acid, and sugar added at least 10%) by using UV/Vis spectrophotometer coupled with unsupervised learning techniques (PCA) and supervised learning techniques (k-nearest neighbor (k-NN), RF, DT, SVM, and MLP).

2. Material and Methods

2.1. Samples

Adulterated samples were prepared by adding different dosages of sugar and acid as the most common type of adulteration in lime juice. For this purpose, three solutions were prepared by D(+)-Glucose-monohydrate, Sucrose (extra pure), and D(-) Fructose (99%) of Merck company, citric acid (99.5%) of SIGMA-ALDRICH company, and the water in our laboratory. These solutions were prepared according to the concentration of the component in natural lime juice for preparing adulterated samples, as follows:

Solution α: 6% w/w citric acid in distilled water.

Solution β: 1.5% w/w glucose, 1.5% w/w fructose, and 0.5% w/w sucrose in distilled water.

Solution γ: 6% w/w citric acid, 1.5% w/w glucose, 1.5% w/w fructose, and 0.5% sucrose in distilled water.

26 lime samples, including Mexican and Persian, were collected in two consecutive years, 2018 and 2019 from the orchards of 4 main lime-cultivated areas in Iran (Jahrom, Darab, Rodan, and Dezful). Conventional methods in the industry prepared juice samples and then stored at -18°C until the tests were performed (Lorente et al., 2014). All 26 samples of natural lime juice were considered authentic samples. 8 natural lime juice (one Mexican and one Persian of each region) were selected randomly and then diluted by adding 10, 20, 30, 40, and 50 percent of solution α to have 40 adulterated samples. In the same way, 120 more adulterated samples were prepared using distilled water, Solution β, and Solution γ separately.

Ultimately, this study was conducted by a total of 186 samples in five groups, including 160 adulterated samples prepared by adding distilled water (A1), solution α (A2), solution β (A3), and solution γ (A4), along with 26 natural samples of the primary lime juice (A5).

2.2. spectrophotometric Measurements

The samples were centrifuged for 15 minutes at 3500 rpm by Heraeus (Labofuge 400) instrument, and then they were diluted 40 times using distilled water to prevent spectra saturation. Since samples should be transparent as much as possible to decrease light scattering, remained pulp was removed using Whatman 42 filter paper after dilution. UV/Vis spectra of samples were recorded in 190-800 nm with a sensitivity of 1 nm using Lambada 25 spectrometer (Perkin-Elmer Co. USA). Juice spectra were obtained using the standard rectangular quartz cuvettes with a path length of 10 mm in a cell volume of 3.5 mL in two replicates. The average of replicates was recorded and then used as data for analysis in the next steps (Boggia et al., 2013).

2.3. Statistical analysis

In the first step, the model may overfit the data if we have redundant variables, leading to narrower prediction intervals and biased predictions. Hence absorption at 551-800 nm was removed for all samples due to the insignificant absorption, and the considered range was reduced to 210-550 nm.

In the second stage, an unsupervised learning technique, PCA, was used to assess the data obtained from the sample absorption analysis. PCA explores the data to identify the possible clusters and reduce variables of the experiment based on its correlation towards observation (called PC). As the most used tool in non-target analysis, PCA makes it possible to visualize a two-dimensional data matrix comprehensively, reducing the visualization dimensions (Ropodi et al., 2016). First, we performed weighted generalized least squares (GLS) (Manson) procedure (alpha 0.03) to deal with artifacts caused by changes in spectroscopic instrumentation, which typically require the development of entirely new calibration models. Then, we executed PCA.

The data were normalized using the Euclidean norm formula by the L2 method of Python programs so that the sum of the absorption squares of each sample is equal to one. The achieved data was used in the classification investigation part. k-NN, DT, RF, MLP, and SVM were applied as supervised learning techniques to investigate and compare the sensitivity and specificity of classification. Each algorithm has its advantages and disadvantage in a specific case. The

k-NN, as a nonlinear classification, needs memory intensive because it must keep track of all training data and find the neighbor nodes. It is slow due to executing expensive real-time. It works better linear regression when the data signal-to-noise ratio is high. DT supports nonlinearity and is faster than k-NN. RF is slow at training but more robust and accurate than DT. SVM supports both linear and nonlinear solutions (Boateng et al., 2020).

The k-NN algorithm is a non-parametric pattern recognition method in which the unknown sample is placed in a group where most of the k nearest samples (usually an individual) around this unknown sample belong to that group (Bizzani et al, 2020; Pérez-Caballero et al., 2017).

The DT goal is to create a training model that can predict the class or value of the target variable via learning simple decision rules derived from training data. The DT algorithm is like the structure of a tree with internal nodes. Each node represents the type of test performed to determine each characteristic. Each branch shows the result(s) of the tests, and each leaf (leaf node) implies a group of samples. The classification rules in the decision tree represent a path from root to leaf. So, the decision tree has three parts: roots, internal, and leaf nodes (Pérez-Caballero et al., 2017; Saha & Manickavasagan, 2021).

The RF method is a set of different decision trees. When classifying, each decision tree of the forest determines its decision independently of the other trees. The decision that receives the most votes is considered the final result (Pérez-Caballero et al., 2017; Saha & Manickavasagan, 2021).

The function of artificial neural networks is similar to that of brain neurons. MLP, as the simplest artificial neural network model, is one of the deep learning methods in the Feed Forward Artificial Neuron group. The MLP algorithm consists of an input, hidden, and output layer (Saha & Manickavasagan, 2021).

SVM is used for cases in which research data are not linearly separable and their classification is impossible. SVM changes the problem set by changing the properties to classify the samples with new properties. In addition, SVM separates the classes so that the distance between the types reaches its maximum. For this reason, it is called a classifier with a large margin. SVM uses kernel functions such as linear, polynomial, radial basis function (RBF) (Saha & Manickavasagan, 2021; Shafiee & Minaei, 2018).

2.4. *Performance evaluation method*

In this study, the models were optimized and developed by two conventional approaches (Callao & Ruisánchez, 2018; Ropodi et al., 2016).

In the first one, the results were divided into 30% test set and the other 70% as a training set to validate the models. The second approach used the K-Fold type cross-validation method to calibrate the models.

In general, the performance of each model was evaluated by calculating the accuracy index (number of correctly classified samples to the total samples), precision index (capability of the model not to place other classes in the given class), recall or sensitivity index (capability of the model to correctly identify samples belonging to the given class), and F-score (precision and sensitivity harmonic means) (Pérez-Caballero et al., 2017; Ropodi et al., 2016).

2.5. *Software*

PCA tool of PLS toolbox (Manson) was employed in MATLAB 2013b (Natick, 2019), as an orthogonal transformation to convert a set of correlated variables into a set of uncorrelated variables (principal components), in order to preprocess and to explore the data in the unsupervised part. Python 3.7.4 (6.3.0 Jupyter) and packages of Pandas, Numpy, Scikit-learn, and Matplotlib, Skopt were applied for data preprocessing and implementation of supervised learning techniques.

3. Results and Discussion

This investigation can be divided in three-part: A discussion of the spectral difference of adulterated solutions, unsupervised elaboration, and supervised analyses and classification.

3.1. *The spectral difference of adulterated solutions*

Fig. 1 shows the spectra of solutions α, β, and γ. Accordingly, the absorption start points are somewhat different: Solution α at 260 nm, Solution β at 310 nm, and solution γ after 310 nm.

Fig. 2a is the raw spectra of lime samples. Natural lime juice samples have UV/Vis absorption in the areas of 270 to 275 nm and 300 to 345 nm. Furthermore, absorption intensity in the same wavelengths as the adulterated solutions is also observed in their spectrum. The main organic compounds in lime juice are citric acid, malic acid, isocitric acid, sugars (glucose, fructose, and sucrose), and phenolic compounds. Phenolic compounds effectively absorb 270-345 nm (Sanches et al., 2022) and organic acids and lime juice sugars absorb light at 210-270 nm (L. Kaijanen, 2015; Yu et al., 2020). Adding α, β, and γ solutions to prepare adulterated samples varies the concentration of all compounds. This not only changed their peak intensity but also affected other absorption. The average absorption for each group of lime juice is shown in Fig. 2b, to imply such changes in the spectra better.

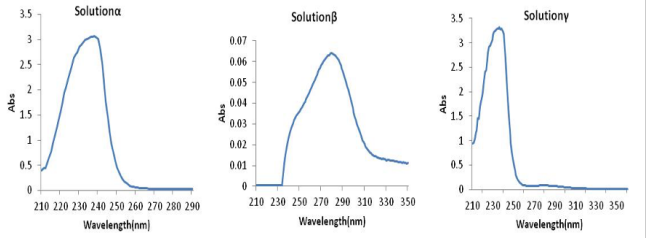

Fig. 1. Absorption spectra of solutions α, β, and γ used for the preparation of adulterated samples of lime juice.

3.2. *Unsupervised analyses*

In the unsupervised elaboration part, the samples with added solution γ were excluded from the data matrix to reduce the overlapping factors between groups. The PCA score plot is represented in Fig. 3a, illustrating the differentiation perfectly of the other four groups. The two first main components in the score plot (PC1, PC2) cumulated 21% of the variance. Although the PCA model was acquired for 20 PCs contains 76% of the total variance. Q Residual Reduced and Hoteling T2 Reduced with a 95%

confidence limit, include 24% of the total variance remains in the residuals and 76% of the total variance represented in the PCA model, respectively (Fig. 3b). This data helps to explain how well a model is describing a given sample and why that sample has its observed scores in the model. According to the influence plot of Fig. 3b, the outliers of the model belong to A2, A3, and A5 groups, mostly. The model outliers are spots outside the confidence level along the Q axis, and high-scoring points are spots outside the confidence level along the T2 axis.

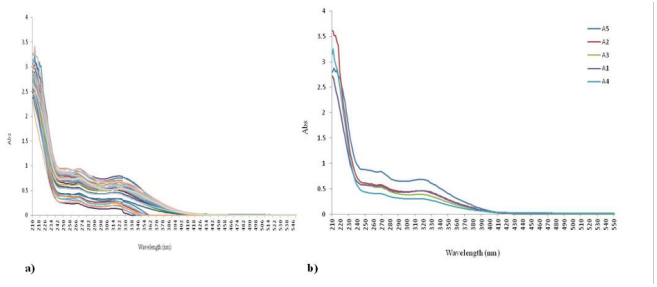

Fig. 2. Absorption spectra of lime samples. (a) Average absorption spectra of lime juice based on the group of samples: A1 (water added), A2 (acid added), A3 (Sugar added), A4(Mix A1, A2, and A3), and A5 (authentic) group. (b)

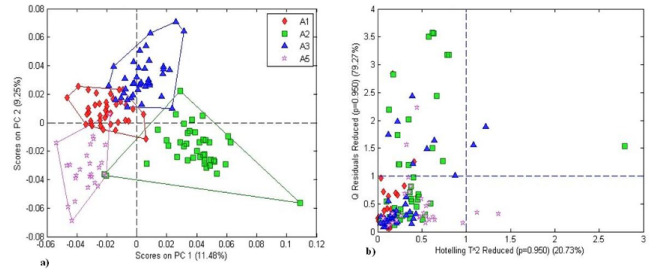

Fig. 3. Score plots of the first two PCs (PC2 vs. PC1) absorption spectra of four groups of lime juice samples, A1 (water added), A2 (acid added), A3 (Sugar added), and A5 (authentic) group (a); Q Residual Reduced vs. Hoteling T2 Reduced (b).

3.3. Supervised analyses and classification

For supervised analyses and classification, k-NN, DT, RF, MLP, and SVM were employed. Fig. 4 shows changes in the accuracy of the k-NN model for test and training sets for k = 5. The model accuracy for test and training sets was 83% and 84%, respectively. Its accuracy was 79% after cross-validation (Table 1 and Fig. 5a), which was lower than SVM (92%) and MLP (87%) models. The sensitivity and F-score in identifying group A3 were lower than other groups, while the precision of its identification, with this classification method, was the highest (Fig. 5).

Using k-NN for products such as meat, coffee, etc. has been reported (Saha & Manickavasagan, 2021). It is also used for beverages. This method has been used for monitoring the amount of pectin with MIR and Td-NMR spectroscopy in orange juice. The accuracy of this model was 85% when the data was divided into a 30% test set and 70% training set to validate the model (Bizzani et al., 2020). This method was able to classify three traditional Mexican drinks with an accuracy of 98% in another study (Pérez-Caballero et al., 2017).

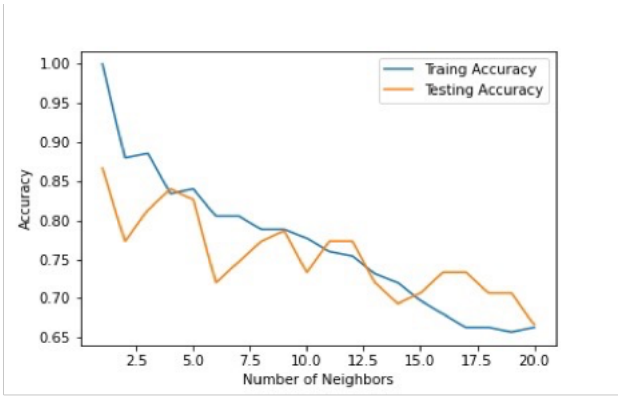

Fig. 4. Changing of train and test accuracy based on k-value.

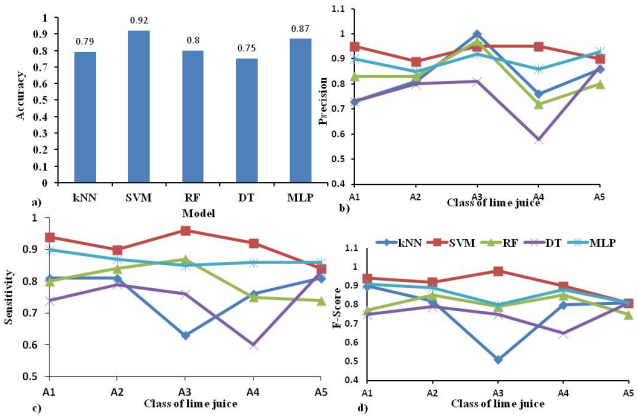

Fig. 5. Comparison of accuracy (a), precision (b), sensitivity (c), and F-score (d) value of different models in separating five groups of lime juice samples using absorption spectra. A1 (Water added), A2 (Citric acid added), A3 (Sugar added), A4 (Mix A1, A2, and A3), and A5 (authentic) group.

For the decision tree, the optimal value of the parameters affecting the accuracy of the model, such as the maximum depth of the tree and the minimum number of samples required to split an internal node, should be determined. The value of these factors selected as default in Python for this model is none and 2 respectively. The model accuracy is 80% and 75% in validation and calibration, respectively, to classify into 5 groups (Table 1). The DT classification has the lowest level of accuracy in separating adulterated lime juice samples (Table 1 and Fig. 5a). The precision of this model was also the lowest compared to all other models in identifying all five classes of lime juice samples Fig. 5b. As well as, the value of F-score and sensitivity of this model were lower than other models in classifying four groups: A1, A2, A4, and A5 (Fig. 5c & d). The sensitivity, accuracy, and F score in the model obtained from this method were consistent, and A4 and A5 among all classes had the lowest and highest value in sensitivity, precision, and F-score, respectively (Fig. 5).

This method was used for various studies in the food industry, such as classifying honey according to its geographical origin with 54% accuracy (Chudzinska & Baralkiewicz, 2011), and identifying added other oils to olive oil, with receiver operating characteristic (ROC) of 98% (Dankowska & Kowalewski, 2019), the authentication of organic grape juice with 86% accuracy using inductively coupled plasma mass spectrometry (ICP-MS) data

(Maione et al., 2016), and determining the types of traditional Mexican drinks, with 98% accuracy (Pérez-Caballero et al., 2017).

Another classification technique used in this study was RF to classify lime juice samples using UV/Vis absorption data. The set of trees (forests) in decision trees was 100 according to the Python Random Forest function default. The validation and calibration accuracies of the model were 78% and 80%, respectively (Table 1), which was more accurate than k-NN (79%) and DT (75%) (Fig. 5a).

RF has been used for the quality control and authentication of many food products (Dasenaki & Thomaidis, 2019; Saha & Manickavasagan, 2021). Some of them were about juices and beverages, such as determining the cultivation area of Argentine lemon (Gaiad et al., 2016) with 71% accuracy, classifying three types of a traditional Mexican drink (Pérez-Caballero et al., 2017) with 98% accuracy, distinguishing natural and synthetic lime juice (Shafiee & Minaei, 2018) was as accurate as 70% accuracy, and identifying additives in orange juice by electronic noise (Qiu & Wang, 2017) with 90% accuracy.

The first layer had 341 nodes in MLP model, based on the number of absorptions read for each sample at 210-550 nm. The hidden layer had 14 nodes or neurons, and output layer had five nodes, according to the number of groups of the lime juice samples (A1, A2, A3, A4, and A5). Solver = lbfgs and Random state = 1 in Python were selected to acquire the model. The validation and calibration accuracies of the model were higher than k-NN, DT, and RF (Table 1). MLP could classify samples of different groups of lime juice with 87% accuracy, which ranks second after the SVM model (Fig. 5a). The difference in values between groups obtained for sensitivity, precision, and F- score was insignificant (Fig. 5). Fig. 6 also compares the class separation decision boundary for MLP and three other models. Due to high accuracy, MLP had a smoother and clearer decision boundary than DT, RF, and k-NN.

MLP has been used for quality control of food products such as coffee (Barbosa et al., 2014) , honey, flour, and rice with over 92% accuracy (Saha & Manickavasagan, 2021). This method could authenticate organic and ordinary grape juices using ICP-MS data as accurately as 86%. (Maione et al., 2016).

Optimal values for the kernel (linear, polynomial, and RBF), gamma, and degree are important terms in the SVM method. Typically, machine learning techniques in the food industry specify the kernel, followed by a grid search cross-validation method to determine the optimal values of the gamma and degree parameters (Bizzani et al., 2020; Ropodi et al., 2016; Saha & Manickavasagan, 2021). We leveraged the Bayesian cross-validation scheme to find the optimal values for the parameters as mentioned earlier.

As demonstrated in Fig. 7, the highest accuracy of the SVM was obtained with RBF kernel when c, gamma, and degree parameters are equal to 5.55 and 51.08, and 1, respectively. Fig. 7 shows the optimum values of these parameters in optimizing the SVM model. Table 1 shows this model's validation and calibration accuracy were 88% and 92%, respectively.

This model had the highest accuracy (92%) in separating 5 groups of lime juice samples among all the supervised methods used in this study.

The SVM has the highest F-score (96%) and sensitivity (98%) for A3 adulterated samples. Also, as shown in Table 1 and Fig. 5, SVM had the highest F-score, sensitivity, and precision compared to other models in identifying four groups of adulterated samples. However, the model F-score (84%) and precision (90%) for identifying authentic samples were slightly lower than MLP.

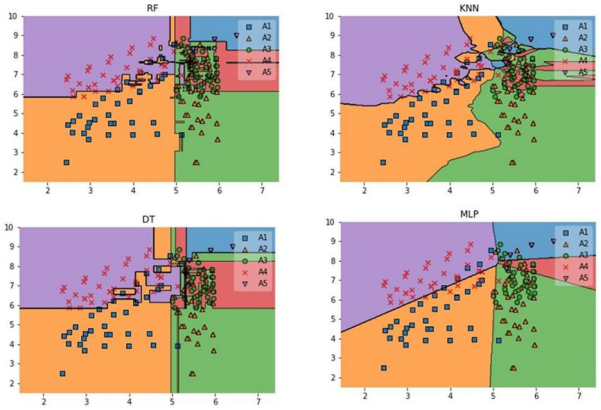

Fig. 6. Decision regions of RF, k-NN, DT, and MLP Classifier for the separation of five groups of lime juice samples. A1 (Water added), A2 (Citric acid added), A3 (Sugar added), A4 (Mix A1, A2, and A3), and A5 (authentic) group.

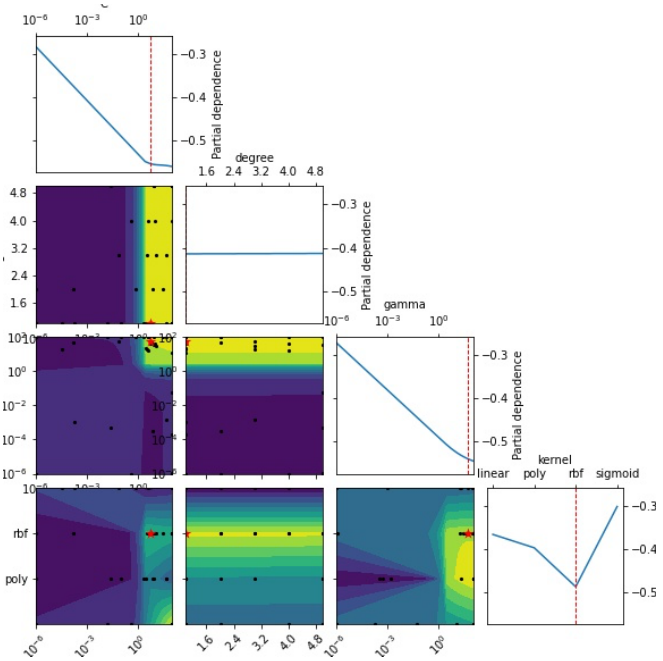

Fig. 7. Bayesian cross-validation to find the optimal values for effective parameters (kernel type, gamma, and c) of SVM model in classification lime juice samples.

The SVM is one of the methods used in many studies on the quality control of food products (Jiménez-Carvelo et al., 2019). The models obtained from SVM and MLP had high accuracy in most of these studies. This has caused these two deep-learning techniques to become great methods in lime juice quality control and authentication.

Table 1. Performance result of supervised learning techniques based on UV/Vis spectra data `for classification adulterated lime juice samples.

Classifier	Group of lime juice samples	Validation	Calibration (cross validation)			
		Accuracy	Accuracy	Precision	Sensitivity	F-score
KNN	A1	0.83	0.79	0.73	0.9	0.81
	A2			0.81	0.82	0.81
	A3			1	0.51	0.63
	A4			0.76	0.8	0.76
	A5			0.86	0.81	0.81
SVM	A1	0.88	0.92	0.95	0.94	0.94
	A2			0.89	0.92	0.9
	A3			0.95	0.98	0.96
	A4			0.95	0.9	0.92
	A5			0.9	0.81	0.84
RF	A1	0.78	0.8	0.83	0.77	0.8
	A2			0.83	0.85	0.84
	A3			0.97	0.79	0.87
	A4			0.72	0.85	0.75
	A5			0.8	0.75	0.74
DT	A1	0.8	0.75	0.73	0.75	0.74
	A2			0.8	0.79	0.79
	A3			0.81	0.75	0.76
	A4			0.58	0.65	0.6
	A5			0.87	0.81	0.83
MLP	A1	0.85	0.87	0.9	0.91	0.9
	A2			0.85	0.89	0.87
	A3			0.92	0.8	0.85
	A4			0.86	0.88	0.86
	A5			0.93	0.81	0.86

A1(Water added), A2(Citric acid added), A3(Sugar added), A4(Mix A1, A2, and A3), and A5(authentic) group.

4. Conclusion

Exploring the data by PCA revealed the four clusters separately while plotted in the PC1-PC2 score plot, which cumulated 24% of the variance. PC1 and PC2 are the model components that contained 76% of the variables, and the value of Hotelling's T2 and Q Residual was not satisfactory. The clustering revealed by PCA with such results indicates the power of this method in the early recognition of such studies. Various classification methods were used in the supervised investigation section (k-NN, RF, DT, SVM, and MLP) to determine the type of lime juice adulteration. SVM and MLP detected at least 10% water, acid, and sugar added to lime juice with the highest accuracy than others (92% and 87%, respectively), while DT had the lowest accuracy with 75%. According to these results, applying machine learning methods to the spectroscopy data can be considered a green, fast, and non-destructive method in controlling the quality and determining the type of adulteration of lime juice with high accuracy.

Acknowledgment

The authors would like to thank the University of Tehran's Vice-chancellor of Research and Technology. We are also grateful to Iran National Science Foundation (INSF). This work is based upon research founded by INSF with project No.: 4000794.

Sincerely appreciate the advice of Mr. Goleayn, a faculty member of Iran's citrus and subtropical fruits center.

Conflict of interest

The authors declared that they have no conflict of interest.

References

Adeli, H., Khorasani, M. T., & Parvazinia, M. (2019). Wound dressing based on electrospun PVA/chitosan/starch nanofibrous mats: Fabrication, antibacterial and cytocompatibility evaluation and in vitro healing assay. *International Journal of Biological Macromolecules, 122*, 238-254.

AIJN. (2016). Cod of practice for evaluation of fruit and vegetable juices 6.26. reference guideline for lime juice.AliAbadi, M. H. S., Karami-Osboo, R., Kobarfard, F., Jahani, R., Nabi, M., Yazdanpanah, H., . . . Faizi, M. (2022). Detection of lime juice adulteration by simultaneous determination of main organic acids using liquid chromatography-tandem mass spectrometry. *Journal of Food Composition and Analysis, 105*, 104223.

Barbosa, R. M., Batista, B. L., Varrique, R. M., Coelho, V. A., Campiglia, A. D., & Barbosa, F. (2014). The use of advanced chemometric techniques and trace element levels for controlling the authenticity of organic coffee. *Food Research International, 61*, 246-251. doi:https://doi.org/10.1016/j.foodres.2013.07.060

Boateng, E. Y., Otoo, J., & Abaye, D. A. (2020), Basic Tenets of Classification Algorithms K-Nearest-Neighbor, Support Vector Machine, Random Forest and Neural Network: A Review. *Journal of Data Analysis and Information Processing, 08*(04), 341-357.

Bizzani, M., William Menezes Flores, D., Alberto Colnago, L., & David Ferreira, M. (2020). Monitoring of soluble pectin content in orange juice by means of MIR and TD-NMR spectroscopy combined with machine learning. *Food Chemistry, 332*, 127383. doi:https://doi.org/10.1016/j.foodchem.2020.127383

Boggia, R., Casolino, M. C., Hysenaj, V., Oliveri, P., & Zunin, P. (2013). A screening method based on UV–Visible spectroscopy and multivariate analysis to assess addition of filler juices and water to pomegranate juices. *Food Chemistry, 140*(4), 735-741.

Callao, M. P., & Ruisánchez, I. (2018). An overview of multivariate qualitative methods for food fraud detection. *Food Control, 86*, 283-293.

Chang, J. D., Zheng, H., Mantri, N., Xu, L., Jiang, Z., Zhang, J., . . . Lu, H. (2016). Chemometrics coupled with ultraviolet spectroscopy: a tool for the analysis of variety, adulteration, quality and ageing of apple juices. *International Journal of Food Science & Technology, 51*(11), 2474-2484.

Chudzinska, M., & Baralkiewicz, D. (2011). Application of ICP-MS method of determination of 15 elements in honey with chemometric approach for the verification of their authenticity. *Food and Chemical Toxicology, 49*(11), 2741-2749. doi:https://doi.org/10.1016/j.fct.2011.08.014

Dankowska, A., & Kowalewski, W. (2019). Comparison of different classification methods for analyzing fluorescence spectra to characterize type and freshness of olive oils. *European Food Research and Technology, 245*(3), 745-752. doi:10.1007/s00217-018-3196-z

Dasenaki, M. E., & Thomaidis, N. S. (2019). Quality and authenticity control of fruit juices-A review. *Molecules, 24*(6), 1014.

FAO. (2020). *The Citrus Bulletin*.

Fidelis, M., Santos, J. S., Coelho, A. L. K., Rodionova, O. Y., Pomerantsev, A., & Granato, D. (2017). Authentication of juices from antioxidant and chemical perspectives: A feasibility quality control study using chemometrics. *Food Control, 73*, 796-805.

Gaiad, J. E., Hidalgo, M. J., Villafañe, R. N., Marchevsky, E. J., & Pellerano, R. G. (2016). Tracing the geographical origin of Argentinean lemon juices based on trace element profiles using advanced chemometric techniques. *Microchemical Journal, 129*, 243-248. doi:https://doi.org/10.1016/j.microc.2016.07.002

González-Molina, E., Domínguez-Perles, R., Moreno, D. A., & García-Viguera, C. (2010). Natural bioactive compounds of Citrus limon for food and health. *Journal of pharmaceutical and biomedical analysis, 51*(2), 327-345. doi:https://doi.org/10.1016/j.jpba.2009.07.027

Guyon, F., Auberger, P., Gaillard, L., Loublanches, C., Viateau, M., Sabathié, N., . . . Médina, B. (2014). 13C/12C isotope ratios of organic acids, glucose and fructose determined by HPLC-co-IRMS for lemon juices authenticity. *Food Chemistry, 146*, 36-40.

Jandrić, Z., & Cannavan, A. (2017). An investigative study on differentiation of citrus fruit/fruit juices by UPLC-QToF MS and chemometrics. *Food Control, 72*, 173-180.

Jiménez-Carvelo, A. M., González-Casado, A., Bagur-González, M. G., & Cuadros-Rodríguez, L. (2019). Alternative data mining/machine learning methods for the analytical evaluation of food quality and authenticity – A review. *Food Research International, 122*, 25-39. doi:https://doi.org/10.1016/j.foodres.2019.03.063

Jittanit, W., Suriyapornchaikul, N., & Nithisopha, S. (2013). The comparison between the quality of lime juices produced by different preservation techniques. *Procedia-Social and Behavioral Sciences, 91*, 691-696.

L. Kaijanen, M. P., S. Pietarinen, E. Jernström, S. Reinikainen. (2015). Ultraviolet Detection of Monosaccharides: Multiple Wavelength
Strategy to Evaluate Results after Capillary Zone
Electrophoretic Separation. *Int. J. Electrochem. Sci, 10*, 2950 - 2961.

Lorente, J., Vegara, S., Martí, N., Ibarz, A., Coll, L., Hernández, J., . . . Saura, D. (2014). Chemical guide parameters for Spanish lemon (Citrus limon (L.) Burm.) juices. *Food Chemistry, 162*, 186-191.

Lubinska-Szczygieł, M., Różańska, A., Namieśnik, J., Dymerski, T., Shafreen, R. B., Weisz, M., . . . Gorinstein, S. (2018). Quality of limes juices based on the aroma and antioxidant properties. *Food Control, 89*, 270-279.

Lyu, W., Yuan, B., Liu, S., Simon, J. E., & Wu, Q. (2022). Assessment of lemon juice adulteration by targeted screening using LC-UV-MS and untargeted screening using UHPLC-QTOF/MS with machine learning. *Food chemistry, 373*, 131424.

Maione, C., de Paula, E. S., Gallimberti, M., Batista, B. L., Campiglia, A. D., Jr, F. B., & Barbosa, R. M. (2016). Comparative study of data mining techniques for the authentication of organic grape juice based on ICP-MS analysis. *Expert Systems with Applications, 49*, 60-73. doi:https://doi.org/10.1016/j.eswa.2015.11.024

Natick. (2019). MATLAB. *MA, USA: The Mathworks, Inc*.

Pérez-Caballero, G., Andrade, J., Olmos, P., Molina, Y., Jiménez, I., Durán, J., . . . Miguel-Cruz, F. (2017). Authentication of tequilas using pattern recognition and supervised classification. *TrAC Trends in Analytical Chemistry, 94*, 117-129.

Qiu, S., & Wang, J. (2017). The prediction of food additives in the fruit juice based on electronic nose with chemometrics. *Food Chemistry, 230*, 208-214. doi:https://doi.org/10.1016/j.foodchem.2017.03.011

Ríos-Reina, R., Azcarate, S. M., Camiña, J., Callejón, R. M., & Amigo, J. M. (2019). Application of hierarchical classification models and reliability estimation by bootstrapping, for authentication and discrimination of wine vinegars by UV–vis spectroscopy. *Chemometrics and Intelligent Laboratory Systems, 191*, 42-53.

Rivera-Cabrera, F., Ponce-Valadez, M., Sanchez, F., Villegas-Monter, A., & Perez-Flores, L. (2010). Acid limes. A review. *Fresh produce, 4*(1), 116-122.

Ropodi, A., Panagou, E., & Nychas, G.-J. (2016). Data mining derived from food analyses using non-invasive/non-destructive analytical techniques; determination of food authenticity, quality & safety in tandem with computer science disciplines. *Trends in Food Science & Technology, 50*, 11-25.

Saha, D., & Manickavasagan, A. (2021). Machine learning techniques for analysis of hyperspectral images to determine quality of food products: A review. *Current Research in Food Science, 4*, 28-44. doi:https://doi.org/10.1016/j.crfs.2021.01.002

Sanches, V. L., Cunha, T. A., Viganó, J., de Souza Mesquita, L. M., Faccioli, L. H., Breitkreitz, M. C., & Rostagno, M. A. (2022). Comprehensive analysis of phenolics compounds in citrus fruits peels by UPLC-PDA and UPLC-Q/TOF MS using a fused-core column. *Food Chemistry: X, 14*, 100262. doi:https://doi.org/10.1016/j.fochx.2022.100262

Shafiee, S., & Minaei, S. (2018). Combined data mining/NIR spectroscopy for purity assessment of lime juice. *Infrared Physics & Technology, 91*, 193-199.

Yu, C., Wang, Y., Cao, H., Zhao, Y., Li, Z., Wang, H., . . . Tang, Q. (2020). Simultaneous Determination of 13 Organic Acids in Liquid Culture Media of Edible Fungi Using High-Performance Liquid Chromatography. *BioMed Research International, 2020*, 2817979. doi:10.1155/2020/2817979